DEVELOPMENT OF OLIGONUCLEOTIDE-BASED MICROARRAYS FOR THE DETECTION OF PLANT VIRUSES AND THE FIRST CHARACTERIZATION OF A PLANT VIRUS BELONGING TO THE FAMILY *TOTIVIRIDAE*

By

VEENITA GROVER SHAH

Bachelor of Science in Biochemistry
University of Delhi
Delhi, India
2003

Master of Science in Biosciences
Jamia Milia Islamia University
Delhi, India
2005

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2010

DEVELOPMENT OF OLIGONUCLEOTIDE-BASED MICROARRAYS FOR THE DETECTION OF PLANT VIRUSES AND THE FIRST CHARACTERIZATION OF A PLANT VIRUS BELONGING TO THE FAMILY *TOTIVIRIDAE*

Dissertation Approved:

Dr. Ulrich K. Melcher

Dissertation Adviser

Dr. Marilyn J. Roossinck

Dr. Michael W. Palmer

Dr. Peter Hoyt

Dr. Patricia Ayoubi

Dr. Mark E. Payton

Dean of the Graduate College

DEDICATION

to my

Father Mr. **Madan Grover**

and

Mother Mrs. **Veena Grover**

ACKNOWLEDGMENTS

I take this opportunity to thank a number of outstanding individuals and scholars who helped me throughout the period of my graduate program at Oklahoma State University. I will always cherish my memory of all those, to only some of whom it is possible to give particular mention here, who helped in enriching the experiences of my life during this time.

This thesis would have not have been possible without the help, support and patience of my major adviser, Dr. Ulrich K. Melcher. I would like to extend my deep gratitude and appreciation to Dr. Melcher for his constant support in guiding me to a deeper understanding of knowledge work, and his invaluable supervision and encouragement during my entire Ph.D. study and research. His mentoring helped me in all the time of research and writing of this thesis. I owe my most sincere gratitude to Dr. Marilyn J. Roossinck for providing me with the opportunity to work at the Noble Foundation, as a part of her group and for her constant guidance, help and motivation throughout the term of my research. I further thank the rest of my thesis committee: Dr. Michael Palmer, Dr. Peter Hoyt and Dr. Patricia Ayoubi for their insightful comments and advice during the course of this work.

An expression of gratitude goes to all my friends and fellow lab members at Oklahoma State University and Noble Foundation for their helpful suggestions, encouragements and contribution of humor during my academic endeavors. A special note of thanks is due to the faculty and staff members of the Department of Biochemistry and Molecular Biology at Oklahoma State University and Plant Biology Division at Noble Foundation, Dr. Rick S. Nelson in particular, for

all their help and assistance.

I would like to express my deepest and most sincere gratitude to my parents Mr. Madan Grover and Mrs. Veena Grover, and other family members (Mr. Dhananjay Shah, Mrs. Rashmi Shah, Mr. Vidur Grover, Mrs. Neha Grover, Ms. Heta Shah and Ms. Yuvika Grover) for their incessant love, affection and encouragement which enabled the completion of this degree. Finally, I would like to thank and appreciate my wonderful husband, Mr. Viraj Shah for his tremendous support, endless patience and continual dosage of love and stimulation all through my graduation which made it possible for me to achieve my goal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

GENERAL INTRODUCTION

Most of the recognized groups of living organisms harbor viruses, which are obligate symbionts known to depend on their hosts for their replication and survival. Studies, primarily in marine environments, have revealed an enormous diversity of DNA as well as RNA viruses, indicating viruses as one of the most abundant biological entities (Azam & Worden, 2004; Breitbart & Rohwer, 2005b; Fuhrman, 1999; Suttle, 2005; Suttle, 2007). The discovery of the virus that marked the beginning of the science of virology was made in the 1890s with the detection of a causal agent, later named as *Tobacco mosaic virus*, of a disease causing mosaic and distortion on tobacco plants (Beijerinck, 1898; Iwanowski, 1892; Mayer, 1886). *Tobacco mosaic virus* played a significant role in the history of virology for many other reasons apart from being the first virus to be discovered. These include it being the first virus to be purified (Stanley, 1935), determination of the chemical composition of the virus, the isolation of its protein and nucleic acid components, reconstitution from its dissociated parts and determination of the sequence of the first viral coat protein. Despite the fact that with the further discovery of every novel plant virus and their characteristics, new conceptual grounds are laid, the view of viruses as pathogens still dominates the field of plant virology due to an evident bias in plant virus studies towards acute and chronic viruses. This perception has less to do with what we know about viruses and more to do with ignorance of what we do not know. Virus studies in wild plant communities are often underestimated relative to cultivated systems since viruses in wild plants are generally considered not to harm the host. It is likely that most viruses are commensals, demanding no apparent fitness cost from the host (Roossinck, 2005). The growing list of viruses being discovered in wild plants

suggests an important role for viruses in our ecosystem. Due to the symbiotic associations of viruses with their hosts, a role of viruses in the evolution of the host is suggested by researchers (Filée *et al.*, 2003; Villarreal, 2005). Recent examples of beneficial effects of viruses on their host plants are slowly but steadily broadening the horizons of plant virology.

It has been long known that viruses can be latent or symptomless in some host plants even though they may cause symptomatic diseases in others. Symptoms of virus infection can vary greatly in expression and severity depending on a number of factors including age of the host, host cultivar, strain of the virus, environmental conditions and the presence or absence of other viruses (Department of Crop Sciences, 1991 ; Zitter & Murphy, 2009). In the phenomenon of tolerance (Kang *et al.*, 2005), infected plants may have high titers of virus without exhibiting obvious symptoms of infection. In such cases, viruses may move through the plant in a manner similar to that in susceptible hosts, but disease symptoms are just not observed. It is possible that such plants are nevertheless affected by the virus through a decrease in fitness not detectable by simple visual observation (Pagan *et al.*, 2008). Most of the viral sequences discovered during a recent plant virus biodiversity survey, Plant Virus Biodiversity and Ecology project (PVBE), came from asymptomatic, healthy-looking plant samples (Melcher *et al.*, 2008; Muthukumar *et al.*, 2009; Roossinck *et al.*, 2010), again supporting the idea that the presence or absence of symptoms should not be related to virus infection. The mechanisms by which these symptoms are generated, and by which plants resist these effects, are still largely unknown.

Since viruses are considered one of the most agriculturally important and biologically intriguing groups of plant pathogens, it was not surprising that most of the viruses cataloged in the Viral Identification Data Exchange database (1996) were derived from cultivated crop species and symptomatic hosts (Wren *et al.*, 2006). Although our knowledge of viruses of non-cultivated plants remains far from comprehensive, there have been some surveys for viruses in plant populations from non-managed ecosystems (Bodaghi *et al.*, 2004; Fraile *et al.*, 1997; Kawakami

et al., 2007; Raybould *et al.*, 1999). These surveys studied the distribution of several known crop plant viruses in non-cultivated plants using specific assays to screen the plants for those viruses. Most of the unknown viruses in non-cultivated plants caught scientists' attention and were discovered because the plants exhibited novel disease symptoms (Ciuffo *et al.*, 2008; Gibbs, 1980; Hassan *et al.*, 2009; Ooi *et al.*, 1997; Robertson, 2005, 2007). The finding of the first vascular plant virus, Stilbocarpa mosaic bacilliform badnavirus (SMBV) causing mild to severe yellow mosaic symptoms, in an extreme environment of the subantarctic Macquarie Island (Skotnicki *et al.*, 2003) was followed by successful detection of several other known plant viruses in host plants growing in the near Antarctic region (Polischuk *et al.*, 2007).

Recent ecogenomic studies of plant viruses in the wild conducted at the Area de Conservacion Guanacaste (ACG) in Costa Rica (Roossinck *et al.*, 2010) and the Tallgrass Prarie Preserve (TGPP) in northeastern Oklahoma (Melcher *et al.*, 2008; Muthukumar *et al.*, 2009; Roossinck *et al.*, 2010) targeted native plants regardless of the presence or absence of symptoms, and discovered several hundreds of new potential viral species. These studies were able to link each of the obtained viral sequence to its specific plant host in a terrestrial system. Only very few viruses with sequence similarities high enough to be considered a strain of a known species were discovered from these studies. In most of the other cases, similarity was only high enough to indicate that the potential virus in the plant belonged to the same genus or family as the matched virus sequence in GenBank. These results supported the concept originally suggested by Harrison in 1981 (Harrison, 1981) of categorizing viruses of non-cultivated plants (WILPAD, wild-adapted) and crop plants (CULPAD, cultivated-adapted) into two groups based on their adaptation to existence. Virus taxa commonly associated with crop plant diseases were relatively absent in these non-cultivated plants. The exception to this case was one particular crop plant virus, *Zucchini yellow mosaic virus,* a member of the family *Potyviridae* that was found with high frequency in the ACG in Costa Rica (Saha *et al.*, Unpublished data).

One of the high incidence sets of viruses found in the TGPP were from the genus *Endornavirus*, a group of persistent plant viruses. The term "persistent virus" in plants reflects one of the four lifestyles a virus can adopt, namely persistent, acute, chronic and endogenous (Roossinck, 2010). The important characteristics of a persistent plant virus are 1) it usually causes no or very slight disease symptoms, 2) it is present in low concentrations in the host, and 3) it is not horizontally transmitted 4) it does not move from cell to cell due to the lack of movement protein (MP) and 5) it is seed transmitted, and maintains infection for the lifetime of the host. Though all of the viruses belonging to this group have double-stranded (ds) RNA genomes, this characteristic is not a criterion for persistence. In animals, life of a persistent virus is suggested to begin with an initial phase of productive infection and antiviral host responses. However, the virus is not fully cleared out by these host responses, and maintains the capacity of constant or periodic reproduction (Villarreal *et al.*, 2000). These viruses establish species-specific persistent and inapparent infections that are stable on an evolutionary time scale. The persistence of a virus tends to be highly host specific, possibly due to the need for a close coordination of the virus with host regulatory systems (Villarreal, 2005). Virus persistence appears to rely on host mechanisms for virus maintenance and probably for competition and exclusion of other viral agents as well. Currently, the mechanisms of virus persistence and maintenance, and their consequence for the infected host are not very clear.

The best studied of the plant persistent viruses are those previously called cryptic viruses (Boccardo *et al.*, 1987). Cryptic plant viruses are seed-borne dsRNA-viruses which co-exist with their host plants throughout their lives without inducing any symptoms. Cryptic viruses went undetected for a long period of time due to lack of research attention since they do not appear to cause any economically important diseases. This sadly is still the case with persistent plant viruses which remain one of the most poorly characterized kinds of plant viruses. Cryptic viruses were first noticed in plants in the late 1960s by Pullen, who observed small virus-like particles (VLPs) in all plants of seven beet species (Pullen, 1968; Pullen, 1969). These particles were

neither mechanically transmittable to other herbaceous plants, nor eliminated from the host by heat treatment. They were suspected to be an obligatory constituent of plants until a couple of years later, when a few beet plants not containing the VLPs were found. The VLPs were purified then and named as *Beet cryptic virus* (BCV) (Kassanis et al., 1977). The virus was shown to be present in 90% of plants from three different sugar beet cultivars at a concentration of 1µg or less per gram of leaf tissue. Kassanis et al. also established that BCV was seed transmissible via both the ovule and the pollen (Kassanis *et al.*, 1978). Lisa et al. 1981 first showed that particles of *Carnation cryptic virus* contained double stranded (ds) RNA genomes, as was also true for other cryptic viruses in alfalfa, meadow fescue, hop trefoil, red clover, white clover and ryegrass (Boccardo *et al.*, 1987; Lisa *et al.*, 1981). Cryptic viruses were confirmed in 1985 to be plant viruses and not viruses of fungi contaminating the host plants by Boccordo et al. (Boccardo *et al.*, 1985). The report showed that the incidence and concentration of *White clover cryptic virus* was unaffected by systemic and surface fungicides. Viruses were recovered in similar amounts from seedlings grown from surface-sterilized seeds. Further reports provided supporting evidence favoring a lack of association of cryptic viruses with fungi (Boccardo *et al.*, 1987). One of the not so infrequent characteristics of cryptic viruses is that two or three viruses can be found in an individual plant host. However, they do not appear to be dependent on each other.

Cryptic viruses now belong to the genus *Alphacryptovirus* of the family *Partitiviridae*, which include viruses not only in plants but also in fungi and more recently protozoa (Nibert *et al.*, 2009). In algae, the association of their dsRNAs with chloroplasts and mitochondria raised the possibility of prokaryotic origin (Ishihara *et al.*, 1992). Cryptic viruses have a small segmented double-stranded RNA genome in the size range of 1-3 Kbp per segment. In plants, cryptic viruses appear to persist indefinitely. They seem to be so well adapted to their hosts that they can survive dramatic changes in growth and environmental conditions. A study with *Carnation cryptic virus* and *Beet cryptic virus* showed that 16 years of continuous *in vitro* tissue culturing and thermotherapy were unable to cure plants of these partitiviruses (Szegö *et al.*, 2005). Presence of

Pepper cryptic virus in all the tested cultivars of Jalapeño peppers indicated its presence at the time of origin of the Jalapeño cultivar (Arancibia *et al.*, 1995). The other group of persistent plant viruses mentioned above, the Endornaviruses, infect plants, fungi and oomyctes, and are similar to cryptoviruses in that 1) they are efficiently transmitted through seed, 2) most of them do not show horizontal spread, and 3) they are not associated with disease symptoms. However, unlike cryptoviruses, which contain virions and consist of two dsRNAs each about 1-3 kb in length, none of the endornaviruses are associated with particles and each of them is associated with a large non-segmented ds RNA of more than 10 kbp, encoding a single open reading frame with recognizable RNA helicase and polymerase domains. A recent phylogenetic analysis of the available isolates of Phytophthora endornaviruses from Europe and USA showed that these viral isolates are highly conserved even in different hosts suggesting a population bottleneck during their emergence (Kozlakidis *et al.*, 2010). Like partitiviruses, endornavirus persistence seems to be very stable as well. All cultivars of bell pepper contain an endornavirus that can be vertically transmitted to other cultivars of pepper through crosses (Valverde & Gutierrez, 2007). It is suggested that crosses among different pepper genotypes by plant breeders in order to generate new cultivars may have resulted in the spread of the bell pepper-dsRNA to different horticultural types of pepper. It seems pretty likely that examples of such persistent viruses exist much more commonly than recognized.

Viruses belonging to the family *Totiviridae*, known to infect fungi, protozoa and arthropods (Zhai *et al.*, 2010), have unsegmented dsRNA genomes that contain two partially overlapping open reading frames (ORFs), encoding the viral coat protein (CP) and RNA dependent RNA polymerase (RdRp) (Ghabrial & Suzuki, 2009). A potential persistent plant virus belonging to the family *Totiviridae* was found during a recent ecogenomic study of plant viruses at the TGPP (Roossinck, 2010). In addition to high incidence of infection, the putative virus was also widely spread in several different plant host families making it the second most prevalent viral family in the prairie. Due to their high abundance at various locations in the prairie, six plant

host species including *Ambrosia psilostachya, Asclepias viridis, Panicum virgatum, Ruellia humilis, Sorghastrum nutans* and *Vernonia baldwinii* were chosen for repeated sampling for four consecutive years of study (2005, 2006, 2007 and 2008). The incidence of occurrence of this putative virus was most frequent and the titers highest in the plant species *Ruellia humilis* among the six targeted host species. *R. humilis,* a species native to United States is a perennial herb belonging to the family *Acanthaceae* with the common name Wild petunia. *R. humilis* is known to tolerate a wide range of conditions including full or partial sunlight, moist to dry conditions, and almost any kind of soil. The multi-branched stem of this plant with hairy leaves may reach a height of 60 cm but is usually less than 30 cm tall. The short internodes give the plant a compact, leafy and bushy appearance. The plant blooms from late spring until fall with lavender to purple colored petals. In the field, after the seed dispersal in autumn, seeds are exposed to effective stratification temperatures during winter, and germination starts in late spring.

The viral sequence retrieved for this putative virus was similar to that of Black raspberry virus F, a dsRNA totivirus assumed, without any supporting evidence, to be a fungal virus. Given the very small amount of fungal tissue that is found in plants harboring endophytes, the number of putative totivirus sequence reads obtained from individual plant samples was high enough to suspect that it is a plant virus, replicating in the plant cells, rather than a virus of a colonizing fungal endophyte. In addition, the sampled host plants did not exhibit any apparent symptoms of infection, and the retrieved sequence information for the virus showed no obvious movement protein gene encoded by the virus, suggesting the likelihood of it being a persistent plant virus.

Several recent reports have mentioned totiviruses in plants (Alioto *et al.*, 2003; Covelli *et al.*, 2004; Cox *et al.*, 2000; Kozlakidis *et al.*, 2006; Marais *et al.*, 2009; Martin *et al.*, 2006) but no strong evidence has been provided to designate them as plant or fungal viruses. These reports have been discussed in detail in the latter part of the study. However, it still remained to be determined whether these are plant viruses or viruses of fungal endophytes colonizing the host plants. Recent studies on partitiviruses, one of the well-studied kinds of persistent plant viruses,

show that some plant partitiviruses are phylogenetically more similar to fungal partitiviruses than other plant partitiviruses (Li *et al.*, 2009; Sabanadzovic & Ghanem-Sabanadzovic, 2008; Strauss *et al.*, 2000; Veliceasa *et al.*, 2006), suggesting evolutionary connections between fungal and plant viruses. One of the proposed hypotheses is that these persistent plant viruses were originally mycoviruses that escaped their normal host and became trapped in plants during an endophytic association between fungal endophytes and plants. Hence, these viruses are unable to move cell to cell due to lack of appropriate movement proteins. The other proposed situation is that these viruses are fungal viruses which use plants as their vectors, and thus can replicate in either of the two hosts (Roossinck, 1997, 2010). Some of the phylogenetic analyses have shown evolutionary connections between plant and fungal viruses with the direction of evolution from plant viruses to fungal viruses. One such example is that of hypoviruses of plant-pathogenic fungi which were shown to be related to the plant potyviruses (Koonin *et al.*, 1991) but in this case, the CP gene seems to be lost during the transition to the new host (Koonin *et al.*, 2008). Another such example is that of *Botrytis cinerea virus F* and *Sclerotinia sclerotiorum debilitation-associated RNA virus*, which are hypothesized to be derived from a plant virus through loss of MP in the first case, and loss of MP and CP in the second case (Martelli *et al.*, 2007). In contrast to the hypovirus, plant viruses in the family *Partitiviridae* are suggested to be related to fungal viruses (Li *et al.*, 2009; Roossinck, 2010; Veliceasa *et al.*, 2006), and appear to retain their fungal character lacking the movement protein. However, in the case of ourmiaviruses, a fungal virus appears to have evolved into a typical plant virus by capturing RNA segments encoding the proteins for virion formation and systemic spread in plants (Rastgou *et al.*, 2009).

In addition to exploring the diversity and understanding the ecology of the virus, it is also important to investigate the positive roles played by viruses in the environment. The time has come to broaden our horizons and open our minds to ideas such as “If there is no selective advantage for the virus to cause disease, there might be a selective advantage for it to make its host healthy” (Powledge, 1999). The complexity of ecological systems presents considerable

challenges for experiments to assess the benefits of viruses in our ecosystem. While there are some reports on beneficial viruses in the animal kingdom (Jiu *et al.*, 2007; Renault *et al.*, 2005; Stoye, 2006; Tillmann *et al.*, 2001), there are comparatively fewer descriptions of viruses beneficial to plants. These beneficial viruses can act as obligate mutualists to conditional mutualists. Infection of *Kennedya rubicunda* plants by *Kennedya yellow mosaic virus* was shown to make the wild legume host less attractive to herbivores, increasing the plant's longevity and fitness (Gibbs, 1980). The observation of extended survival of virus-infected plants under abiotic stress (drought and cold) indicated the involvement of viruses with their hosts in potential mutualistic relationships contributing to plant fitness under such extreme conditions (Xu *et al.*, 2008). In another example, white clover plants in soil infested with fungus gnats are shown to produce more biomass and more ramets when infected with *White clover mosaic virus* than when virus-free, due to the viruses' induction of repellent volatile emissions by the plant (van Mölken, 2009).

Persistent viruses appear to share a close relationship with their hosts, hinting at crucial functions that can be played by these viruses in their hosts. One such example was observed in the case of a plant–fungus–virus interaction, where a persistent virus in an endophytic fungus is required for thermal tolerance of plants growing in geothermal soils with temperatures over 50 °C (Márquez *et al.*, 2007; Redman *et al.*, 2002). The fungus was shown to not be able to grow in high temperatures in cultures, and fungal strains cured of the virus could not provide thermal tolerance to plants, reflecting three-way mutualistic symbiosis. Due to hypovirulence, a phenomenon in which fungal viruses significantly reduce the virulence of pathogenic fungi, *Cryphonectria parasitica* has become a model system for studying fungus-virus interactions. In chestnut blight, caused by the fungus *C. parasitica*, hypoviruses in the family *Hypoviridae* are known to be most significant in controlling the pathology of the fungus, thus acting as a mutualist to plants (Milgroom & Cortesi, 2004). Recently, a hypovirulence-associated DNA mycovirus, *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus, has been reported from the plant

pathogenic fungus *Sclerotinia sclerotiorum* (Yu *et al.*, 2010). Another example is the cytoplasmic male sterility (CMS) trait of the '447' line of *Vicia faba* which is correlated with the presence of a dsRNA of high molecular mass of unknown origin (Pfeiffer, 1998). In sugar beet crops, the effect of BCV infection was investigated in field trials in 1990. The presence of BCV appeared not to affect yield loss of beets at four locations which also suffered from drought stress but yield losses were found at one site where the crop was grown in land with good moisture retention properties. Root and sugar yield were reduced by up to 17% and 20%, respectively, by BCV infection (Xie *et al.*, 1994). The question that arises here is: Do persistent viruses have a role in drought tolerance as observed with acute viruses (Xu *et al.*, 2008)?

Trifolium repens early nodulin downregulation 1 (*TrEnodDR1*), down-regulated in response to expression of the *nod* genes of *Rhizobium leguminosarum* (Suzuki *et al.*, 2001) was later shown to encode for the coat protein of *White clover cryptic virus* (Boccardo & Candresse, 2005). Artificial expression of the *White clover cryptic virus* CP gene in *Lotus japonicus* showed suppression of nodulation by the virus CP gene (Nakatsukasa-Akune *et al.*, 2005). In comparison to the control, the growth of transformants was reduced and the number of root nodules per unit root length was greatly decreased along with an increase in the concentration of endogenous abscisic acid (ABA), which controls nodulation. The results suggested the suppression of nodulation by increasing the endogenous ABA concentration, possibly by activating the plant's innate immune response. It seems likely that similar examples exist, where plants may have acquired the genes of persistent plant viruses for their own benefit, as suggested for some endogenous pararetroviral sequences (Staginnus *et al.*, 2007). These viral sequences may be widespread in the EST databases, considered as a useless fraction of the plant genome but the mRNAs may be utilized by plants as epigenetic material for several beneficial functions (Roossinck, 2010). Moreover, persistent viruses may be affecting their hosts in more subtle ways, putting a selective pressure for the persistent virus to eliminate the host colonization by other competing native viruses (Villarreal, 2005).

It has been suggested that the evolution of plant viruses has been modular (Gibbs, 1987; Roossinck, 2005). Thus it is possible that that the commonly found persistent viruses may lead to the generation of novel viruses by recombining with infecting acute viruses. This is well exemplified by a recent study on the members of plant virus genus *Ourmiavirus* which appear to have evolved from a typical fungal virus into a plant virus by capturing RNA segments encoding the proteins responsible for virion formation and systemic spread in plants (Rastgou *et al.*, 2009). This supports the potential role of persistent plant viruses in the evolution of acute viruses and broadens our vision on the diversification of the ways of viral evolution. In the case of animals, asymptomatic persistent viruses can act as a source of emerging acute diseases in new host species. There exist examples which support the idea that some of the acute epidemic agents originated from persistent infections of other hosts. One such example is that of a *human influenza A virus* which is not persistent and displays high genetic variability, but all 14 HA subtypes of influenza A appear to have originated from a persistent virus that causes an asymptomatic gut infection in aquatic bird populations and maintains a remarkable level of genetic homogeneity for years (Webster *et al.*, 1993). Another example is that of Hantavirus in humans that can cause fatal acute hemorrhagic fever or pulmonary disease. The virus seems to have made a host jump from rodents in which it establishes an inapparent persistent infection of kidneys and other tissues. The virus is genetically stable in its natural host justifying its persistent nature (Feuer *et al.*, 1999; Villarreal *et al.*, 2000). The above mentioned cases and a few more represent examples of persistent viruses resulting in acute diseases in adapted host species. Such a switch between persistent and acute lifestyles has not yet been clearly exhibited in plant viruses. The relationship between endornaviruses and the ssRNA plus-sense alpha-like viruses suggests that either the dsRNAs evolved from an ancestral ssRNA virus or vice versa (Gibbs *et al.*, 2000). So, either these endornaviruses have evolved from acute, infectious viruses with a loss of CP gene, or the acute alphaviruses have evolved from an ancestral endornavirus, consistent with the idea that emerging acute viruses can evolve from persistent viruses.

Though roles of viruses seem apparent in contributing to their hosts' fitness by creating mutualistic relationships with plants, virus epidemics in wild plant communities should not be underestimated. Viruses may not be overtly damaging in communities of wild plants, but they can decrease the fitness of host plants leading to potential impacts on the wild plant community. (Friess & Maillet, 1996; Malmstrom *et al.*, 2005b; Maskell *et al.*, 1999) Emerging infectious diseases (EIDs) of newly evolved plant viruses pose a constant threat to economically important species, emphasizing the need to develop novel methods for viral identification. According to data in ProMED for 1996 to 2002, plant viruses were the cause of 47% of the reported emerging infectious diseases of plants (Anderson *et al.*, 2004). Introduction of new host plants or increase in vector populations can result in increase in viral disease emergence in plants. With global climate changes, conditions affecting development and distribution of vectors are bound to change (Canto *et al.*, 2009; Garrett *et al.*, 2006). Since temperatures for virus replication show optima, temperature variations will also influence the proliferation of viruses. Global climate change or other factors can also lead to changes in agricultural practices and land use patterns, replacing crops in one location with crops from another region. The crops newly introduced in to a region can possibly be subjected to infection by a variety of new indigenous viruses, or may spread their own viruses to native plants (Webster *et al.*, 2007). In Africa, several crops introduced from other continents suffered from infections of viruses including *Cacao swollen shoot virus*, Cassava mosaic viruses, *Maize streak virus*, and *Sugarcane streak virus*, presumably originating from native plants (Bosque-Perez, 2000; Fargette *et al.*, 2006).

Plant virus outbreaks can greatly affect both yield and quality of agricultural products, leading to significant economic damages and reduced public health. Plant biosecurity has been more stringent than ever worldwide, as the potential employment of plant pathogens as agents of bioterrorism threatens food security (Rodoni, 2009). The role of agriculture in the health and economy of a country coupled with the ease of deploying biological weapons makes agriculture an attractive target for a bioterrorist attack (Casagrande, 2000). Enhancement and ease of

transportation of goods leads to increased chances of cross-border movement of viruses and other microbes. Plant biosecurity efforts aim at preventing the entry of undesirable microbes and viruses along with plant importations (Rodoni, 2009). Prior virus biodiversity surveys in the neighborhood of agricultural fields will allow investigators to decide whether the infection was intentional or natural. Genomic approaches for rapid and accurate detection of plant viruses become extremely important during the time of viral outbreaks. Knowledge about prevailing viruses becomes particularly crucial since the undetected presence of a virus may invalidate most of the tests to confirm that plants are virus-free. Ongoing studies for discovery of viral species in cultivated and non-cultivated plants utilize many genomic approaches for systematic searches for viruses in an unbiased fashion, although the identification of a completely unrelated novel plant virus is always a bigger challenge.

The thesis has been divided into three principal chapters. The first chapter, for the most part, represents a paper "Oligonucleotide-based microarrays for detection of plant viruses employing sequence-independent amplification of targets" that has been published in the *Journal of Virological Methods* in January, 2010. The work demonstrates the utility of array hybridization, its development and application in plant virus detection studies. The main aim of this study was to develop a microarray based on a sequence-independent amplification method with the potential to detect a broad group of plant viruses including detection of new virus species, strains and variants. The method described demonstrates a viable procedure for nucleic acid amplification and hybridization that should be effective in detecting most plant RNA or DNA viruses as long as the virus has representative sequence information available. The second chapter represents a review "Genomic approaches to discovery of viral species diversity of non-cultivated plants" submitted as a co-author with Dr. Ulrich Melcher for the book titled as "Recent Advances in Plant Virology". The Chapter summarizes several genomic approaches employed for plant virus detection studies, discussing their suitability for investigations of viruses in non-cultivated plants. A part of the conclusions from the second chapter is included in the Concluding

remarks section of the thesis. The third chapter represents the work that is in its manuscript draft stage, and is titled as “First characterization of a plant virus belonging to the family *Totiviridae*”. This study primarily focused on two main objectives. The first objective was to characterize the putative totivirus, discovered in the native plants of Oklahoma, as 1) a plant virus or 2) a mycovirus of a fungal endophyte colonizing the plants or 3) a virus that can be transmitted between plants and fungi. The second objective was to determine, if it is a plant virus, whether the virus is a persistent plant virus. In case it is a persistent plant virus, it is possible that the virus might be spreading in plants, perhaps through fungal endophytes, something which has been hypothesized but not observed before.

Research Goals and Objectives

The goals of this research were to:

1. To develop a microarray based on a sequence-independent amplification method with the potential to detect a broad group of plant viruses including detection of new virus species, strains and variants.

   Specific Objectives:

   i) To examine the influence of various factors such as probe length, target length, spacer length and temperature on hybridization specificity and detection sensitivity.

   ii) To develop an extraction and amplification protocol suitable for sequence-independent amplification of total nucleic acids from a variety of plant materials.

   iii) To produce a microarray comprising an appropriate subset of genus and virus-specific oligonucleotides, and validate the efficacy of detection using appropriate and characterized virus isolates and field samples.

2. To characterize the putative totivirus as a fungal or a plant virus or one that can be transmitted between plants and fungi and to test for the persistence of the virus in the plant, if it is a plant virus.

Specific Objectives

i) To screen harvested *R. humilis* samples for the presence or absence of the putative totivirus.

ii) To test for the mechanical transmission of the virus.

iii) To determine the precise localization of the putative totivirus in plant tissue, using *in situ* hybridization, in order to determine host of the virus.

iv) To test for the association of the virus with fungi by performing fungal staining.

v) To isolate fungal endophytes from the infected plants and characterize any dsRNAs (i.e. viruses) that they harbor.

vi) To germinate seeds from the infected and uninfected plant samples, and test them for seed transmission of the virus.

vii) To determine the complete sequence of the putative totivirus genome including the precise 5´- and 3´-ends.

CHAPTER I

OLIGONUCLEOTIDE-BASED MICROARRAYS FOR DETECTION OF PLANT VIRUSES EMPLOYING SEQUENCE-INDEPENDENT AMPLIFICATION OF TARGETS

INTRODUCTION

Development of sensitive and multiplexed detection tools capable of rapidly and economically identifying a broad spectrum of plant viruses is critical in epidemiological and ecological investigations, reacting to agricultural outbreaks and biodefense (Boonham *et al.*, 2003; Webster *et al.*, 2004; Wheelis *et al.*, 2002)**.** Common methods for plant virus detection include variations of the polymerase chain reaction (PCR), serological assays such as enzyme linked immunosorbent assays (ELISA) and immunofluorescent antibody tests (Menzel *et al.*, 2002; Webster *et al.*, 2004), dsRNA and metagenomic approaches (reviewed in Chapter 2). PCR-based techniques have improved tremendously and are preferred often for definitive identification of the causative agent. Most of the molecular techniques, whether protein or nucleic acid based, have limitations, including a requirement for prior knowledge or presumption regarding identities of viruses present in samples and detection restriction to a limited number of candidate viruses. Perhaps more importantly, most of these techniques lack the ability to detect novel viruses. For broad-spectrum identification of plant viruses, there is a need for complementary and

comprehensive multi-targeted approaches for virus detection.

Microarrays, first developed to assay the differential expression of mRNAs in different tissues or developmental stages (Schena *et al.*, 1995), were recognized soon for their potential to identify pathogens. Arrays have been developed for the detection of animal and plant pathogens (Jääskeläinen & Maunula, 2006; Seifarth *et al.*, 2003; Sengupta *et al.*, 2003; Wilson *et al.*, 2002), including a remarkable application of the technique in identification of the severe acute respiratory syndrome (SARS) virus as a member of the genus *Coronavirus* (Wang *et al.*, 2002; Wang *et al.*, 2003). Fewer arrays have been developed for the detection of plant viruses, the earliest of which were for the detection and discrimination between potato virus isolates (Boonham *et al.*, 2003). Early arrays consisted of PCR products amplified from cDNA libraries (Boonham *et al.*, 2003; Lee *et al.*, 2003) and were improved later using high purity artificially synthesized oligonucleotides (Bystricka *et al.*, 2005; Deyong *et al.*, 2005). Oligonucleotide probes of 20-70 nt have been used successfully depending upon the desired level of detection specificity (Bystricka *et al.*, 2005; Deyong *et al.*, 2005; Pasquini *et al.*, 2008).

In this article, 25- to 30-mer probes will be referred to as short oligonucleotide probes and 50- to 70-mer probes as long oligonucleotide probes. Literature data suggest that while long probes provide better detection sensitivity, only short probes allow efficient discrimination between closely related sequences (Chou *et al.*, 2004; Letowski *et al.*, 2004; Urakawa *et al.*, 2003). Arrays with both types of probes targeting several different taxonomic groups of viruses should provide both high sensitivity as well as strong discrimination ability.

Target preparation methods and their resulting lengths influence the stability of duplex formation and hybridization signal intensity (Liu *et al.*, 2007; Peplies *et al.*, 2003; Peytavi *et al.*, 2005; Southern *et al.*, 1999)**.** Secondary structure formation in longer targets can cause a decrease in hybridization efficiency by reducing the binding constant with probes by 10^5 to 10^6-fold, increasing false-negative signals (Lima *et al.*, 1992). To mitigate the effects of target secondary

structure hindrances, determination of an optimum target length and optimized technical conditions are critical to achieve an efficient and discriminating hybridization. A recent study (Liu *et al.*, 2007) examined the effects of target length on hybridization efficiency using different length targets against *Escherichia coli* gene probes. Unfortunately, effects of both target and probe length on hybridization specificity and detection sensitivity in plant virus detection studies have not been investigated.

Inefficient hybridization can result also from low target nucleic acid concentrations. In the case of microarrays for RNA viruses occurring at high concentrations, labeled cDNA targets can be generated by direct (Boonham *et al.*, 2003; Lee *et al.*, 2003) or indirect (Bystricka *et al.*, 2005; Pasquini *et al.*, 2008) incorporation of the label during reverse transcription reactions using random primers, without amplification. However, for viruses present in lower titers, target amplification is needed to increase the probability of virus detection. The use of group- or genus-specific primers (Deyong *et al.*, 2005; Sugiyama *et al.*, 2008) for amplification of viral sequences is not suitable for detection of emerging unknown viruses. In addition, there are many groups of plant viruses for which no effective generic primers are available due to extreme nucleotide sequence variability of genomes. Thus, there is a significant need for the application of sequence-independent amplification methods for detection of plant viruses, especially when prior information about the identity of the virus(es) is not available. A method developed for non-specific amplification of DNA (Bohlander *et al.*, 1992) was modified recently and used in a macroarray system for detection of plant RNA viruses (Agindotan & Perry, 2007).

The present study demonstrates the use of sequence-independent amplification starting from viral nucleic acid (VNA) (Melcher *et al.*, 2008) or total RNA followed by *in vitro* transcription to generate cRNA targets for detection of plant viruses using microarrays. Though the method was validated using either VNA or total RNA as substrates, VNA has a two-fold advantage for detection of encapsidated viruses. First, targets derived from VNA, isolated from

virus-like particles, will contain lower proportions of host-derived nucleic acids reducing the background and improving target specificity and sensitivity of hybridization. Second, VNA, as the substrate for random amplification, targets both DNA and RNA plant viruses. This study describes the validation of an array constituting both short and long oligonucleotide probes using tymoviruses as model pathogens. Tymoviruses were chosen for the study because they are one of the most prominent viral genera present in non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma (Muthukumar *et al.*, 2009) (Min *et al.*, Unpublished data). Initial experiments were performed with *Turnip vein clearing virus* (TVCV, GenBank accession no. U03387) and *Cauliflower mosaic virus* (CaMV, GenBank accession no. M90541) to examine the effects of probe and target length variations on hybridization efficiency.

METHODS AND MATERIALS

PCR amplification and purification of DNA targets

Different length DNA products for TVCV (100, 300 and 1000 bp) and CaMV (92 and307 bp) were amplified from *E. coli* derived TVCV (Zhang *et al.*, 1999) and CaMV (Armour *et al.*, 1983) plasmids using specific primers (Table 1). Total reaction mixtures of 25 µl comprised of 16 µl of nuclease-free water, 2.5 µl of 10X Taq polymerase buffer, 2 µl of 25 mM $MgCl_2$, 0.35 µl of dNTPs/aa-dUTP (a mixture of 10 mM dGTP, dATP, dCTP each, 5 mM dTTP and 5 mM aminoallyl-dUTP), 2.5 U of Taq polymerase (Invitrogen, Carlsbad, CA, USA), 1µl each of 0.4 mM forward and reverse primers and 2 µl of the template plasmid. Cycle parameters for the PCR amplification were as follows: 94 °C for 2 min followed by 30 cycles of 30 s at 94 °C, 40 s at 48 °C and 1 min at 72 °C with a final 10 min extension step at 72 °C. Synthesized PCR products were analyzed using 2% agarose gel electrophoresis in 40 mM Tris acetate, 1mM EDTA. DNA fragments of 307 bp or below were purified using QIAquick Nucleotide Removal Kit (Qiagen, Valencia, CA, USA) while Qiaquick PCR Purification Kit (Qiagen) was used for 1000 bp products. Purified samples were dried and suspended in 8 µl of nuclease-free water. All oligonucleotides including primers and probes used in this study were synthesized commercially (Integrated DNA Technologies, Inc., Coralville, IA, USA and Midland Certified Reagent Co., Midland, TX, USA).

Plant materials, viruses and RNA synthesis

The three *Tymovirus* species used were *Kennedya yellow mosaic virus* (KYMV), *Turnip yellow mosaic virus* (TYMV) and a novel tymovirus (Min *et al.*, Unpublished data), designated in this study as *Asclepias* asymptomatic virus (AsAV). The cDNAs for TYMV and KYMV were provided by Dr. Yannis Tzanetakis, University of Arkansas, USA. Briefly, the purified viral RNA

Table 1: Primer sequences used to amplify cDNA targets

Primer	Sequence 5' to 3'	Product Size (bp)
TVCV100F	CAACCCAGGCGATGG	100
TVCV100R	AACTTTTCCCAGATCTTGTACTCTA	
TVCV300F	CACCAGAAAGACACCTGCGA	300
TVCV300R	GCAATGATGATGGTA	
TVCV1000F	CACCAGAAAGACACCTGCGA	1000
TVCV1000R	CTAGCCACTCTCCGG	
CaMV92F	ATGTCCACAAGGTCACT	92
CaMV92R	GAAATGCTTCGTCCAT	
CaMV307F	CGAGAAGCGAAGAGGAAAGAA	307
CaMV307R	TCTGAACACACGAAATGCTTC	

of KYMV was extracted from virus particles prepared from an infected legume, *Kennedya rubicunda* (Dale & Gibbs, 1976), and total RNA was extracted from TYMV-infected *Brassica pekinensis* (Tzanetakis *et al.*, 2007). The extracted RNAs were reverse transcribed using random hexamers. *Asclepias viridis* infected with AsAV was collected from the Tallgrass Prairie Preserve, Oklahoma. Uninfected *A. viridis* tissue sample was provided by Dr. Richard S. Nelson, Samuel Roberts Noble Foundation, USA.

Virus-like particle isolation and subsequent VNA extraction from infected and uninfected *A. viridis* plant tissue samples were performed as previously described (Melcher *et al.*, 2008). The strategy employed to convert viral RNA or DNA into a form suitable for hybridization to detect viral sequences is shown in Figure 1. The sequence-independent amplification method described previously (Bohlander *et al.*, 1992; Wang *et al.*, 2002) was modified slightly to make amplified targets from mixed populations of single-stranded or double-stranded RNA and DNA. VNA obtained from plant tissues was reverse-transcribed using an SP6 anchor primer with twelve 3'-end random nucleotides (5'-ATTTAGGTGACACTATAGAAN12). The second strand cDNA synthesis was carried out using two rounds of Sequenase (USB, Cleveland, OH, USA), which also can synthesize cDNA from viral DNA genomes. The double-stranded cDNA was then PCR amplified for 30 cycles using the SP6 anchor primer (5'-ATTTAGGTGACACTATAGAA) with Taq polymerase. To incorporate the SP6 promoter on the 5'-ends of TYMV and KYMV cDNAs obtained by reverse-transcription using random hexamers, another round of PCR amplification was performed using SP6-N_{12} random primer and SP6 primer. To prepare targets for labelling, amino-allyl UTP was incorporated into all three PCR amplified cDNA samples by *in vitro* RNA transcription with SP6 RNA polymerase using the MEGAscriptTM high yield transcription kit (Ambion, Austin, TX, USA) at 37 °C for 16 h. The synthesized cRNA mixture was treated with DNase to remove template cDNA, purified using a Mega ClearTM kit (Ambion), dried and suspended in 10 μl of nuclease-free water.

.

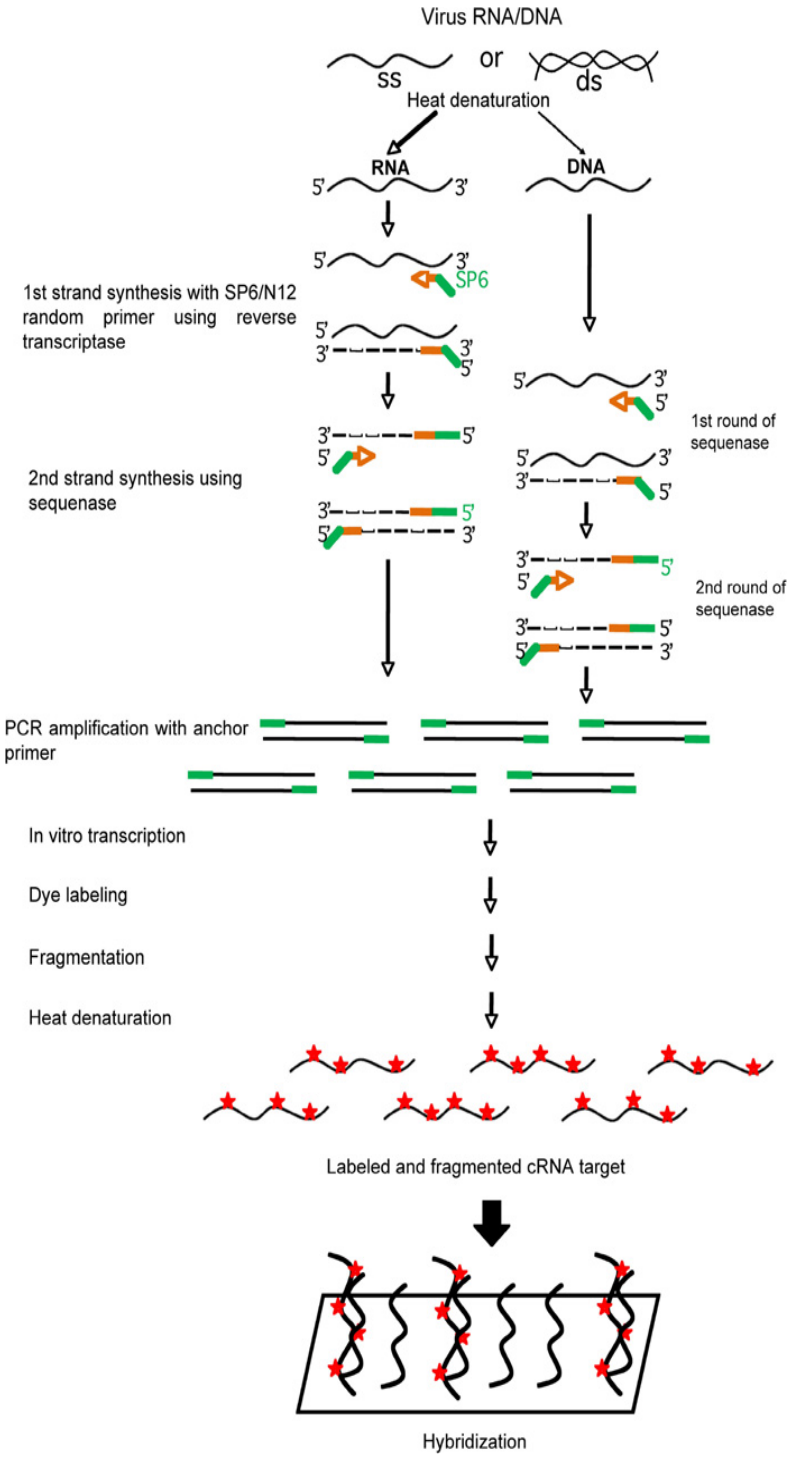

Figure 1. Schematic outline of the strategy used to convert viral RNA and DNA into labeled cRNA for the microarray detection of viral sequences. The outline is abbreviated, with blocking and washing steps not depicted.

Design of oligonucleotide probes and printing

A collection of different oligonucleotide probes ranging from 25-mers to 70-mers were designed for this study (Table 2 and Supplementary Table 1). Conserved regions at a genus or subgenus level were identified from most viral species and used for designing short degenerate probes for members of *Tobamovirus, Caulimovirus, Potexvirus, Marafivirus, Alphacryptovirus* and *Furovirus* genera. Degenerate probes were designed for genera, or if too complex, for subgenera, by aligning sequences and submitting the alignment to Primo Degenerate (Chang Biosciences, Castro Valley, CA, USA). The program finds probe sequences with the least degeneracy to pair with all probe members of the set. In the design, G-T mispairing was allowed and inosine was used in positions requiring all four bases.

Conserved 70-mer probes designed previously (Wang *et al.*, 2002) were used as long probes for TVCV (10003781) and CaMV (10000830). Five probes of different lengths (TV3781-21, TV3781-24, TV3781-27, TV3781-30 and TV3781-50) were designed from within the conserved 70-mer TVCV probe (10003781). TVCV-specific spacer probes were designed with a run of 20 consecutive thymidylates (T_{20}) to provide separation of the hybridization sequence from the substrate. Spacers were located at 3'- (TV3781-30-T_{20}) or 5'- (T_{20}-TV3781-30) ends of the 30-mer probe. Three short 30-mer probes corresponding to *Marafivirus* (Marafi.4636), *Alphacryptovirus* (Acrypto2.66) and *Furovirus* (Furo1.773) were designed with and without a spacer at their 3'-ends. Short probes specific for *Ambrosia* asymptomatic virus 1, AAV1 (Melcher *et al.*, 2008) were designed with no spacer or 5-, 10- and 20-mer thymidylate spacers at their 3'-ends. All the short TVCV-specific probes were plus-sense probes. Plus-sense probes represent the plus-sense viral sequence which will bind to the complementary minus-sense viral RNA of the incoming target sample. Conversely, minus-sense probes will bind to complementary plus-sense viral RNA. Ten plus-sense and ten minus-sense virus-specific short probes with terminal oligo-d(T) spacers were designed for the novel tymovirus, AsAV. In this study, the term “virus-specific” indicates that probe design was based on a specific virus sequence and that its

hybridization will not necessarily discriminate against other closely related species. Hence, cross-hybridization to these probes by targets from related species of viruses was both expected and observed. The ten minus-sense virus-specific short probes for AsAV were designed both with and without spacers for comparison purposes. Five plus- and minus-sense pairs for each of the 50-mer and 70-mer virus-specific probes were designed for each of the three species: AsAV (Min et al., in preparation), KYMV (GenBank accession no. D00637) and TYMV (GenBank accession no. X16378). The complete genome sequences of these species were aligned using Clustal W (Thompson *et al.*, 1994). Regions of high sequence similarity for the three species were identified from alignments and used to design long oligonucleotide probes. The 50-mer probes were designed internal to the 70-mer probes for all three species.

Each oligonucleotide was suspended at a concentration of 20 μM in 3X SSC (Invitrogen, 1X SSC = 150 mM NaCl, 15 mM sodium citrate, pH 7.0). An Omni GridTM DNA microarray printer (Gene Machines, San Carlos, CA, USA) with Stealth SM3B pins (TeleChem International, Inc., Sunnyvale, CA, USA) was used to print arrays on polycarbodiimide-coated slides (Carbo StationTM, Nisshinbo Industries, Inc., Tokyo, Japan). Each probe was printed as four replicates in different areas of the array, to give both adequate replication and location randomization. A Cy3-labeled oligonucleotide was printed on the slides once in each block to provide positional information on the array. The printed oligonucleotide spots had an average diameter of 100-110 μm and 250 μm center to center spacing. The humidity was maintained around 57% during printing.

The printed arrays were subjected to UV irradiation (0.6 J/cm^2) (Kimura *et al.*, 2004) using a UV StratalinkerTM 1800 (Stratagene, La Jolla, CA, USA). The arrays were treated then with a blocking solution (3% bovine serum albumin, 0.1 M Tris-HCl pH 7.5, 0.2 M NaCl and 0.1% sarcosyl) for 30 min, washed in TE buffer (10 mM Tris-HCl pH 7.2, 1 mM EDTA) for 20 min, rinsed briefly with gently flowing nanopure water and dried using a slide centrifuge (TeleChem International, Inc.). All of these steps were performed at room temperature. During

the course of this study, several versions of the array were fabricated incorporating new oligonucleotides at different stages.

Sample labeling, hybridization and image analysis

The purified cDNA samples for TVCV and CaMV as well as *in vitro* transcribed cRNA samples for tymoviruses containing aminoallyl-moieties were coupled to NHS-ester derivatized fluorescent dye. The cDNA or cRNA samples dissolved in nuclease-free water were denatured at 90 °C for 2 min followed by snap cooling on ice. The denatured targets were mixed with 3 µl of 0.1 M sodium bicarbonate and 2 µl (14 nanomoles) of alexa647 dye (Invitrogen) suspended in anhydrous dimethylsulfoxide (EMD Chemicals, Inc., Gibbstown, NJ, USA). While protected from light, the coupling reaction proceeded for 1.5 h at room temperature. The labeled cRNA was purified using the Mega Clear Kit while QIAquick Nucleotide Removal Kit and Qiaquick PCR kit were used for purification of 100-300 bp and 1000 bp labeled cDNA samples, respectively. The fluorescently labeled cRNA was treated with a fragmentation buffer (Ambion) as per manufacturer's instructions to produce shorter RNA products of 60-200 bp. Non-specific target interactions were blocked by addition of 0.08 µg oligo-d(A_{20})/µg target, prior to hybridization against probes with thymidylate spacers. The targets were dried, resuspended in 10 µl of water, denatured at 95 °C for 5 min and snap-cooled on ice for 30 s. After addition of 20 µl of pre-heated Unihyb hybridization buffer (TeleChem International, Inc.), the targets were applied to the slide by flowing underneath a 25 X 40 mm lifter slip (Erie Scientific Company, Portsmouth, NH, USA). The slide was placed in a sealed hybridization cassette plate (Corning Life Sciences, Lowell, MA, USA). The available slots in the hybridization cassettes were filled with 10 µl of 3.5X SSC to maintain humidity during the reaction. DNA targets were hybridized at 42 °C for 16-18 h and cRNA targets were hybridized at 46 °C and 60 °C for 2 h. After hybridization, the arrays were washed sequentially once in 2% SDS, 2X SSC and once in 1X SSC

(Sengupta *et al.*, 2003). The slides were dried and scanned using a Scan Array™ Express scanner (Packard Bioscience, Meriden, CT, USA). Array image acquisition and signal analysis were performed using GenePix Pro 4.0 software (Molecular Devices, Sunnyvale, CA, USA). Data analysis was performed essentially as previously described (Sengupta *et al.*, 2003).

RESULTS

Effect of target and probe length variation on hybridization signal intensity

To test the effects of probe and target length variation on hybridization efficiency, DNA targets of different lengths were hybridized to arrays containing different length oligonucelotide probes specific for the targets (Grover *et al.*, 2007). Hybridizations of three TVCV cDNA targets (100, 300, 1000 bp) and two CaMV cDNA targets (92, 307 bp) were examined against short conserved degenerate 30- and 25-mer probes (Tobamo I-III 4557, Caulimoa.4734) and long virus-specific 70-mer probes (10003781, 1000830) (Table 2). In all experiments, the fluorescence value of an oligonucleotide was required to be at least twenty times above the average background signal to be considered positive. Short degenerate probes did not show detectable signals with any of the target lengths for either of the two species (Fig. 2). On the other hand, the longer 10003781 TVCV probe produced a positive hybridization signal with the TVCV target but not with the CaMV target, while the opposite was true for the CaMV probe 1000830, indicating the expected specificity had occurred. Positive signals were strongest with shorter TVCV and CaMV targets (100 bp and 92 bp, respectively).

To test whether poor hybridization of targets to shorter probes was due to degeneracy present in the short probes, the 100 and 300 bp TVCV targets were hybridized to an array containing six virus-specific probes with lengths ranging from 21 nt (TV3781-21) to 70 nt (10003781) (Table 2). Once again, the longer target (300 bp) produced hybridization signals close to background with all of the six different length probes (Fig. 3). Shorter targets (100 bp) did not produce strong hybridizations with shorter probes but the hybridization efficiency improved approximately five-fold as the probe length increased from 21 nt to 70 nt. To explain the consistent higher efficiency patterns observed with longer targets, relationship between the targets' hybridization efficiencies and overall Gibbs free energies (ΔG), which consider the

Table 2: Sequences of DNA oligonucleotide target-specific probes used in the microarray study

Probe	Target Species/Genus	Sequence 5' - 3'	Probe Length (nt)
10000829	*Cauliflower mosaic virus*	GTCACTACGAATGGAATGTGGTCCCTTTCGGCTTAAAGCAAGCTCCATCCATATTCCAAAGACACATGGA	70
10000830	*Cauliflower mosaic virus*	TCCATGTGTCTTTGGAATATGGATGGAGCTTGCTTTAAGCCGAAAGGGACCACATTCCATTCGTAGTGAC	70
10003781	*Turnip vein clearing virus*	AAATTCTGGAACTCGACATTTCGAAGTACGATAAGTCACAAAACGAGTTCCATTGTGCTGTAGAGTACAA	70
10003782	*Turnip vein clearing virus*	TTGTACTCTACAGCACAATGGAACTCGTTTTGTGACTTATCGTACTTCGAAATGTCGAGTTCCAGAATTT	70
Caulimoa.4734	*Caulimovirus*	TGCCTTTTGGITTAAAGCAAGCGCC	28
TobamoI-III 4557	*Tobamovirus*	CAGAATGAGTTTCATTGTGCWGTIGAGTAT	30
TV3781-21	*Turnip vein clearing virus*	GAGTTCCATTGTGCTGTAGAG	21
TV3781-24	*Turnip vein clearing virus*	GAGTTCCATTGTGCTGTAGAGTAC	24
TV3781-27	*Turnip vein clearing virus*	AACGAGTTCCATTGTGCTGTAGAGTAC	27
TV3781-30	*Turnip vein clearing virus*	CAAAACGAGTTCCATTGTGCTGTAGAGTAC	30
TV3781-50	*Turnip vein clearing virus*	GATAAGTCACAAAACGAGTTCCATTGTGCTGTAGAGTACAAGATCTGGGA	50
(T_{20})-TV3781-30	*Turnip vein clearing virus*	TTTTTTTTTTTTTTTTTTTTCAAAACGAGTTCCATTGTGCTGTAGAGTAC	30
(TV3781-30)-T_{20}	*Turnip vein clearing virus*	CAAAACGAGTTCCATTGTGCTGTAGAGTACTTTTTTTTTTTTTTTTTTTT	30
Acrypto2.66	*Alphacryptovirus*	GACTGCTCTACCTCAACTTTTTACTTACT	29
Acrypto2.66-T_{20}	*Alphacryptovirus*	GACTGCTCTACCTCAACTTTTTACTTACTTTTTTTTTTTTTTTTTTTT	28
Furo1.773	*Furovirus*	CTATCCATAGTATTTATGATATTG	24
Furo1.773-T_{20}	*Furovirus*	CTATCCATAGTATTTATGATATTGTTTTTTTTTTTTTTTTTTTT	24
Marafi.4636	*Marafivirus*	CCTGGAAAGCTTGCCAGACCCTCGCTCTCATGCACGATG	39
Marafi.4636-T_{20}	*Marafivirus*	CCTGGAAAGCTTGCCAGACCCTCGCTCTCATGCACGATGTTTTTTTTTTTTTTTTTTTTT	39
AAV1-T05P	*Ambrosia* asymptomatic virus 1	GGTGAGGGGCCCACCTTCGACGCAAACACCGAGTTTTT	33
AAV1-T10P	*Ambrosia* asymptomatic virus 1	GGTGAGGGGCCCACCTTCGACGCAAACACCGAGTTTTTTTTTT	33
AAV1-T_{20}P	*Ambrosia* asymptomatic virus 1	GGTGAGGGGCCCACCTTCGACGCAAACACCGAGTTTTTTTTTTTTTTTTTTTTTT	34
AAV1-T_{20}M	*Ambrosia* asymptomatic virus 1	CTCGGTGTTTGCGTCGAAGGTGGGCCCCTCACCTTTTTTTTTTTTTTTTTTTTT	34
Tymo.3202P (T)	*Asclepias* associated virus	AACATGAAAAATGGCTTCGATGGATTTTTTTTTTTTTTTTTTTT	44
Tymo.3202M (T)	*Asclepias* associated virus	TCCATCGAAGCCATTTTTCATGTTTTTTTTTTTTTTTTTTTT	42
Tymo.3202M (NT)	*Asclepias* associated virus	TCCATCGAAGCCATTTTTCATG	22
Tymo.5391P (T)	*Asclepias* associated virus	ACTTACGACGACAACACTGACTATAACTTTTTTTTTTTTTTTTTTTT	47
Tymo.5391M (T)	*Asclepias* associated virus	GTTATAGTCAGTGTTGTCGTCGTAAGTTTTTTTTTTTTTTTTTTTT	46
Tymo.5391M (NT)	*Asclepias* associated virus	GTTATAGTCAGTGTTGTCGTCGTAAG	26
Tymo.544P (T)	*Asclepias* associated virus	CATGCACGACGCTCTCATGTATTTTTTTTTTTTTTTTTTT	41
Tymo.544M (T)	*Asclepias* associated virus	AATACATGAGAGCGTCGTGCATGTTTTTTTTTTTTTTTTTTTT	43
Tymo544M (NT)	*Asclepias* associated virus	AATACATGAGAGCGTCGTGCATG	23
Tymo.829P (T)	*Asclepias* associated virus	TCCTGGAATCCTGGGGCCCCCTTTTTTTTTTTTTTTTTTT	41

Tymo.829M (T)	*Asclepias* associated virus	GGGGGCCCCAGGATTCCAGGATTTTTTTTTTTTTTTTTTT	41
Tymo.829M (NT)	*Asclepias* associated virus	GGGGGCCCCAGGATTCCAGGA	21
Tymotp.3295P (T)	*Asclepias* associated virus	AACATGAAAAATGGCTTCGATGGAATTCTCTTTTTTTTTTTTTTTTTTTTTT	51
Tymotp.3295 M (T)	*Asclepias* associated virus	GAGAATTCCATCGAAGCCATTTTTCATGTTTTTTTTTTTTTTTTTTTT	48
Tymotp.3295 M (NT)	*Asclepias* associated virus	GAGAATTCCATCGAAGCCATTTTTCATG	28
Tymotp.4978P (T)	*Asclepias* associated virus	AACGACTATGCTCAGCTCTCCTCCAAAACCTTTTTTTTTTTTTTTTTTTT	51
Tymotp.4978 M (T)	*Asclepias* associated virus	GGTTTTGGAGGAGAGCTGAGCATAGTCGTTTTTTTTTTTTTTTTTTTT	48
Tymotp.4978 M (NT)	*Asclepias* associated virus	GGTTTTGGAGGAGAGCTGAGCATAGTCG	28
Tymotp.5007P (T)	*Asclepias* associated virus	CAAATCCACCATTGTCGCCAATGCTTCCCGTTTTTTTTTTTTTTTTTTTT	51
Tymotp.5007 M (T)	*Asclepias* associated virus	CGGGAAGCATTGGCGACAATGGTGGATTTGTTTTTTTTTTTTTTTTTTTT	50
Tymotp.5007 M (NT)	*Asclepias* associated virus	CGGGAAGCATTGGCGACAATGGTGGATTTG	30
Tymotp.5488P (T)	*Asclepias* associated virus	GGCACTTACGACGACAACACCGACTACAACTTTTTTTTTTTTTTTTTTTT	51
Tymotp.5488 M (T)	*Asclepias* associated virus	GTTGTAGTCGGTGTTGTCGTCGTAAGTGCCTTTTTTTTTTTTTTTTTTT	50
Tymotp.5488 M (NT)	*Asclepias* associated virus	GTTGTAGTCGGTGTTGTCGTCGTAAGTGCC	30
Tymotp.5512P (T)	*Asclepias* associated virus	TACAACATTGCCGTGCTCTACTCTCAATACTTTTTTTTTTTTTTTTTTTT	51
Tymotp.5512 M (T)	*Asclepias* associated virus	GTATTGAGAGTAGAGCACGGCAATGTTGTATTTTTTTTTTTTTTTTTTT	50
Tymotp.5512 M (NT)	*Asclepias* associated virus	GTATTGAGAGTAGAGCACGGCAATGTTGTA	30
Tymotp.5725P (T)	*Asclepias* associated virus	CCTCGCTCTGTTCGCCAAGCTGATGATCGCTTTTTTTTTTTTTTTTTTTT	51
Tymotp.5725 M (T)	*Asclepias* associated virus	GCGATCATCAGCTTGGCGAACAGAGCGAGGTTTTTTTTTTTTTTTTTTTT	50
Tymotp.5725 M (NT)	*Asclepias* associated virus	GCGATCATCAGCTTGGCGAACAGAGCGAGG	30
AsAV1_70P.1005	*Asclepias* associated virus	CCTCTTCACCTACACAAGAGCGGTTCGAACCCTCCGAACCTCCGACCCCGCCGGCTTCGTTCGAACCCAG	70
AsAV1_70M.1074	*Asclepias* associated virus	CTGGGTTCGAACGAAGCCGGCGGGGTCGGAGGTTCGGAGGGTTCGAACCGCTCTTGTGTAGGTGAAGAGG	70
AsAV1_50P.1010	*Asclepias* associated virus	TCACCTACACAAGAGCGGTTCGAACCCTCCGAACCTCCGACCCCGCCGGC	50
AsAV1_50M.1059	*Asclepias* associated virus	GCCGGCGGGGTCGGAGGTTCGGAGGGTTCGAACCGCTCTTGTGTAGGTGA	50
AsAV2_70P.3220	*Asclepias* associated virus	CTTCAATCAGAAACTGAGAGACTCTCGCAATTCATCGACCACTATTGTTGGTGGACGTACAGAGTCCCAT	70
AsAV2_70M.3289	*Asclepias* associated virus	ATGGGACTCTGTACGTCCACCAACAATAGTGGTCGATGAATTGCGAGAGTCTCTCAGTTTCTGATTGAAG	70
AsAV2_50P.3240	*Asclepias* associated virus	ACTCTCGCAATTCATCGACCACTATTGTTGGTGGACGTACAGAGTCCCAT	50
AsAV2_50M.3289	*Asclepias* associated virus	ATGGGACTCTGTACGTCCACCAACAATAGTGGTCGATGAATTGCGAGAGT	50
AsAV3_70P.4399	*Asclepias* associated virus	TCCACCATCGTCGCCAATGCTTCCCGGTCTGACCCAGATTGGAGACATACTGCCGTCAAGATATTCGCCA	70
AsAV3_70M.4468	*Asclepias* associated virus	TGGCGAATATCTTGACGGCAGTATGTCTCCAATCTGGGTCAGACCGGGAAGCATTGGCGACGATGGTGGA	70
AsAV3_50P.4407	*Asclepias* associated virus	CGTCGCCAATGCTTCCCGGTCTGACCCAGATTGGAGACATACTGCCGTCA	50
AsAV3_50M.4456	*Asclepias* associated virus	TGACGGCAGTATGTCTCCAATCTGGGTCAGACCGGGAAGCATTGGCGACG	50
AsAV4_70P.3129	*Asclepias* associated virus	TTCCGACCCTTCCATTTCTCTCATCATCCTCCTTGGAGACCCTCTCCAGGGAGAGTATCATTCCACTTCC	70
AsAV4_70M.3198	*Asclepias* associated virus	GGAAGTGGAATGATACTCTCCCTGGAGAGGGTCTCCAAGGAGGATGATGAGAGAAATGGAAGGGTCGGAA	70
AsAV4_50P.3149	*Asclepias* associated virus	TCATCATCCTCCTTGGAGACCCTCTCCAGGGAGAGTATCATTCCACTTCC	50

AsAV4_50M.3198	*Asclepias* associated virus	GGAAGTGGAATGATACTCTCCCTGGAGAGGGTCTCCAAGGAGGATGATGA	50
AsAV5_70P.4824	*Asclepias* associated virus	CTCCACCCAATTCGGACCCCTCACCTGCATGCGCCTTACTGGAGAGCCCGGCACTTACGACGACAACACT	70
AsAV5_70M.4893	*Asclepias* associated virus	AGTGTTGTCGTCGTAAGTGCCGGGCTCTCCAGTAAGGCGCATGCAGGTGAGGGGTCCGAATTGGGTGGAG	70
AsAV5_50P.4833	*Asclepias* associated virus	ATTCGGACCCCTCACCTGCATGCGCCTTACTGGAGAGCCCGGCACTTACG	50
AsAV5_50M.4882	*Asclepias* associated virus	CGTAAGTGCCGGGCTCTCCAGTAAGGCGCATGCAGGTGAGGGGTCCGAAT	50
KYMV1_70P.1012	*Kennedya yellow mosaic virus*	TCTCTTCACCTACACGCGAGCCGTCAGAACGCTCCGCGTCTCCGACCCCGGCAGGCTTCGTTCGGACCCAA	70
KYMV1_70M.1081	*Kennedya yellow mosaic virus*	TTGGGTCCGAACGAAGCCTGCGGGGTCGGAGACGCGGAGCGTTCTGACGGCTCGCGTGTAGGTGAAGAGA	70
KYMV1_50P.1016	*Kennedya yellow mosaic virus*	TCACCTACACGCGAGCCGTCAGAACGCTCCGCGTCTCCGACCCCGCAGGC	50
KYMV1_50M.1065	*Kennedya yellow mosaic virus*	GCCTGCGGGGTCGGAGACGCGGAGCGTTCTGACGGCTCGCGTGTAGGTGA	50
KYMV2_70P.3425	*Kennedya yellow mosaic virus*	CTGCAGTCCGAGACCACCCGACTTCTCCCCTTCATTGATCACTACTGTTGGTGGACTTATCGTGTCCCCT	70
KYMV2_70M.3494	*Kennedya yellow mosaic virus*	AGGGGACACGATAAGTCCACCAACAGTAGTGATCAATGAAGGGGAGAAGTCGGGTGGTCTCGGACTGCAG	70
KYMV2_50P.3445	*Kennedya yellow mosaic virus*	ACTTCTCCCCTTCATTGATCACTACTGTTGGTGGACTTATCGTGTCCCCT	50
KYMV2_50M.3494	*Kennedya yellow mosaic virus*	AGGGGACACGATAAGTCCACCAACAGTAGTGATCAATGAAGGGGAGAAGT	50
KYMV3_70P.4594	*Kennedya yellow mosaic virus*	AACCCAAGCCACTCTCGTGGCCAACCACTCCCGTTCTGACCCCGACTGGCGCCACACAGCAGTCAAA	67
KYMV3_70M.4660	*Kennedya yellow mosaic virus*	TTTGACTGCTGTGTGGCGCCAGTCGGGGTCAGAACGGGAGTGGTTGGCCACGAGAGTGGCTTGGGTT	67
KYMV3_50P.4609	*Kennedya yellow mosaic virus*	CGTGGCCAACCACTCCCGTTCTGACCCCGACTGGCGCCACACAGCAGTCA	50
KYMV3_50M.4658	*Kennedya yellow mosaic virus*	TGACTGCTGTGTGGCGCCAGTCGGGGTCAGAACGGGAGTGGTTGGCCACG	50
KYMV4_70P.3334	*Kennedya yellow mosaic virus*	GGCAGACCCCTGTCTTGAACTGGTCATCATTCTCGGCGACCCTCTACAAGGCGAGTACCACTCCACTTCC	70
KYMV4_70M.3403	*Kennedya yellow mosaic virus*	GGAAGTGGAGTGGTACTCGCCTTGTAGAGGGTCGCCGAGAATGATGACCAGTTCAAGACAGGGGTCTGCC	70
KYMV4_50P.3354	*Kennedya yellow mosaic virus*	TGGTCATCATTCTCGGCGACCCTCTACAAGGCGAGTACCACTCCACTTCC	50
KYMV4_50M.3403	*Kennedya yellow mosaic virus*	GGAAGTGGAGTGGTACTCGCCTTGTAGAGGGTCGCCGAGAATGATGACCA	50
KYMV5_70P.5026	*Kennedya yellow mosaic virus*	TGCAACGCAGTTCGGCCCTCTGACCTGCATGCGCCTCACTGGCGAACCTGGCACCTACGACGACAACTCA	70
KYMV5_70M.5095	*Kennedya yellow mosaic virus*	TGAGTTGTCGTCGTAGGTGCCAGGTTCGCCAGTGAGGCGCATGCAGGTCAGAGGGCCGAACTGCGTTGCA	70
KYMV5_50P.5035	*Kennedya yellow mosaic virus*	GTTCGGCCCTCTGACCTGCATGCGCCTCACTGGCGAACCTGGCACCTACG	50
KYMV5_50M.5084	*Kennedya yellow mosaic virus*	CGTAGGTGCCAGGTTCGCCAGTGAGGCGCATGCAGGTCAGAGGGCCGAAC	50
TYMV1_70P.1033	*Turnip yellow mosaic virus*	CCTGTTCACCTATACCAGAGCAGTCCGCACACTCCGAACTTCAGACCCAGCAGCATTCGTAAGGATGCAC	70
TYMV1_70M.1102	*Turnip yellow mosaic virus*	GTGCATCCTTACGAATGCTGCTGGGTCTGAAGTTCGGAGTGTGCGGACTGCTCTGGTATAGGTGAACAGG	70
TYMV1_50P.1038	*Turnip yellow mosaic virus*	TCACCTATACCAGAGCAGTCCGCACACTCCGAACTTCAGACCCAGCAGCA	50
TYMV1_50M.1087	*Turnip yellow mosaic virus*	TGCTGCTGGGTCTGAAGTTCGGAGTGTGCGGACTGCTCTGGTATAGGTGA	50
TYMV2_70P.3356	*Turnip yellow mosaic virus*	CTTCCCTCTGAAACTCTCAGGCTGCTACCATACATCGACATGTACTGCTGGTGGAGTTACCGCATTCCTC	70
TYMV2_70M.3425	*Turnip yellow mosaic virus*	GAGGAATGCGGTAACTCCACCAGCAGTACATGTCGATGTATGGTAGCAGCCTGAGAGTTTCAGAGGGAAG	70
TYMV2_50P.3376	*Turnip yellow mosaic virus*	GCTGCTACCATACATCGACATGTACTGCTGGTGGAGTTACCGCATTCCTC	50
TYMV2_50M.3425	*Turnip yellow mosaic virus*	GAGGAATGCGGTAACTCCACCAGCAGTACATGTCGATGTATGGTAGCAGC	50
TYMV3_70P.4520	*Turnip yellow mosaic virus*	TCCACCATAGTGGCCAACGCTTCACGCTCCGACCCAGACTGGCGACACACCACCGTCAAGATCTTCGCGA	70
TYMV3_70M.4589	*Turnip yellow mosaic virus*	TCGCGAAGATCTTGACGGTGGTGTGTCGCCAGTCTGGGTCGGAGCGTGAAGCGTTGGCCACTATGGTGGA	70

TYMV3_50P.4528	*Turnip yellow mosaic virus*	AGTGGCCAACGCTTCACGCTCCGACCCAGACTGGCGACACACCACCGTCA	50
TYMV3_50M.4577	*Turnip yellow mosaic virus*	TGACGGTGGTGTGTCGCCAGTCTGGGTCGGAGCGTGAAGCGTTGGCCACT	50
TYMV4_70P.3265	*Turnip yellow mosaic virus*	CGCCGACCCCGCCCTCGAGCTCGTCATAATTCTCGGCGATCCTCTMCAGGGCGAGTACCACTCCCAATCG	70
TYMV4_70M.3334	*Turnip yellow mosaic virus*	CGATTGGGAGTGGTACTCGCCCTGKAGAGGATCGCCGAGAATTATGACGAGCTCGAGGGCGGGGTCGGCG	70
TYMV4_50P.3285	*Turnip yellow mosaic virus*	TCGTCATAATTCTCGGCGATCCTCTMCAGGGCGAGTACCACTCCCAATCG	50
TYMV4_50M.3334	*Turnip yellow mosaic virus*	CGATTGGGAGTGGTACTCGCCCTGKAGAGGATCGCCGAGAATTATGACGA	50
TYMV5_70P.4948	*Turnip yellow mosaic virus*	CTCCACCCAGTTCGGCCCCCTCACATGCATGCGCCTAACCGGGGAACCCGGAACTTACGACGACAACACT	70
TYMV5_70M.5017	*Turnip yellow mosaic virus*	AGTGTTGTCGTCGTAAGTTCCGGGTTCCCCGGTTAGGCGCATGCATGTGAGGGGGCCGAACTGGGTGGAG	70
TYMV5_50P.4957	*Turnip yellow mosaic virus*	GTTCGGCCCCCTCACATGCATGCGCCTAACCGGGGAACCCGGAACTTACG	50
TYMV5_50M.5006	*Turnip yellow mosaic virus*	CGTAAGTTCCGGGTTCCCCGGTTAGGCGCATGCATGTGAGGGGGCCGAAC	50

70-mer probes for *Turnip vein clearing virus* and *Cauliflower mosaic virus* are probes described by Wang *et al.*, 2003; M-Minus sense probe, P-Plus sense probe, (T)-Tailed/Spacer probe, (NT)-Non-tailed/Non-spacer probe. Probes not indicated as P or M are plus-sense probes.

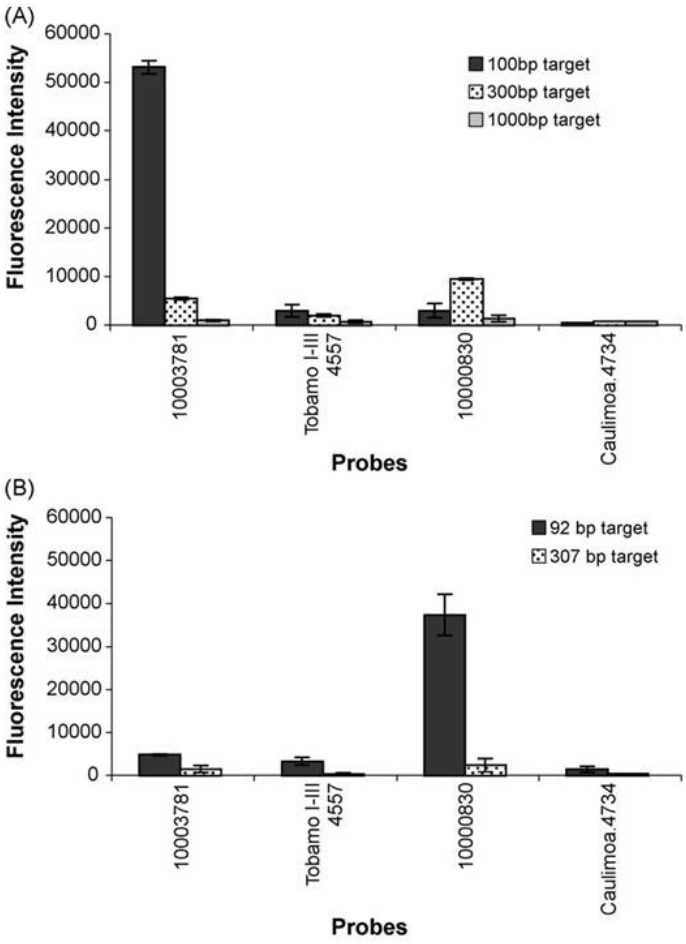

Figure 2. TVCV and CaMV cDNA hybridization fluorescence intensity as a function of target and probe length. Signal patterns observed after microarray hybridization of (A) TVCV cDNA targets of three different lengths and (B) CaMV cDNA targets of two different lengths to a set of short and long TVCV- and CaMV-specific oligonucleotide probes. The long probes (10003781 and 10000830) are TVCV- and CaMV-specific perfect match 70-mer probes. Tobamo I–III 4557 and Caulimoa.4734 are short degenerate probes for TVCV and CaMV with five and two nucleotide mismatches, respectively. Error bars represent the standard deviations for analyzed probe replicates.

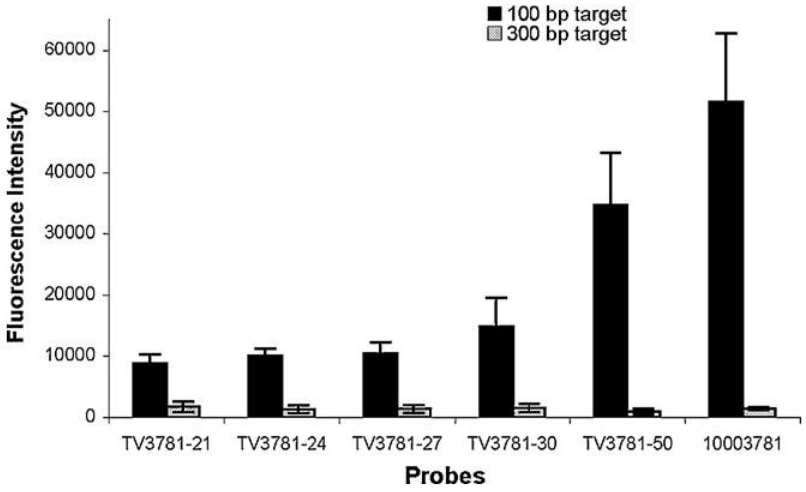

Figure 3. Signal patterns after hybridization of 100 and 300 bp cDNA targets of TVCV to six different length probes ranging from 21- to 70-mer for the same target. All six probes are sequence-specific perfect match probes. The suffix numbers of probes specify their length, for e.g. TV3781-21 is a 21 nt long probe. 10003781 is the long TVCV-specific 70-mer probe. Error bars represent the standard deviations for analyzed probe replicates.

occurrences of secondary structures in the DNA targets, was evaluated. As calculated using the DNA mfold server (Zuker, 2003), the negative values of ΔG for 300 and 1000 bp targets were twice and seven times lower than for the 100 bp target, respectively. Furthermore, both of the longer targets contain an additional sequence capable of forming a hairpin whose loop could pair in pseudoknot fashion with sequence complementary to the probe used. The observed hybridization inefficiency of probes with longer targets may thus be due to formation of secondary structures in longer DNA targets, which is a well documented factor affecting probe binding for both DNA and RNA molecules (Lima *et al.*, 1992; Liu *et al.*, 2007; Peplies *et al.*, 2003; Southern *et al.*, 1999). The superior hybridization of TV3781-50 relative to shorter probes could be due to its substantially higher calculated melting temperature (T_m, 66 °C vs. 53 to 58 °C) or to its extra length. The extra length could circumvent possibly limited accessibility of short surface-bound DNA probes to targets.

Spacer Effect

To test the theory of limited accessibility of shorter probes hindering hybridization efficiency, a spacer molecule was introduced to increase the distance between the DNA probe sequence and the slide surface. To find an optimum spacer length, oligo-d(T) spacers of different lengths (5-mer, 10-mer and 20-mer) were attached on an AAV1-specific probe. The probes were hybridized to the complementary target synthesized after virus purification from *Ambrosia psilostachya* (Melcher *et al.*, 2008). The results showed that 20-mer spacer length produced the strongest and most specific hybridization signals (data not shown). Further experiments were performed using the selected 20-mer oligo-d(T) spacer. TVCV cDNA targets of 100 and 300 bp were hybridized to probes TV3781-30, T_{20}-TV3781-30 and TV3781-30-T_{20}, providing no spacer, a 5'-end T_{20} spacer and a 3'-end T_{20} spacer, respectively. No effect was observed on the hybridization efficiency of the longer target (300 bp) when hybridized to TVCV-specific spacer-probes, whereas the hybridization efficiency of the shorter target (100 bp) improved with spacer-probes

(Fig. 4). The signal intensity of the shorter target with probe TV3781-30-T_{20} was 6.2-fold higher than that of the same probe without the spacer, whereas the signal for probe T_{20}-TV3781-30 increased only 2.5-fold relative to the non-spacer probe, indicating that spacers were optimal when placed on the 3'-end. Although the calculated T_m values for TV3781-30-T_{20} and T_{20}-TV3781-30 are the same, and slightly higher (less than 3 °C) than that of TV3781-30, there was a significant difference among the hybridization efficiencies of these three probes, suggesting that the increased hybridization efficiency was due not to an effect of T_m, but to the increased length of the probe. To ensure that the increase in intensities were not due to non-specific hybridization of targets to the spacer, hybridization intensities of targets to three target-irrelevant probes (Marafi.4636, Acrypto2.66 and Furo1.773) were compared with and without 3'-end spacers. Regardless of the presence or absence of spacers, these probes produced intensities less than 20% of target-specific probes with spacers at either ends. Together the results suggested that the hybridization efficiency of short probes could be improved to produce detectable and specific signals by addition of oligo-d(T) spacers at 3'-ends. These findings were in agreement with similar previous reports using spacers with different slide chemistries (Chou *et al.*, 2004; Peplies *et al.*, 2003; Southern *et al.*, 1999).

Detection of tymoviruses singly and in mixtures

To validate the DNA array with material from plant samples, the array was tested for sequence-specific detection using *tymovirus* species: AsAV, KYMV and TYMV. Labeled and fragmented cRNA targets of pure and mixed samples were hybridized on separate arrays.

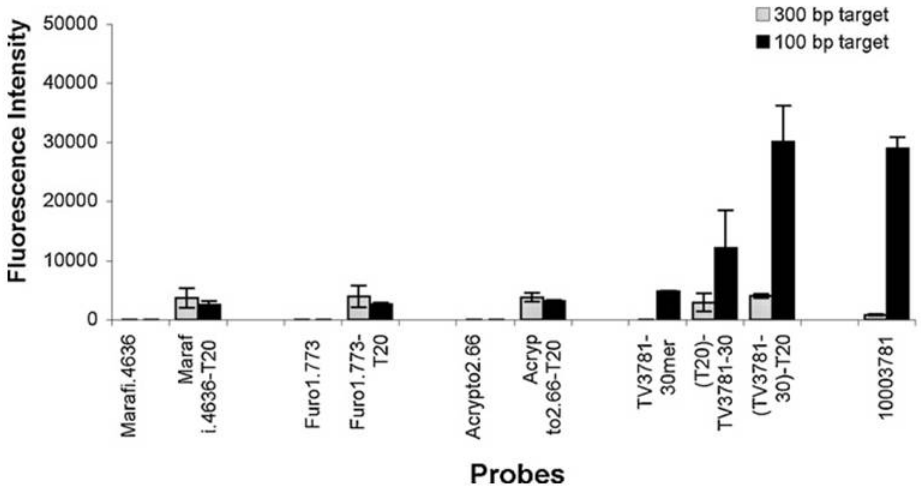

Figure 4. Impact of spacers on hybridization efficiency of shorter probes. Hybridization of 100 and 300 bp TVCV cDNA targets to probes TV3781, T_{20}-TV3781-30 and TV3781-30-T_{20} provided with no spacer, 5'-end T_{20} spacer and 3'-end T_{20} spacer, respectively. 10003781 is the long conserved TVCV-specific 70-mer probe. Hybridization of incoming targets to three target-irrelevant control probes (Marafi.4636, Acrypto2.66 and Furo1.773) with and without 3'-end spacers was also monitored to test for non-specific pairing of targets to spacers.

Hybridization with short oligonucleotide probes

To test the hybridization method using short oligonucleotide probes, cRNA targets derived from an AsAV-infected and uninfected control *A. viridis* were hybridized to an array containing a set of ten AsAV-specific probe pairs along with other unrelated viral probes. The cRNA target from an uninfected plant did not hybridize with any of the viral probes on the array, including 25-mers to 70-mers, validating the design of the array and the hybridization protocol (Fig. 5A). Labeled AsAV target demonstrated highly specific hybridization with short AsAV-specific probes (Fig. 5B). All minus-sense probes hybridized with strong signals to the target while plus-sense probes did not hybridize or hybridized poorly. Hybridization with long oligonucleotide probes (described below) also showed such preferential hybridization to minus-sense probes. The poor hybridization performance of plus-sense probes (discussed below) caused us to focus on the minus-sense probes in what follows.

As in Figure 4, short minus-sense probes without spacers displayed weaker hybridization signals than corresponding probes with spacers (Fig. 6), when hybridized to cRNA targets from infected *A.viridis*, confirming the importance of spacers for short oligomers. A possible disadvantage of using an oligonucleotide spacer is the potential base pairing between the spacer and the target molecule. A 20-mer oligo-d(A) was added to the fragmented cRNA target just prior to hybridization to bind to the complementary oligo-d(T) spacer and prevent any random pairing between targets and spacers. The false positive signals observed in earlier hybridizations were lowered to near background levels, resulting in a decline in non-specific hybridizations without a loss in signals for specific hybridizations (Fig. 7).

Hybridization with long oligonucleotide probes

Long oligonucleotide probes are becoming employed widely in arrays for pathogen detection studies (Agindotan & Perry, 2008; Pasquini *et al.*, 2008; Wang *et al.*, 2003). The study used two types of long probes, 50-mer and 70-mer, to compare their array performance against cRNA

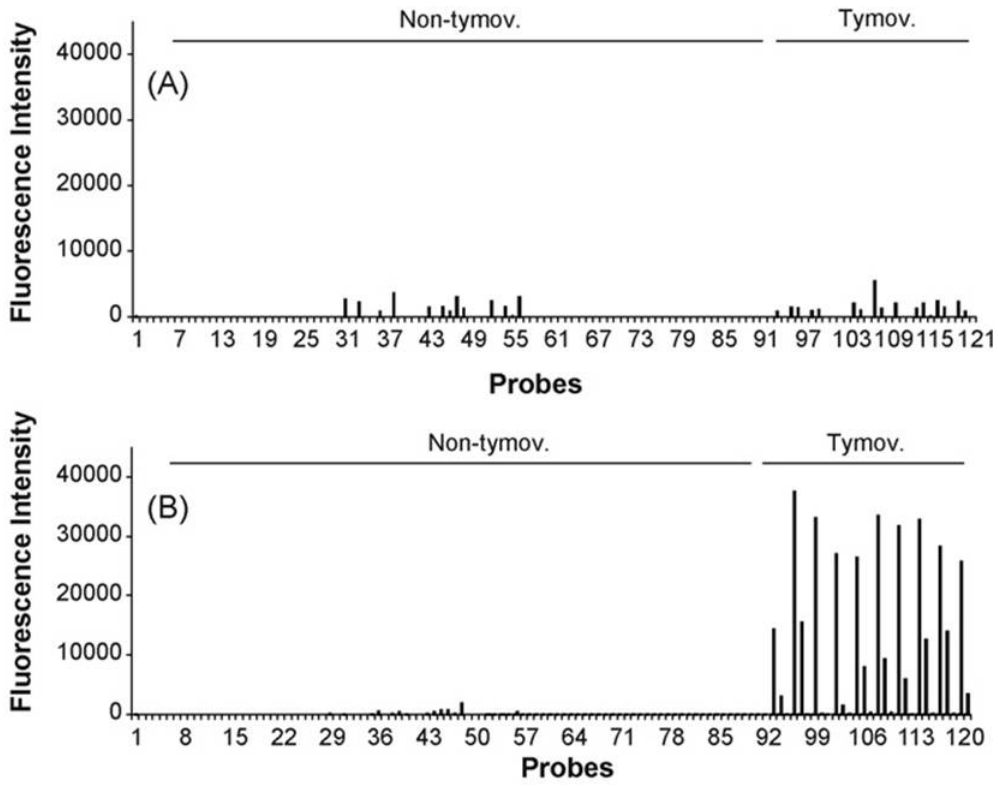

Figure 5. Signal patterns after microarray hybridization of labeled cRNA from (A) uninfected A. *viridis* used as a negative control and (B) AsAV-infected *A. viridis* at 46 °C to a set of short target-specific and non-target probes (Table 2). Probe numbers 1–91 in both A and B are the non-tymoviral probes on the array ranging from 25- to 70-mers, whereas probe numbers 92–121 are the specific tymoviral probes. There are ten tymoviral probe sets in triplets, P(T), M(T), and M(NT) as shown in Table 2. P, M, T and NT stand for positive-sense, minus-sense, tailed/spacer and non-tailed/non-spacer probes, respectively. The results for each triplet are presented in the same order, P(T), M(T), and M(NT), with M(T) probe showing the strongest signal in each set.

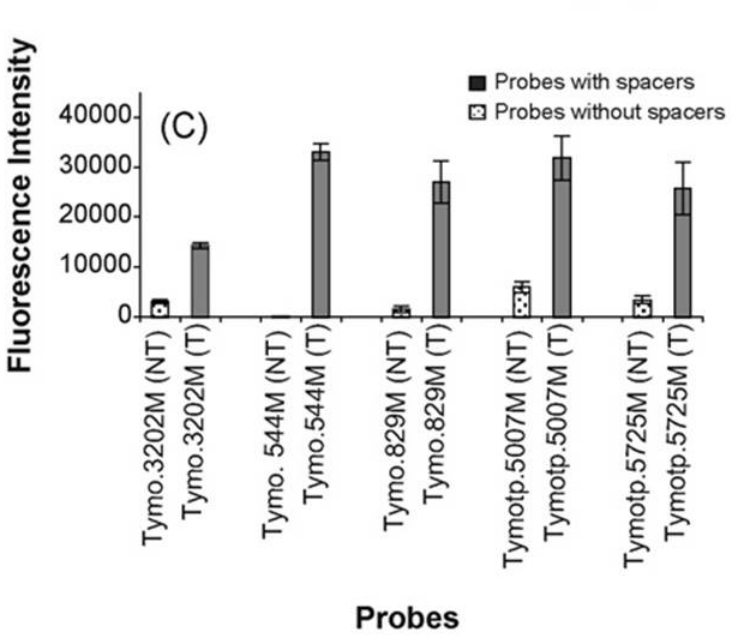

Figure 6: Comparison of hybridization efficiency of targets against probes with and without spacers. The figure displays the signal intensity comparison between the AsAV-specific short probes with and without spacers, when hybridized to labeled cRNA targets from AsAV-infected *A. viridis*. Only five of the ten with vs. without spacer probe comparisons are shown. Error bars represent the standard deviations for analyzed probe replicates.

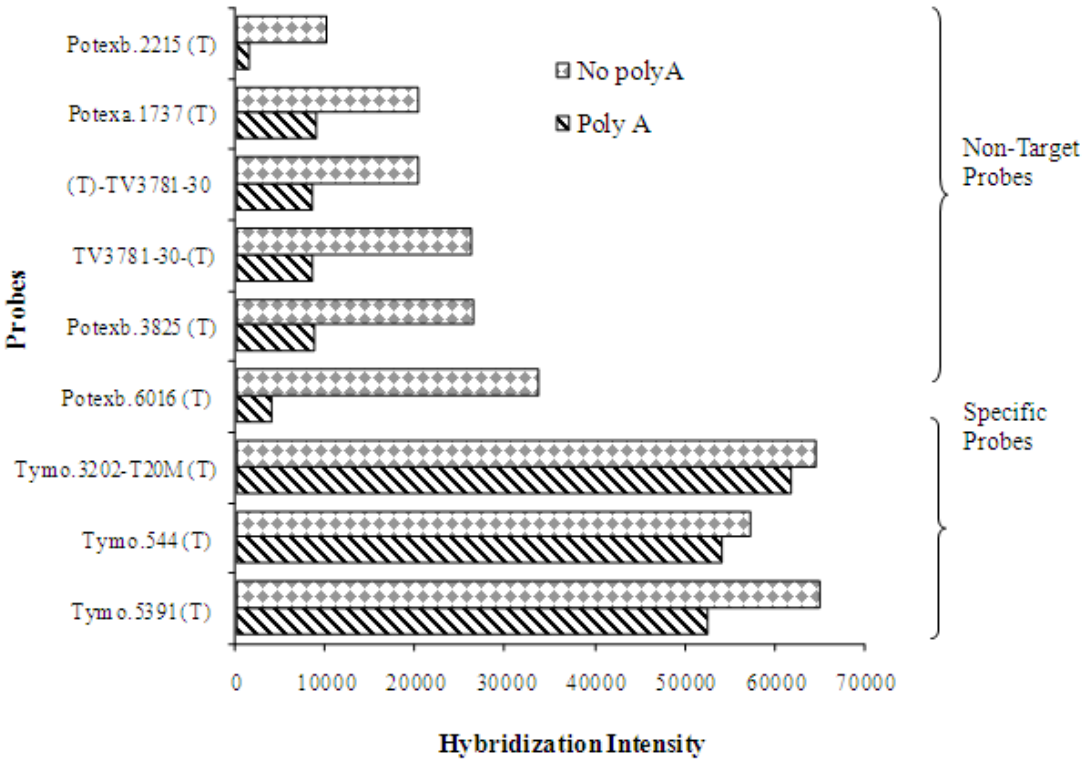

Figure 7: Addition of oligo-d(A) to hybridization mixture helps reduce non-specific hybridization. The figure shows hybridization of labeled AsAV cRNA to some of the specific and non-target short viral probes in presence and absence of oligo-d(A) in the reaction mixture.

targets derived from three *Tymovirus* species (AsAV, KYMV and TYMV) and an uninfected *A. viridis* using a hybridization temperature of 60 °C (Telechem International). Figure 8A shows the compiled results from five individual hybridizations. The cRNA target from an uninfected plant did not hybridize with any of the viral probes on the array. As shown in figure 5B for AsAV target hybridized with short probes, cRNA targets for all three species when hybridized to an array containing longer probes also demonstrated a lack of hybridization to non-tymoviral probes on the array (data not shown). For AsAV and TYMV targets, all specific long probes hybridized strongly to their respective viral targets. In contrast, only three out of five KYMV probe pairs (50-mer and 70-mer) were able to detect the target species. The other two probe-pairs (KYMV2-50M/KYMV2-70M and KYMV4-50M/KYMV4-70M) produced signals below the detection threshold and did not qualify as positives. These two probe pairs were not considered in further analyses. The average ratio of mean median intensities for 70-mer to 50-mer probes was about 1.1 for 13/15 of the probe pairs, indicating the ability of 50-mer probes to produce as strong signals as 70-mer probes under optimal hybridization conditions. As predicted, cross-hybridizations to probes with targets derived from heterologous species were observed, reflecting the successful representation of conserved regions within the genus *Tymovirus* on the array (Fig. 8A). Cross-hybridization signals resulted from probe sequence identities ranging from 60 to 88% and increased approximately linearly with sequence identity values. For example, the AsAV5, KYMV5 and TYMV5 probe group has the highest (greater than 78%) sequence identities of all probe pairs, and produced the strongest cross-hybridizations with viral targets. In general, 50-mer probes with less than 75% overall sequence identity and 70-mer probes with less than 70% overall sequence identity with non-target sequences were virus species-specific under the described hybridization conditions. Cross-hybridizations of targets from heterologous species were more intense with 70-mer probes than with 50-mer probes, which was expected since shorter probes provide greater discrimination between hybridizing nucleic acids. In total, ~34% (9/26) of the heterologous 50-mer probes and ~46% (12/26) of the heterologous 70-mer probes

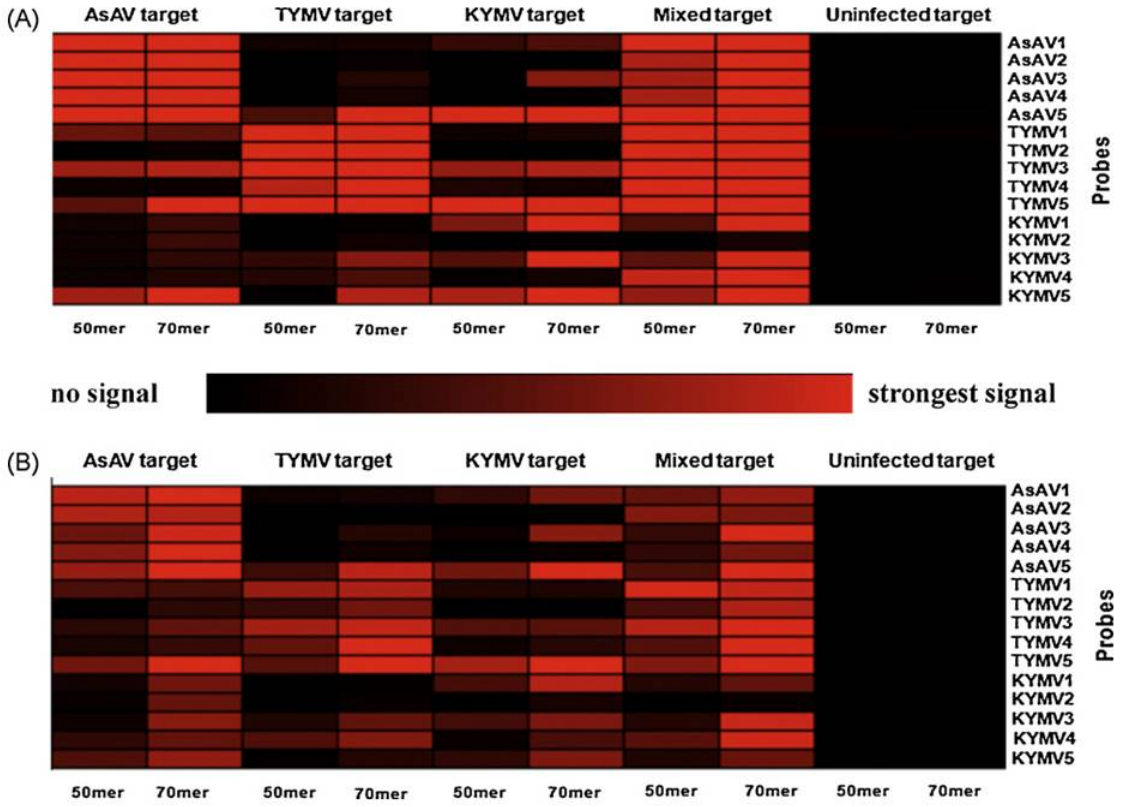

Figure 8. Hybridization results of AsAV-, TYMV- and KYMV-infected samples as single infections or mixture. Uninfected *A. viridis* sample was a negative control target. The figure shows a composite overview of signal patterns in the form of a heat map for five individual hybridization reactions performed at (A) 60 °C and (B) 46 °C. Each column represents the signal intensities of the fifteen 50- or 70-mer species-specific oligonucleotide probes hybridized to the incoming viral targets.

produced a hybridization signal greater than 35% of the strongest signal for that probe. The observed cross-hybridizations did not hinder the identification of individual target species in the respective infected samples since multiple homologous probes hybridized with their targets with stronger signals.

To test for simultaneous detection of multiple viruses in a single sample, cDNAs of AsAV, TYMV and KYMV were mixed prior to *in vitro* transcription and the labeled cRNA mixture was tested on an array (Fig. 8A). The results showed that the presence of multiple viruses did not interfere with the detection of any single virus in the sample. Probe pairs; 1, 2, 3 and 4 of AsAV and TYMV as well as 1 and 3 of KYMV, which achieved high signal intensities with mixed species targets, were virus species-specific in single hybridizations, as cross-species hybridizations were absent or weak (Fig. 8A). The signatures of all three viral species were readily detected by 14 out of 15 probe pairs in the mixture. These results demonstrate that the array approach can reliably detect multiple viruses present in individual plants, and has a potential for screening of viral species in environmental samples.

Influence of temperature on signal intensities of long oligonucleotide probes

Hybridizations of the uninfected control target and all three viral targets were repeated at 46 °C instead of 60 °C to test the effect of temperature on hybridization. The uninfected sample did not hybridize with any viral probe on the array (Fig. 8B). The decrease in hybridization temperature was accompanied by a decrease in signal intensities of target-specific long oligonucleotide probes. The temperature decrease did not result in positive hybridization to KYMV2 and KYMV4 probes, false negatives at 60 °C. However, a variation in sensitivity of hybridization between 50-mers vs. 70-mers was observed at 46 °C. The average ratio of mean median intensities for 70-mer to 50-mer probe rose to 1.6 for 86% (13/15) of the probe pairs when hybridized at 46 °C compared to an average of 1.1 when hybridized at 60°C. Two of the probe pairs AsAV2 50M vs. AsAV2 70M and TYMV1 50M vs. TYMV1 70M, hybridized to their

targets at 46 °C, with almost equally strong signals. Concurrent with a decrease in the hybridization temperature, the percentage of heterologous probes producing hybridization signals greater than 35% of the strongest signal for that probe also increased from 34% (9/26) to 46% (11/26) in the case of 50-mer probes and from 46% (12/26) to 57% (15/26) in the case of 70-mer probes. Thus, comparison of hybridizations performed at two different temperatures showed that hybridization performed at 60 °C produced more sensitive and specific detection signals.

DISCUSSION

One aspect of this work was to investigate and optimize parameters that could influence the hybridization efficiency of oligonucleotide probes using polycarbodiimide slide chemistry for microarray detection of plant viruses. The oligonucleotide probes attached to polycarbodiimide-coated slides are bound most likely via thiamine bases forming covalent bonds in the presence of UV irradiation (Kimura *et al.*, 2004). Thus, it is possible that immobilized DNA containing thymine bases on polycarbodiimide-coated slides may limit oligonucleotide accessibility to the DNA target. This risk is higher presumably for smaller immobilized probes than longer probes. The use of terminal thymidylate spacers produced an improvement in the hybridization efficiency of shorter probes. Although the exact mechanism is not proven, the suggested hypothesis is that the spacers are extending these probe sequences away from the slide surface, making the probes accessible for interaction with the target. Another benefit of oligothymidylate spacers is that the spacer itself decreases the possibility that a thymidine internal to the virus sequence will be used for attachment.

The observation that targets hybridized preferentially to probes of one polarity was highly reproducible. Investigations elsewhere (David Wang, personal communication) have indicated similar observations with double stranded fluorescent targets, whereas tests using single-stranded fluorescent targets of both polarities produced signals with appropriate complementary oligonucleotides. The reason for such extreme strand preference for target-probe hybridization in presence of a double-stranded fluorescent target remains to be elucidated.

Target length is also an important parameter in hybridization studies (Liu *et al.*, 2007; Peplies *et al.*, 2003; Peytavi *et al.*, 2005; Southern *et al.*, 1999). Shorter fragments of around 100 bp target length produced stronger hybridization signals on the array than longer targets for both TVCV and CaMV species. The observations above suggest that stronger signals could be due to secondary structure formation in the longer target strands making the target regions inaccessible

to probes. The present results were in agreement with a recent study using *E. coli* 16S rRNA gene probes which showed enhanced hybridization with PCR amplicons of less than 150 bp and fragmented rRNA between 20-100 nt (Liu *et al.*, 2007). In summary, these results contributed to the establishment of efficient probe design and target synthesis strategy to improve the sensitivity and specificity of virus detection for the microarray format. The method described herein provides a viable procedure for nucleic acid amplification and hybridization that should be effective in detecting most plant RNA or DNA viruses as long as the virus has representative sequence information available. In it, viral nucleic acid concentrations for hybridization are increased by preliminary particle enrichment and by synthesis of *in vitro* transcribed cRNA containing aminoallyl moieties. In previous reports using microarrays to detect plant viruses, labeling of targets produced using random primers was either achieved by incorporation of labeled nucleotides during reverse transcription of the total RNA (Boonham *et al.*, 2003; Lee *et al.*, 2003) or using indirect fluorochrome labeling (Bystricka *et al.*, 2005; Pasquini *et al.*, 2008). These procedures did not include a PCR amplification step that could increase the sensitivity of this technique. Direct incorporation of the fluorophore at the reverse transcription step can result in a lower amount of DNA obtained than by indirect labeling, due to poor incorporation of fluorophore-labeled nucleotides into DNA during polymerization. Combining sequence-independent target amplification and *in vitro* transcription with indirect labeling ensures a highly efficient label incorporation as well as sufficient target yield of the final cRNA product. A fragmentation step was added to decrease the formation of possible secondary structures in labeled cRNA target molecules and increase the diffusion rate of the target molecules.

While long probes are reported to be superior in sensitivity, short oligonucleotide probes are suitable for efficient discrimination between closely related species. Hence, a potential exists for the utilization of longer probes for detection of viruses at higher taxonomic levels like genus or family level, along with shorter probes for discrimination between closely related viral species or strains. The study successfully validated the use of both long and short probes (with spacers)

under the described hybridization method and conditions. Comparison was made also between the two types of long probes (50-mers vs. 70-mers) under two different hybridization temperatures using three *Tymovirus* species. An augmentation in hybridization signals occurred with an increase in hybridization temperature (60 °C). This could be explained by the destabilization of secondary structures within target molecules, increasing their accessibility to probes. These results disagree with an earlier report (Chou *et al.*, 2004) that observed a reduction in hybridization signal intensities at higher hybridization temperatures (50 °C and 63 °C) for both 50-mer and 70-mer probes. One explanation for this discrepancy could be the excellent signal-to-noise ratio provided by the polycarbodiimide slide chemistry (Kimura *et al.*, 2004). The results demonstrate the use of 50-mer oligonucleotide probes as an attractive choice, especially for plant virus detection studies given the inherent nucleotide variability in genomes of most plant viruses. The 50-mer probes can produce an ideal balance between probe sensitivity and specificity making the assay specific enough, but not too specific to overlook closely related viral species.

Because the emphasis of this report is on the description of methods, a broader testing of many viral strains was not undertaken. However, several features of this microarray are particularly promising with regards to its ultimate use as a simple, accurate hybridization method for detection of a broad group of viruses. First, the reproducible absence of false hybridization by targets prepared from uninfected plant samples made the interpretation of results simple and reliable. Second, all three individual species used in this report were readily detected by hybridizations to the appropriate oligonucleotides without any non-specific hybridization to unrelated viral probes. Virus-specific hybridizations produced strong signals for multiple virus-specific probes providing explicit interpretations. Since the probe design for all long oligonucleotide probes was focused on regions conserved among the three species, it was not surprising to observe some cross-hybridizations between heterologous species. Indeed, they demonstrated the ability of the array to detect and differentiate between closely related uncharacterized plant viruses. Third, signature sequences of all three viral species were detected

readily in the mixed viral target validating the feasibility of our microarray for simultaneous detection of multiple viruses in a single plant sample.

Although most of the oligonucleotide probes performed as predicted, some probes worked better than others. Since it has already been reported that oligonucleotide probes binding to different regions of a genome yield different signal intensities (Li & Stormo, 2001; Lockhart *et al.*, 1996) the ability of an oligonucleotide probe to yield a good hybridization signal is unpredictable just on the basis of sequence information alone. Thus, multiple probes per species should be used in oligonucleotide array designs to obtain reliable information because seldom do they all prove effective (Agindotan & Perry, 2008).

In conclusion, the report illustrates a significant step forward in plant virus diagnostics by detailing for the first time, a microarray method with the potential to detect a broad group of plant viruses. Such a hybridization approach can facilitate the development of a powerful multi-viral detection system of considerably expansive application for identification of both known and related uncharacterized emerging viruses.

CHAPTER II

GENOMIC APPROACHES TO DISCOVERY OF VIRAL SPECIES DIVERSITY OF NON-CULTIVATED PLANTS

INTRODUCTION

The study of virus infections in non-cultivated ('wild') plants to explore the ecological roles, diversity, and distribution of these plant-associated viruses in unmanaged ecosystems is a long-emerging field in science, bringing together the disciplines of ecology, vector biology, epidemiology and virology. In 1981, Harrison noted that the kinds of viruses associated with cultivated plants (CULPAD, cultivated-adapted) were distinct from those that tended to be associated with plants in natural habitats (WILPAD, wild-adapted) (Harrison, 1981). Unfortunately, the plant viruses known to Harrison were overwhelmingly those that caused diseases in crop plants, with but a few that caused diseases in other plants. There was very little knowledge of viruses that did not call our attention to them by causing obvious symptoms in plants of economic interest to us. This is still largely true today. Examination of the initial plant source of viruses catalogued in the Viral Identification Data Exchange database (Brunt *et al.*, 1996) shows that most known viruses are from cultivated crop species and symptomatic hosts (Wren *et al.*, 2006).

The study of plant viruses was initiated in plant pathology (Scholthof *et al.*, 1999), a field

in which interest is in what makes a plant diseased and how that disease can be prevented. The view of plant viruses as pathogens still dominates plant virology despite the realization that disease is not a necessary consequence of virus infection. Under some circumstances, virus infection of plants actually contributes to improving their fitness. Beneficial effects of viruses on plants often only become apparent under particular ecological circumstances. An indirect example is the Curvularia thermal tolerance virus whose presence in a fungus in *Curvularia* plants increases the thermal tolerance of the plant–fungal complex, allowing it to grow closer to the edges of hot pools in Yellowstone National Park (Márquez *et al.*, 2007; Rodriguez *et al.*, 2004). The observation of increased drought and cold tolerance of virus-infected plants (Xu *et al.*, 2008) also indicates that viruses can make mutualist contributions to plant fitness under some circumstances. Infection of *Kennedya rubicunda* plants by *Kennedya yellow mosaic virus* makes the plants less attractive to herbivores, thus increasing the plant's longevity and fitness (Gibbs, 1980). Virus infection may negatively affect the fitness of one plant species more than it will that of another species (Malmstrom *et al.*, 2005a; Malmstrom *et al.*, 2005b). As a result, virus infection of the second species provides it a competitive advantage when it is in mixed species populations. The relative success of ecotypes of *Arabidopsis thaliana* infected with *Cucumber mosaic virus* (CMV) is influenced by the density of plants (Pagan *et al.*, 2008). Infection of some *A. thaliana* ecotypes by CMV accelerates their production of seed (Pagan *et al.*, 2008). White clover plants in soil infested with fungus gnats produce more biomass and more ramets when infected with *White clover mosaic virus* than when virus-free, due to the viruses' induction of repellent volatile emissions by the plant (van Mölken, 2009). Further mutualistic interactions between plants and viruses await identification and description. The exploration of such interactions and the mechanisms involved in possible adaptation of plants to new environments is changing the world's view of viruses simply as pathogens.

The practical importance of discovering viruses in native plants is well illustrated by agricultural history in Africa, where numerous crops introduced from other continents have fallen

victim to viruses (*Cacao swollen shoot virus*, *Maize streak virus*, *Sugarcane streak virus*, viruses of cassava mosaic) presumably originating from native plants (Bosque-Perez, 2000; Fargette *et al.*, 2006). A phylogenetic study of potyviruses of Australia suggested that they emerged by transfer of these viruses from native ecosystems into cultivated crops when cultivation began in human history (Gibbs *et al.*, 2008a; Gibbs *et al.*, 2008b). Another practical driver of explorations of virus biodiversity occurs when native plants are considered for promotion as alternative food or forage crops. Such situations result in screening the plants for each of multiple viruses (Odedara *et al.*, 2008; Odedara *et al.*, 2007). Non-cultivated plant species growing near crop lands (often called weeds) are often surveyed for the presence of viruses that may affect crop plants (Roye *et al.*, 1997; Sampangi *et al.*, 2007; Sivalingam & Varma, 2007).

One view of the origin of viruses (Jaspars, 1999) suggests that viruses have had a long association with their hosts, consistent with vertical transmission of viruses. Some viruses have been suggested to have co-diverged with wild plants long before domestication of plants (Gibbs *et al.*, 1999; Lartey *et al.*, 1996). Comparative analysis of viral genes has led to the concept that viral genomes consist of certain hallmark genes that were present in the earliest stages of life and numerous accessory genes acquired from a variety of organisms in more recent times (Koonin *et al.*, 2006). Since viruses are obligatorily dependent on hosts for replication, and since they have survived eons of evolution, the net effect of viruses on ecosystems must, in general, be positive. Therefore, it is unrealistic to expect that viruses always cause disease. The supposed long history of plant-virus association also requires that under most circumstances viruses are in a kind of equilibrium with their hosts (Thresh, 1982). The viruses do not strongly negatively affect the plants, nor do the plants strongly negatively affect virus replication. The equilibrium is often disrupted by introductions of vectors or of other viruses or plants in naive areas, resulting in outbreaks (Webster *et al.*, 2007). Comparison of virus phylogenies to plant phylogenies to understand evolution and the influence of ecosystem properties on the distribution and evolution of plant viruses is hampered by incomplete knowledge of existing viruses, as well as the inability

to find fossil viruses.

Studies, primarily of marine sediments, have shown a great diversity of viruses at individual localities, and evidence suggests considerable transfer of these agents from one environment to another (Breitbart & Rohwer, 2005b; Djikeng *et al.*, 2009; Hubert *et al.*, 2009; Short & Suttle, 2005; Suttle, 2007). Viruses of crops show both worldwide distributions for some known pathogenic viruses, through continentally limited distributions, to highly local distributions. Distributions of viruses not associated with crop disease are uncharacterized. Does each locality have its unique catalogue of viruses associated with plants, implying a very large diversity? Or is diversity more limited, with viruses being naturally globally distributed?

Although dwarfed by the number of studies on viruses in crop species, some knowledge of viruses of non-cultivated plants has accumulated. There have been limited surveys for viruses in plant populations from non-managed ecosystems (Bodaghi *et al.*, 2004; Fraile *et al.*, 1997; Raybould *et al.*, 1999). These surveys studied the distribution of known viruses of crop plants in non-cultivated plants, including orchids (Kawakami *et al.*, 2007), using specific assays (serological or RT-PCR) to screen the plants for those viruses. Incidence rates of infection of single plant species with single viral species cover a wide range, but frequently were between 30% and 70%. Previously unknown viruses have been discovered in non-cultivated plants because those plants exhibited novel symptoms (Ciuffo *et al.*, 2008; Gibbs, 1980; Hassan *et al.*, 2009; Ooi *et al.*, 1997; Robertson, 2005, 2007). To investigate the diversity of viruses in extreme environments, plants growing near the Antarctic circle have been examined for the presence of viruses (Polischuk *et al.*, 2007; Skotnicki *et al.*, 2003). Recent inventories of non-cultivated plant viral biodiversity sampled plants without regard to symptoms (Melcher *et al.*, 2008; Muthukumar *et al.*, 2009; Roossinck *et al.*, 2010). The Plant Virus Biodiversity and Ecology (PVBE) project (Wren *et al.*, 2006) developed several hypotheses about such viruses. First, viruses are frequent in non-cultivated plants. Between a quarter and a half of all specimens and all species tested in PVBE project had evidence of virus presence. Second, previously unknown viruses are abundant.

Of several hundred potential viral species detected, only 17 represented viruses whose sequences were in the general GenBank/DDBJ/EMBL database. Third, as originally suggested (Harrison, 1981), the taxonomic distribution of the prominent viruses of non-cultivated plants differed from that of crop viruses. Virus taxa, such as *Potyviridae* and *Geminiviridae*, abundant among those causing crop disease, were relatively absent in the PVBE samples. Fourth, viruses of non-cultivated plants often have features not found in their crop relatives. For example, several PVBE putative viruses had additional open reading frames (ORFs) overlapping standard ORFs. Last, viruses naturally exhibit a limited number of types of distributions among plants and plant species. The project undertaken at the Area de Conservación Guanacaste (ACG) in Costa Rica (Roossinck *et al.*, 2010) provides support for many of these hypotheses, except that one particular crop virus was found with high frequency, *Zucchini yellow mosaic virus*, a member of the *Potyviridae.*

This chapter summarizes genomic approaches to the detection or discovery of viruses in plants and of their diversity, and discusses the suitability of the approaches for investigation of viruses in non-cultivated plants. The past three decades have produced genome sequences for most of the known plant viruses. The availability of these gene sequences has made it possible to design genomic approaches to identify viruses related to known taxa. Genomic approaches survey the sequences derived from genomes present in a sample for sequences of interest and include genomics and metagenomics. Genomics refers to the study of genes in a single organism, while metagenomics refers to study of genes present in an environmental sample. Genomic approaches use various combinations of methods for sampling the environment, enriching samples for content of viral genomes, amplifying nucleic acids, and detecting virus-related sequences among the amplified nucleic acid. These methods include, particularly, array hybridization to macroarrays and microarrays, and various megasequencing approaches.

Sampling

As one moves from genomics to metagenomics, there are numerous steps that can serve as a focus for investigation. These levels include individual plant organs, whole plants, assemblages of fresh plants, and the collection of viruses released into an ecosystem.

Individual plants

The study of multiple single plants is exemplified by the PVBE project (Wren *et al.*, 2006) carried out in Oklahoma and modeled after the similar ACG project conducted in Costa Rica (Roossinck *et al.*, 2010). In this approach, care was taken to record the location and species of plants from which individual samples were taken. In this way, it was possible to tie putative viruses to host species. In this approach, usually, samples of young leaves are taken, assuming that viruses tend to accumulate best in younger leaves. A comparison of organs of *Euphorbia marginata* plants revealed random patterns of recovery of a tymovirus among different plants, suggesting that the choice of organ may not be very crucial (Hackett *et al.*, 2009). Individual plants are also the targets of investigations that focus on plants with apparent symptoms of disease or on assay for specific viruses (Ciuffo *et al.*, 2008; Gibbs, 1980; Hassan *et al.*, 2009; Ooi *et al.*, 1997; Robertson, 2005, 2007).

FTA cards

A recent development that promises to facilitate broad surveys of individual non-cultivated plants for virus presence is the ability to recover usable nucleic acid from squashes of plant material on Flinders Technology Associates (FTA) cards (Alabi *et al.*, 2008; Ndunguru *et al.*, 2005; Owor *et al.*, 2007). These cards, designed for storage of nucleic acids, either in their purified form or

within pressed tissue samples, can allow affixing of the tissue sample in the field itself. It has been reported that DNA can be preserved within a pressed plant tissue for more than 14 years and preservation of plant viral RNA has also been demonstrated (Owor *et al.*, 2007).

Lawnmower

At a slightly higher level, collections of plant material from a specified area can be made and processed for analysis of viral content. Unpublished results (Scheets *et al.*, Unpublished data) illustrate this approach, nicknamed the 'lawnmower' approach. All above-ground parts of plants growing in a 30 × 30 cm area were combined, mixed, and aliquots of the mixture were processed for identification of putative viruses. Eight plant species were identified as contributing to the mixture. All three aliquots revealed the presence of a carmovirus, although in different proportions from the three preparations. From this information alone, the virus could not be assigned to a host species. However, analysis of individual plants from other nearby locations revealed a nearly identical virus in *Lespedeza procumbens,* which was one of the species in the mixture. The additional sequences allowed assembly of a complete genome.

Run-off water

At the highest level of metagenomics yet attempted, viruses are recovered from water. Although some of these viruses, like the marine bacteriophages, are likely viruses of marine organisms, there are reports of non-bacteriophage viruses in water. One study examined an Antarctic lake (Lopez-Bueno *et al.*, 2009), and found evidence of phycodnaviruses and mimiviruses, among others. Eukaryotic probable hosts of these viruses inhabit the lake. Another study (Djikeng *et al.*, 2009) focused on a lake in Maryland, and identified sequences belonging to the plant-infecting taxa *Partitiviridae, Bromoviridae, Luteoviridae, Flexiviridae, Tetraviridae, Tymoviridae, Sequiviridae* and *Tobamovirus*. Their presence in the lake sample can be interpreted as run-off from plant material, particularly since the levels of these sequences were much higher in autumn

than in spring. However, it is not known which plant species contributed which virus.

Enrichment

Regardless of the methods used for detection of viral sequences, probabilities of detection are increased by enriching starting material for viral nucleic acids. As demonstrated by the metagenomic studies, viral sequence may be present at low titres. Thus enrichment for viral sequence before initiating sequence analysis is advisable. A variety of techniques can be and have been employed. Virus purification from aqueous sources involves membrane filtration and ultracentrifugation (Djikeng *et al.*, 2009). In vertebrate virology, virus particle concentration by differential and isopycnic centrifugation as a prelude to large-scale sequencing has proven effective in identifying viruses containing single- and double-stranded DNA, and has led to the identification of a novel anellovirus in healthy donors (Breitbart & Rohwer, 2005a). Enrichment is often achieved simply by using a body fluid, such as serum or plasma, poor in cells and thus having fewer cellular nucleic acids to interfere with detection. Supernatants of cell cultures also serve as good starting materials for vertebrate viruses (Djikeng *et al.*, 2008). The equivalent in plant virology, collection of phloem or xylem, has not been used extensively (Simon-Buela & Garcia-Arenal, 1999; Waigmann *et al.*, 2004), but may be necessary for viruses that have their highest titer in phloem or xylem. A further underexplored resource consists of phloem or xylem-feeding insects. They may serve as a concentration mechanism for viruses in vascular streams. Better documented approaches involve extraction of nucleic acids from purifications of virus-like particles by differential or isopycnic centrifugation, purification of double-stranded RNA (dsRNA), immunocapture, group-specific PCR or RT-PCR, isolation of siRNA or cDNA, and representational difference analysis (RDA).

dsRNA

The double-stranded RNA approach is used frequently in identifying causes of apparent viral disease in crops and ornamental plants. For example, the size of dsRNA purified from a grapevine with leafroll symptoms not attributable to any recognized leafroll-associated virus suggested the presence of a member of the *Closteroviridae*, a suggestion subsequently confirmed using family-specific RT-PCR primers for the HSP70 homologue of this family (Alkowni *et al.*, 2004). Because plants do not have appreciable amounts of dsRNA, and dsRNA is almost an obligatory component of plant cells hosting replication of an RNA virus, this method (Roossinck *et al.*, 2010), enriches for dsRNA by binding to CF cellulose. The dsRNA is then converted to dsDNA and amplified using primers with random oligonucleotides at their 3′-ends. Crucial to this procedure is removal of DNA before reverse transcription, ssRNA, and random-terminated primers after dsDNA synthesis. As expected, the procedure is successful in obtaining sequences from putative viruses from groups not known to have capsids. Also as expected, caulimoviruses and geminiviruses are absent.

VLP-VNA

In the virus-like particle (VLP) approach, investigators assume that viral particles are present, and enrich for them by differential centrifugation of plant homogenates (Lane, 1986, 1992). After DNase I treatment to remove external DNA, nucleic acids are extracted by proteinase K-SDS treatment and phenol:chloroform extraction prior to alcohol precipitation with a coprecipitant to yield putative viral nucleic acid (VNA). The extraction procedure was adapted from one designed for *Cauliflower mosaic virus*, considered one of the most stable of virus particles (Hull, 1978). A similar procedure has been developed for clinical samples (Djikeng *et al.*, 2008). The VLP-VNA procedure should enrich RNA-containing plant viruses from a variety of genera as well as DNA-containing viruses. As applied to PVBE samples, plant-infecting taxa detected were *Badnavirus, Flexiviridae, Tymovirus, Comoviridae, Chrysovirus, Luteoviridae* and *Tobamovirus* (Melcher *et*

al., 2008; Muthukumar *et al.*, 2009). With viruses present in high titre, 100% of retrieved sequences were virus-derived in some samples. In other cases, contaminating sequences often included those derived from other symbionts (bacteria and fungi). Surprisingly, despite the large number of samples containing bacterial species, only few bacteriophage sequences were retrieved. Remaining sequences were likely plant sequence, though viruses completely unrelated to sequences of known ones were not ruled out. The unknowns should become less of a problem as more and more plants have their genome sequences determined. Even though the new plant sequences may be those of crop and ornamental species, they will likely be close relatives of non-cultivated plants of interest. Relative to the VLP-VNA method, the dsRNA method (Roossinck *et al.*, 2010) was more successful. This was mostly because amplification of VLP-VNA samples was more sensitive to inhibition of the PCR.

Virus adsorbents

Viruses can also be purified from complex mixtures, by binding specifically to a solid support, such as special plastics (Rowhani *et al.*, 1998). Antibodies can provide specificity. However, for most applications, the specificity of available antibody preparations is too narrow to guarantee discovery of all unknown viruses in samples.

Group-specific primers

For numerous known taxa of plant viruses, genomes of sufficient species have had their sequences determined for the design of universal primers to be attempted. These primers are generally highly degenerate and extremely useful when targeting a particular group of virus genomes in a single or multiplex PCR reaction (Alkowni *et al.*, 2004; Donehower *et al.*, 1990 ; Foissac *et al.*, 2005; Ha *et al.*, 2008; Paximadis *et al.*, 1999; Saldarelli *et al.*, 1998; Teycheneya *et al.*, 2007). For example, RT-PCR screening of several plant viral genera including *Closterovirus*, *Vitivirus*, *Trichovirus*, *Foveavirus* and *Capillovirus* was performed using degenerate primers

followed by cloning and sequencing the amplified products (Foissac *et al.*, 2005; Saldarelli *et al.*, 1998). Polyvalent detection RT-PCR tests were exploited as one of the methods in surveying the plant virus diversity of the sub-Antarctic Kerguellen Island (Marais *et al.*, Unpublished data). Results showed that one of the *Tropaeolum majus* plant samples identified by RT-PCR as being infected by a virus belonging to the genus *Nepovirus* was confirmed positive by sequencing. Degenerate family- or genus-specific amplified products when used as targets against a microarray consisting of a comprehensive set of probes can also accelerate the discovery of novel viruses.

Although simultaneous detection of up to eight viruses by multiplex PCR has been reported (Sanchez-Navarro *et al.*, 2005), the number of targets that can be effectively detected simultaneously is limited. Increasing the number of primer pairs in the reaction mixtures increases the chances of unexpected interactions. Further, smaller products tend to be preferentially amplified over larger fragments, thus causing a bias in identification of viral targets. Amplified fragments of the approximate expected size are usually further analyzed to detect mispriming and otherwise confirm the identity of the product. A frequently used way to confirm and achieve the identification of the specific amplification product to the species level is sequencing.

Degenerate primers should be designed so as to attain maximum coverage of variant sequences, keeping in mind that degeneracy can drastically reduce the concentration of any single defined sequence in the primer mixture. It must also be kept in mind that a negative result does not mean that viruses of the targeted taxon are absent, since the primer design may have not been able to include them. A cautionary example is provided by the case of banana mild mosaic virus and banana virus X, where degenerate primers worked for the former despite its high molecular diversity, but not for the latter which is otherwise highly conserved (Teycheneya *et al.*, 2007).

siRNA

The discovery of RNA silencing has roots in the study of the 'recovery' phenomenon of plants, in which induced siRNAs interfered with virus production (Ding & Voinnet, 2007). As a logical extension of that discovery, several laboratories are using small RNA deep sequencing to discover virus-like sequences. The strength of the megasequencing approach is illustrated by the range of results obtained after laboratory infection of appropriate host plants with nine viruses (Donaire *et al.*, 2009). Tombusvirus infections resulted in over half of the total siRNAs of *Cucumis melo* and *Nicotiana benthamiana* plants being virus derived. In no virus-host combination were the virus-derived sequences less than 1% of the total. Although some genomic regions were overrepresented, especially in the infection of *A. thaliana* with a crucifer-infecting tobamovirus (Qi *et al.*, 2009), all regions participated in siRNA generation. The larger number of reads available through megasequencing, compared with cloning and sequencing, makes possible complete or near complete genome coverage (Donaire *et al.*, 2009; Kreuze *et al.*, 2009). The small lengths of the siRNAs (21–27 nt) do make the risk of misassembly a real concern when multiple infections are possible. Such infections are expected when examining wild plants. Therefore, there must be good coverage throughout the genome sequence and assembly methods must be carefully chosen (Kreuze *et al.*, 2009). Nevertheless, contigs of the order of kbp were assembled readily, and evidence of the presence of low levels of two unexpected viruses (a badnavirus and a mastrevirus) was also obtained (Kreuze *et al.*, 2009). At this point, it is not clear whether the low level siRNAs are signs of an on-going infection or are remnants of a successful silencing attempt by the plant. In either case, the procedure can lead to identification of possibly novel viruses. A bias of plant enzymes to processing small RNAs at GC-rich regions (Donaire *et al.*, 2009; Ho *et al.*, 2008) may interfere with detection of GC-poor viruses and prevent recovery of complete sequences.

RDA

Subtractive hybridization coupled with PCR, in a method called representational difference analysis (RDA), has been used to enrich viral sequences in samples without prior knowledge of which sequences might be present (Chang *et al.*, 1994). The method targets polyadenylated RNAs, and is thus ineffective for plant viral RNAs that are not polyadenylated. The polyA tail is used to exclude the large amounts of rRNA usually present in RNA preparations. The polyA requirement in 'random' hexamer amplification can, however, be bypassed by using a mixture of hexamers that has been depleted of rRNA-complementary oligonucleotides (Endoh *et al.*, 2005). Nevertheless, a prerequisite for RDA is the availability of a pair of samples, one infected and one not infected, the latter to be used as the subtracting agent. This requirement makes RDA difficult to apply to plants from natural environments, unless one is focusing on plants with obvious symptoms of disease and has plants of the same species available that are not infected. RDA can be used to identify novel viruses in symptomatic plants infected with viruses of unknown etiology.

The method (Chang *et al.*, 1994) relies on PCR to differentially amplify non-homologous pathogen sequences present in the 'tester' DNA (from infected sample) but absent from the 'driver' DNA (from uninfected sample). The tester and the driver DNA samples are digested using a restriction enzyme, and an adapter is added only to the tester DNA digest. The driver DNA is used in excess over the tester to drive the reaction. The digests are combined, heated to melt the double-strands and then cooled to anneal them back together. This results in the formation of three kinds of molecules: tester/tester, tester/driver, driver/driver sequences. Due to excess of driver DNA, the tester/tester molecules will be enriched for pathogen sequences because the non-pathogen tester sequences will anneal to the complementary DNA fragments of the driver DNA. The ends of the re-annealed DNA are filled in, followed by PCR amplification with a primer specific for the adapter sequence. The tester/tester molecules with the pathogen sequence will be amplified exponentially since they contain adapter sequences on both ends. The

tester/driver fragments will undergo linear amplification since they have only one adapter sequence, whereas the driver/driver fragments will not be amplified due to lack of adapter sequence. Nuclease digestion (mungbean nuclease) is usually used to remove unwanted ssDNA. More rounds of RDA can be performed by combining the resultant pathogen-enriched amplicons with an excess of driver DNA restriction enzyme fragments.

Amplification

In most genomic approaches for detection and identification of novel viruses, the nucleic acids obtained by the enrichment methods discussed above are not suitable for direct use in detection methodologies. Often, detection methodologies require shorter molecules than are generated by enrichment. Long nucleic acids hybridize poorly to microarrays. Many sequencing methodologies require access to sequences from free ends of molecules. Sequencing requires multiple coverages of the same genomic region. Array hybridization is driven by the concentration of the target nucleic acid. For these reasons, various amplification approaches have been employed.

SISPA and VIDISCA

Since its original description, PCR has been refined in many ways to fulfill the requirements of multiplex amplification, as well as amplification of targets with unknown genome sequences. In humans, viruses have been recovered from clinical specimens by variants of a technique known as sequence-independent single primer amplification (SISPA). DNA obtained by SISPA from serum was cloned in lambda gt11 and clones encoding human astrovirus antigen were recovered after immunoscreening, resulting in the first sequences of the virus (Matsui *et al.*, 1993). This effort was followed by the use of SISPA in the recovery of a hepatitis G virus genome (Linnen *et al.*, 1996). In SISPA, an asymmetric adapter sequence is ligated to both termini of blunt-ended DNA molecules for subsequent PCR amplification with a single primer. The feasibility of the method has been demonstrated for both single-stranded and double-stranded

RNA and DNA viruses (Ambrose & Clewley, 2006). The adapter sequences can contain restriction endonuclease sites to facilitate cloning (Reyes & Kim, 1991). HCoV-NL63, a new human coronavirus, was recovered using a variant of SISPA called virus-discovery-cDNA-amplified fragment length polymorphism (VIDISCA) (van der Hoek *et al.*, 2004). The method follows the same principle as SISPA, except that it uses two primers in the PCR step specific to each adapter attached on the ends of the DNA fragment, as is done in the AFLP technique.

Random PCR

Random PCR, a technique similar to SISPA, uses the first primer with a unique nucleotide sequence at the 5′-end, followed by a random or degenerate sequence at the 3′-end. A subsequent PCR amplification step is carried out with a second, specific primer complementary to the 5′-region of the first random primer. This removes the need for an adapter ligation step required in the SISPA approach. A modified version of such a random PCR amplification strategy was utilized by Wang *et al.* (2003) for amplifying viral nucleic acid to be identified using a microarray. RNA was reverse transcribed using a random primer having a unique sequence at the 5′-end, followed by the second strand synthesis using Sequenase. The product was then used as the template for PCR amplification using just the 5′-sequence of the first random primer. The study revealed the presence of a previously uncharacterized coronavirus in a viral isolate cultivated from a severe acute respiratory syndrome (SARS) patient. Such amplification strategies have also been used for RNA or DNA viral sequence recovery from plant samples through high-throughput sequencing of the cDNA libraries. One such example included successful adaptation of the Wang *et al.* (2002) procedure for virus discovery in the PVBE project (Melcher *et al.*, 2008; Muthukumar *et al.*, 2009). The procedure was developed to sequence nucleic acids amplified and cloned randomly from virus-like particle fractions of plant homogenates. The extracted nucleic acid (RNA and DNA) is, in the first round, subjected to reverse transcription, followed by second strand synthesis, both with random dodecamer 3′-terminated oligonucleotides

whose 5′-ends contain a defined sequence. In the second round, the defined sequence alone is used as a single primer for standard PCR amplification. Evidence was obtained for the presence of several undiscovered viral sequences belonging to different virus families. Array-based techniques utilizing such random amplification methods have also been pursued to prove their worth in detection of previously unknown viruses. A macroarray-based approach using a similar random amplification strategy has been developed for detection of plant RNA viruses (Agindotan & Perry, 2007, 2008). In another approach using microarrays, nucleic acids derived from plant specimens infected with *Tymovirus* species were subjected to a similar version of the random PCR followed by cRNA generation (Grover *et al.*, 2010). The method was adapted to amplify both RNA as well as DNA viral genomes. The results indicated the possible potential of the technique for virus discovery by detecting viral sequences with 70-75% or higher sequence identity.

RCA

One of the most useful methods for amplification of whole circular DNA genomes is rolling circle amplification (RCA) using the phi 29 DNA polymerase. The distinguishing features of this polymerase include its high processivity, strand displacement activity, proof-reading activity and synthesis of long products, which make the enzyme most suitable for the efficient amplification of circular DNA molecules. The application of RCA for detection of unknown circular viral genomes is achieved through multiply primed RCA, where random primers bind to multiple sites of the template molecule. The method entails strand displacement and amplification of the viral genome by the enzyme, followed by analysis of the genome-length amplification products by cloning and sequencing. The technique has successfully been used for the detection of several novel papillomaviruses from different animals (Rector *et al.*, 2004; Rector *et al.*, 2005; Rehtanz *et al.*, 2006). For plant viruses, multiply primed RCA was first applied to amplify the complete circular DNA genome of a bipartite begomovirus, DF-BR2, infecting tomatoes (Rehtanz *et al.*,

2006). RCA can also amplify DNA from the nuclear covalently closed form of hepatitis B virus (Margeridon *et al.*, 2008) and thus, presumably, also of the nuclear genome from members of the *Caulimoviridae*. Recent studies investigating cereal-infecting and tomato-infecting geminiviruses in Germany and Thailand respectively, revealed several new geminivirus species using direct sequencing of RCA products (Agindotan & Perry, 2007; Knierim & Maiss, 2007; Schubert *et al.*, 2007). Shepherd *et al.* (2008) showed that RCA amplification could be successfully applied to amplify and further clone nucleic acids from dried plant tissue samples stored up to 6 months at room temperature, though virus genomes could also be cloned from 47% of 10-year-old samples. The multiply primed RCA approach has become a powerful tool for the detection of unknown viruses since it allows the detection of circular DNA viral genomes without the need of specific primers.

In the PVBE study, surprisingly little evidence of the presence of geminiviruses was found. This could be due to the biodiverse nature of the TPP ecosystem not being conducive to geminivirus establishment, or an inadequacy of the methods used. The VLP-VNA and dsRNA methods have not been tested with geminivirus-infected material. Thus, it is possible that the expected failure of dsRNA in this regard also extends to VLP-VNA. The recent development of rolling circle displacement amplification using the phi 29 DNA polymerase promises to be a method that can fill the gap (if there is one).

Macro/microarray targets

The first step in target sample preparation for microarrays is the extraction of total nucleic acid or viral nucleic acid, which generally entails a combination of phenol-chloroform extraction and nucleic acid precipitation using a few milligrams of suspect tissue. The second step involves enrichment of the extracted nucleic acid for detection. Target nucleic acid concentration is a significant determinant of efficient hybridization. Species-, genus- or family- specific primers (Deyong *et al.*, 2005; Sugiyama *et al.*, 2008) can be used for amplification of known viruses, but

not for detection of emerging viruses of unknown taxa. In the case of microarrays for RNA viruses occurring at high concentrations, labeled cDNA targets can be generated by direct (Boonham *et al.*, 2003; Lee *et al.*, 2003) or indirect (Pagan *et al.*, 2009) incorporation of the label during reverse transcription using random primers, without performing PCR amplification. However, for viruses present in lower titers, target amplification is needed to increase the probability of virus detection. There are many groups of plant viruses for which no effective generic primers are available, due to extreme nucleotide sequence variability of genomes or scanty sequence information.

The rapid sequence-independent amplification approach originally described by Bohlander *et al.* (1992) was adapted and used in a macroarray system for detection of plant RNA viruses (Agindotan & Perry, 2007). Another recent sequence-independent amplification-based microarray approach had the capacity to assess, in one assay, the presence of multiple known or related unknown plant viruses (Grover *et al.*, 2010). The method generated randomly amplified target nucleic acid followed by incorporation of amino-allyl-modified nucleotides during *in vitro* transcription. The resulting cRNA was labeled with a dye by coupling to reactive esters. Reverse transcription was performed using chimeric anchor-random primers followed by second-strand cDNA synthesis using just the anchor primer to aid incorporation of a sequence recognized by SP6 RNA polymerase for subsequent *in vitro* transcription. Since fluorescently labeled nucleotides are not efficiently incorporated during reverse transcription due to steric hindrance caused by dye molecules (Zhu *et al.*, 1994), combining sequence-independent target amplification and *in vitro* transcription with indirect labeling ensured a highly efficient label incorporation as well as a sufficient target yield of the final cRNA product.

Target nucleic acid lengths are well known to influence the ability of duplex formation and consequent hybridization signal intensity (Liu *et al.*, 2007; Peplies *et al.*, 2003; Peytavi *et al.*, 2005; Southern *et al.*, 1999; Zhu *et al.*, 1994). Secondary structure formation in longer targets can cause a decrease in hybridization efficiency by 10^5 to 10^6-fold by reducing the binding constant

with probes, increasing false-negative signals (Lima *et al.*, 1992). Strategies known to be adopted for reducing the effects of secondary structure in the target nucleic acid include fragmentation of the target (Lane *et al.*, 2004; Mehlmann *et al.*, 2005), incorporation of modified bases in the target (Nguyen & Southern, 2000), or including auxiliary oligonucleotides with the target sequence to disrupt secondary structures (Maldonado-Rodriguez *et al.*, 1999).

In conclusion, although sequence-independent amplification and post-synthesis processing methods are sometimes prone to errors producing spurious results, these methods have proved their efficacy in the discovery of previously unknown viruses. They display great potential when combined with suitable end-stage detection methods such as microarrays and sequencing for the identification of unknown viral sequences in wild plants.

Detection

In this section we discuss array hybridization and nucleotide sequencing as the two principle means of detecting novel viruses. The products of the group-specific-primer enrichment approach discussed above are usually analyzed initially by gel electrophoresis, but confirmation of a produced band representing viral sequence requires that the band be sequenced. Similarly, the array analyses, discussed in the following sections, can provide evidence that a virus related to particular known viruses is present in a sample, but confirmation by subsequent sequence-specific amplification and sequencing will be required.

Arrays

Nucleic acid hybridization has already proven to be a powerful tool for detection of virus satellite RNAs as well as viruses which do not produce coat proteins (Harrison & Robinson, 1982; Harrison *et al.*, 1983; Yamaguchi *et al.*, 2005), and are thus non-detectable by serological methods. Array-based hybridization methods have potential utility in discovery of viruses without prior knowledge about the identity of the virus(es) (Grover *et al.*, 2010). Microarrays, first

developed to assay differential expression of mRNAs in different tissues or developmental stages (Schena *et al.*, 1995), are emerging as an important tool in pathogen detection. A number of studies have demonstrated the ability of microarrays to detect both animal and plant pathogens (Chapman *et al.*, 1990; Chiu *et al.*, 2008; Jääskeläinen & Maunula, 2006; Kistler *et al.*, 2008; Mihindukulasuriya *et al.*, 2008; Seifarth *et al.*, 2003; Wilson *et al.*, 2002), including a remarkable application of the technique in the identification of the SARS virus as a member of the genus *Coronavirus* (Wang *et al.*, 2002; Wang *et al.*, 2003). Using current methodologies, microarrays provide the capability for parallel yet specific testing to detect individual viruses or mixtures of viruses in single plant samples with sensitivity comparable to ELISA. Some of the earliest arrays designed for the detection of plant viruses were for potato virus isolates (Boonham *et al.*, 2003; Bystricka *et al.*, 2005) and cucurbit-infecting tobamoviruses (Lee *et al.*, 2003). Since then, numerous successful variations of the technique have been designed for detection of several plant virus groups including characterized and related uncharacterized viruses (Bystricka *et al.*, 2005; Deyong *et al.*, 2005; Grover *et al.*, 2010; Pasquini *et al.*, 2008). These studies demonstrate the value of microarrays as one of the important approaches to be used to identify viral species diversity. They are driven in part by the need to identify pathogenic viruses in economic plants such as grapevine (Engel *et al.*, 2010), known to support a wide variety of viruses, and by the need to screen imported plant varieties for a broad range of viruses (Rao & Singh, 2008; Rodoni, 2009). Microarrays are created by spotting capture probes onto a solid support surface, usually a glass slide. Target nucleic acid is extracted from the test sample, reverse-transcribed, amplified where appropriate, and labeled with a fluorescent dye during one of the processing steps. The labeled target molecules are then hybridized against the arrayed probes. Excess target is washed off from the slide surface and virus presence is detected as fluorescence achieved after hybridization of the labeled target to the sequence-specific spot on the array. Macroarrays, like microarrays, are also based on hybridization and formation of target-probe duplex between the nucleic acid of the pathogen (target) and the complementary pathogen-specific nucleic acid

sequence (probe). One of the main differences between macroarrays and microarrays is scale, with macroarrays typically having tens to hundreds of spots, while microarrays more often have hundreds to thousands of spots. The second distinction is that macroarrays are typically created on membranes, while microarrays are usually spotted on glass or plastic supports. One of the biggest drawbacks of microarrays is the high cost associated with their use, whereas macroarrays are a relatively inexpensive alternative, without a requirement for specialized instrumentation or reagents. On the other hand, macroarrays are limited to a much smaller number of probes per support.

Microarray as well as macroarray hybridization assay systems include the following procedural steps: (1) selection of probes and their immobilization on solid support surfaces (microarray) or membranes (macroarray), (2) target nucleic acid preparation, and (3) hybridization and detection of hybridized products. Variations in array methods include choice of surface support, probe immobilization method, probe type, probe design approach, target nucleic acid processing, target labeling strategy, hybridization and washing conditions, method of scanning and of analysis of the result.

Probe design is of primary importance in development and utilization of array-based detection systems, since probes determine both the sensitivity and specificity of the hybridization reaction. Two different probe types can be used to construct arrays: (1) amplified from genomic DNA or cDNA libraries (Boonham *et al.*, 2003; Lee *et al.*, 2003), and (2) chemically synthesized oligonucleotides (Agindotan & Perry, 2007; Bystricka *et al.*, 2005; Deyong *et al.*, 2005). Using synthetic oligonucleotide probes has advantages over cDNA probes, such as high purity, less intensive preparation and less susceptibility to errors due to cross-contaminating PCR products. In addition, oligonucleotide probes can be modified to orient the binding of probes to the support either by addition of a terminal reactive group or a spacer molecule to reduce steric hindrance during hybridization due to the proximity of the probe to the support surface (Boonham *et al.*, 2003; Grover *et al.*, 2010).

Oligonucleotide probes of 20-70 nt, length depending upon the desired level of detection specificity, have been used successfully (Agindotan & Perry, 2008; Bystricka *et al.*, 2005; Deyong *et al.*, 2005; Pasquini *et al.*, 2008). Short (25- to 30-mer) and long (50- to 70-mer) oligonucleotide probes have their own specific advantages. While long probes provide better detection sensitivity, only short probes allow efficient discrimination between closely related sequences (Chou *et al.*, 2004; Letowski *et al.*, 2004; Urakawa *et al.*, 2003). Several probe design software programs are available (Emrich *et al.*, 2003; Wernersson *et al.*, 2007), and are continually being improved to aid in the choice of oligonucleotide probes from large datasets. Four main criteria considered during the design of probes for microarrays are: (1) the desired level of specificity to their respective targets, (2) inability to form stable secondary structures that may hinder target accessibility to probes, (3) consistency in their thermodynamic properties such as melting temperatures and (4) absence of complementarity to other nucleic acids that might contaminate targets, e.g. host plant DNA.

Focusing on highly variable regions of viral genomes for probe design can result in probes that are highly strain-specific, useful for epidemiological studies of virus distribution and spread. Focusing on moderately conserved regions leads to probes that can detect all viruses belonging to a species. Focusing on highly conserved regions yields probes that recognize viruses at the genus and family levels. There exists a potential for creating arrays containing degenerate family- or genus-specific, species- and strain-specific probe sets to target different taxonomic levels of viruses and increase the accuracy of identification. Probe deposition on the support surface can be achieved either manually (Agindotan & Perry, 2007) or robotically (Boonham *et al.*, 2003). Several factors such as spotting buffer (in which the probes are dissolved), temperature and humidity maintained while printing can influence spot morphology, and must be considered to prevent artifacts during analysis.

Hybridization and washing parameters, pre-hybridization procedures, duration and temperature of hybridization, salt concentration and pH of the hybridization buffer and the

stringency of washing steps must be optimized to achieve the best sensitivity and specificity of detection (Boonham *et al.*, 2007). The free binding sites on the support surface are usually blocked with a protein and/or a non-homologous DNA before hybridization. Choice of an appropriate hybridization temperature and salt concentration is very crucial, since the resulting stringency will determine to what extent near perfect matches will be discriminated from perfect matches. Use of formaldehyde in the hybridization buffer is practiced to reduce the hybridization temperature. Stringency of the washing procedure can be enhanced or reduced by decreasing or increasing the salt concentration of the washing solution. After the washings, target–probe duplexes can be detected using a method dictated by the choice of labeling. High-resolution laser confocal scanners are generally employed for detection of fluorescent dye labels used in microarrays (Boonham *et al.*, 2007). For membrane hybridizations, autoradiography using X-ray film or scintillation detectors are employed to visualize the hybrids in the case of radioactive probes and probes generating chemiluminescence. Some of the non-radioactive labeling detection methods rely on antibodies or other chemicals attached to enzymes that can cause formation of colored precipitates from an appropriate substrate (Agindotan & Perry, 2007; Sugiyama *et al.*, 2008).

In conclusion, exploiting the use of taxonomically high-level probes, such as genus- or family-level probes, with non-specific amplification and labeling methods, provides great potential for microarrays in discovery of new or uncharacterized viruses. Microarrays can enable the detection of unexpected interactions of already known viruses or relatives of known viruses in new plant hosts. Moreover, the use of degenerate family- or genus-specific amplified products as targets for microarrays provides great promise to accelerate the discovery of new or unusual viruses.

Sequencing

Deep sequencing (also known as megasequencing) of cDNAs made from mRNA populations associated with a single organism has been shown to be capable of discovering previously unknown virus-like sequences. In humans, deep sequencing of cDNA from RNA extracted from post-mortem liver samples or serum samples resulted in the identification of a novel arenavirus, Lujo virus (Briese *et al.*, 2009). With plants, in one approach, large quantities of cDNA from a plant specimen were subjected to megasequencing and subsequent bioinformatic assembly of the resulting reads (Adams *et al.*, 2009). Test application of the method to a tomato plant infected with *Pepino mosaic virus* resulted in one-fifth of the sequences deriving from the virus with 97% coverage of the genome. When applied to an unknown virus from *Liatris spicata,* complete genome coverage of a novel cucumovirus was obtained with two-fifths of the sequence reads deriving from this virus. Similarly, cDNAs from a Syrah grapevine undergoing vine decline resulted in the identification of signatures of seven viruses or viroids including a previously undescribed marafivirus (Al Rwahnih *et al.*, 2009). The approach is not specific to crop plants or plants with obvious symptoms of disease and thus could be applied to healthy-appearing non-cultivated plants. At present, the expense may prohibit surveys of hundreds of plants from multiple regions. However, anticipated advances in sequencing promise soon to make such deep analysis possible.

Specifically amplified bands can, of course, be directly sequenced using Sanger chain-termination methods and the primers used in the amplification PCR. The advantage of this approach is the longer reads generated, thus reducing the problem of misassembly of a series of shorter contiguous sequences. Cloning of specifically amplified bands or of randomly amplified sequences, followed by chain-termination sequencing, has the same advantage as direct sequence of minimizing misassembly. The disadvantage is that minor variants in the population may be inadvertently focused on as representing the whole sequence.

In both deep sequencing of cDNA and specific amplification, the amplified targets are

submitted for sequencing. The traditional method of creating a clone library with subsequent sequencing of clone inserts has been used for the VLP-VNA method of the PVBE project (Melcher *et al.*, 2008). More cost effective and higher yielding is the use of pyrosequencing. In this procedure (Roossinck *et al.*, 2010), the final amplification before bead attachment is done using oligonucleotides with sample-specific tags. Four nucleotide tags are sufficient to differentially tag a set of 96 samples. Runs of two or more of the same nucleotide are avoided in tag design due to increased chances of computationally misassigning sequences obtained by pyrosequencing.

It must be stressed that obtaining sequences by these methods from a plant specimen grown outside the laboratory, whether from cultivated or non-cultivated plants, is different from the traditional way of determining definitive sequences of viral genomes. In the latter case, the virus has been purified from the original source and propagated in a plant known to be free of viruses. It is not uncommon to find field grown or non-cultivated plants that have evidence of the presence of multiple viruses. The possibility of multiple infections requires cautions in assembling sequence reads, since without sufficient overlap between reads, sequences from separate viruses of the same plant specimen may be assembled mistakenly. Where the virus is at high titer and the entire sequence run is devoted to a single source of plant material, complete or almost complete coverage of the genome sequence is likely to be obtained. Nonetheless, the 5′ and 3′ ends are likely to remain undefined, unless pursued independently, because of the random nature of the priming used in amplifying the sample nucleic acids. When the titer is low or when multiple samples are processed in a single sequencing run, each with a distinctive tag at the end, the likelihood of internal gaps in the sequence is higher. When the contigs are aligned relative to the framework of a known virus, and they do not overlap, one can infer that they are different parts of the same genome. However, should there be two or more contigs overlapping one region, it will not be possible to match them reliably with one or more contigs from another region of the genome. Gaps between contigs are not randomly distributed. In the PVBE project, some areas of

genomes had eight to ten or more fold coverage than others which were represented by only one or two contigs. The non-randomness is due to the choice of randomly terminated primers. PVBE used primer ends with 12 ambiguous nucleotides. The sequence attached to N_{12}, however, influenced the places at which amplification initiated, judging by the coterminality of many sequence reads at a short sequence strongly resembling that of the primer. This problem can potentially be overcome by training software on known sequences to identify sequences unlikely to favor priming at certain sites unintentionally (Wong *et al.*, 2007).

Analysis of the sequence data to identify sequences of relatives of known viruses is typically begun with BLASTn and BLASTx searches using the nucleotide and deduced protein sequences as query of nucleotide and protein databases, respectively. The BLASTx searches are usually more productive in identifying putative viral contigs, because the nucleotide sequences found in the non-cultivated plants are often considerably diverged from their known relatives. Still, these searches do not always retrieve all sequences of a particular virus. Further sequences may be retrieved by searching the data using amino acid sequences of known viruses as tBLASTn queries. These searches may, however, retrieve sequences of viruses of taxa distantly related to the virus whose amino acid sequence was the query.

In both the dsRNA approach and the VLP-VNA approach, there were also sequences that could be assigned a plant origin and sequences clearly derived from bacteria or fungi. The proportion of unassignable sequences depended on the plant species that gave rise to the sample. For example, a species of *Vitis*, relative of grapevine, whose genome sequence has been completed, had no unassignable sequences, while a sample of a brown alga, considerably less well studied, had the highest proportion of unassignable sequences (Muthukumar *et al.*, 2009). The analysis methods discussed above will identify relatives of known viruses. Viruses belonging to completely unknown taxa will be overlooked by this approach. Theoretically, an approach in which resulting sequences are compared to those from a known uninfected plant, such as using RDA, to identify novel sequences could identify sequences belonging to novel viruses. However,

in dealing with non-cultivated plants, there is no such thing as a certifiably uninfected plant. An approach that should work in cases where multiple specimens with multiple sequences are being examined is to retrieve unassigned sequences and perform a self BLASTn search. These approaches rely on the unknown virus being present in reasonable titer in multiple specimens to exclude the possibility that absence in some specimens is due to statistical randomness.

Many megasequencing studies of viruses produce, as end result, a determination of what fraction of the total nucleotides sequenced were assignable to individual genera or families. Assignment to species is more difficult. If contiguous sequences are generated that cover the region designated by ICTV (Fauquet *et al.*, 2005) as being determinant at the species level (sequences with greater than a specified percentage identity are regarded as belonging to the same species), assignment to a species is straightforward. If not, alignments need to be generated to allow pairwise comparisons. If the sequence of interest shows a level of difference equivalent to that distinguishing the closest related species among themselves, a good tentative conclusion is that the sequence is from a separate species.

Soon to be realized improvements in genomic megasequencing will yield many new plant genome sequences. These are fertile ground for exploration for viruses. Genome searches for certain viruses, especially pararetroviruses, have been performed (Hohn *et al.*, 2008). Directly similar viral sequences will be recognized by the annotation processes, but more distant relationships may require specialized searching of the genomes of these plants. Difficulties in discriminating between actual viruses and endogenous genetic elements have been anticipated (Bousalem *et al.*, 2009). Unfortunately, the plants whose genomes will be sequenced will probably be crop and ornamental plants. Nonetheless, the sequences will expand our knowledge of virus biodiversity and possibly reveal viruses whose lineages have died out leaving no trace among extant viruses.

FUTURE PROSPECTS

Anticipated developments in several areas will make more extensive virus biodiversity inventories likely in the future. Multiple DNA sequencing methods alternative to chain termination and pyrosequencing have been developed recently and more are promised for the near future (Eid *et al.*, 2009; Fuller *et al.*, 2009; Harris *et al.*, 2008). Direct sequencing of RNA has also become a possibility (Ozsolak *et al.*, 2009). These approaches should allow the massively parallel sequencing of plant-associated and environmental nucleic acids to obtain very deep coverage of individual samples or to allow the simultaneous analysis of large numbers of appropriately tagged samples. These developments will require further improvements in computational processing of the large amounts of data, including methods to test for the accuracy and reliability of sorting and assembly processes (Zimin *et al.*, 2008). In particular, analysis of recombination is jeopardized by *in silico* recombination events.

Improvements are also anticipated in virus concentration methods. These may involve specialized coatings capable of adsorbing a variety of viral particles, or the development of molecules able to bind dsRNA selectively from extracts of plant material. Amplification methods may also be improved and standardized (Djikeng *et al.*, 2008). Chromatography matrices that can bind and thus concentrate RNA (Branovic *et al.*, 2003) may be useful in providing sufficiently concentrated starting material from aqueous sources for surveys of viruses in bodies of water. Dielectrophoretic coatings in flow devices may be able to concentrate virus particles from complex mixtures (Davalos *et al.*, 2008).

Eventually, it will be necessary to create profiles of viruses present in large areas. Sampling schemes will need to be employed that will at the same time allow good confidence of recognizing all viruses in an ecological region, and yet be able to identify host or geospatial heterogeneities in the viruses detected. An alternative that may provide the larger-scale

information without the specific knowledge of plant source is the examination of drainage water from watersheds of interest. Monitoring of streams for the presence of specific viruses has been documented (Boben *et al.*, 2007). In addition, virus particles have been purified from water from Lake Needwood by tangential flow filtration and analyzed by metagenomic methods (Djikeng *et al.*, 2009). From 60% to 70% of the source-identifiable sequences of the nucleic acids from this fraction were from viruses. The waters from Lake Needwood contained examples of 28 virus families of which about 45% were likely of plant origin (Djikeng *et al.*, 2009). Imperfect database matches suggested the presence of many previously undescribed plant-associated viruses. Soil may adsorb particles from decayed material and thus be another source. Assemblages of vectors, such as arthropod sweeps or nematode soil filtrates, can be examined to determine virus populations of plants that the vectors have sampled (Martin *et al.*, 2009). In all such cases, though, there is selection for virion particle stability, such that viruses with less stable particles will be underrepresented in the samples.

Although nucleotide sequencing will continue to be the primary method of expanding our knowledge of virus biodiversity, other methods, under development, may also contribute. A major failing of nucleotide sequencing is that it can only recognize sequences as viral sequences when they are related to those of known viruses. Viruses with completely unrelated sequences may exist. In a metagenomic survey of virus diversity, these sequences will be in the category of sequences that have no database hits either with nucleic acid or with protein. In this regard, advances in mass spectroscopy may make it possible to identify likely viral capsid sequences from virus particle preparations. Screening of environmental samples by electron microscopy can reveal viral particles and suggest RT-PCR or PCR tests for detection of the genomes inhabiting those particles.

CHAPTER III

FIRST CHARACTERIZATION OF A PLANT VIRUS BELONGING TO THE FAMILY *TOTIVIRIDAE*

INTRODUCTION

During a recent ecogenomic study of plant viruses in the Tallgrass Prarie Preserve (TGPP) of northeastern Oklahoma, several putative viral signatures were discovered (Melcher *et al.*, 2008; Muthukumar *et al.*, 2009; Roossinck *et al.*, 2010) . Most of these viruses which were determined across a wide range of host plant species are new to science. One of the most prevalent groups of viruses in the TGPP belonged to the family *Totiviridae*. Viruses belonging to the family *Totiviridae*, known to infect fungi, protozoa and arthropods (Ghabrial & Suzuki, 2009; Zhai *et al.*, 2010) are one of the best characterized fungal viruses. The viral members of the family have an undivided dsRNA genome comprising of a coat protein (CP) gene and an RNA dependent RNA polymerase (RdRp) gene, and employ three different mechanisms for expression of the RdRp (Ghabrial & Suzuki, 2009). At present, four genera have been formally recognized in the family *Totiviridae*: *Totivirus, Giardiavirus*, *Leishmaniavirus*, *Victorivirus*. Viruses in the genus *Totivirus* are known to infect yeast, smut fungi, protozoa and arthropods. Members in the genus *Victorivirus* infect filamentous fungi, while those in the genera *Giardiavirus* and *Leishmaniavirus* infect parasitic protozoa. Totiviruses are packaged in isometric particles containing an

unsegmented, uncapped 4.6–6.7 kbp dsRNA genome. The member viruses in the genus *Totivirus* are known to be associated with dsRNA species suspected of being satellite or defective dsRNAs. In yeast and smut totiviruses, satellite dsRNAs can encode killer toxins that provide a benefit to the infected host (Ghabrial, 1998).

The study presented here focuses on a putative virus belonging to the family *Totiviridae,* discovered during the virus biodiversity survey and found in a number of plant host species from several different plant families. The incidence of occurrence of this virus was most frequent and in highest titers in *Ruellia humilis* among the six plant species (*Ambrosia psilostachya, Asclepias viridis, Panicum virgatum, R. humilis, Sorghastrum nutans, Vernonia baldwinii*) chosen for repeated sampling for four consecutive years (2005, 2006, 2007 and 2008) of the survey. Double stranded RNAs, considered a hallmark of RNA virus infection (Dodds *et al.*, 1984), of ~ 5 Kbp were detected from virus-infected plant samples as a first sign of the presence of the virus. The putative virus was recovered in plant samples collected from different locations in the prairie for all four years, suggesting an even temporal and spatial distribution of the virus in the TGPP. Under the assumption that non-overlapping contigs obtained from the same plant specimen derive from the same viral species, the sequence data obtained during the PVBE study suggested several variants and strains of this putative virus, and quite a few related viruses in the TGPP plants. Since the virus-infected plants did not show any obvious viral symptoms, we hereby designate the putative totivirus, found predominantly in *R. humilis*, as *Ruellia* asymptomatic virus 1 (RAV1). Further research on this virus in the present study was conducted using *R. humilis* plants collected from the TGPP.

The genome size of RAV1 was in the range characteristic of a totivirus species, and an almost complete sequence of the dsRNA genome (~ 4700 bp) of the virus was assembled. The putative RdRp encoded by the dsRNA showed highest similarity to the RdRp of Black raspberry virus F, a dsRNA totivirus assumed, without supporting evidence, to be a fungal virus. There were two possibilities involved in this case. First, the virus is a mycovirus of a fungal endophyte

that is colonizing the host plants. This seemed a little unlikely since the virus was often found in high titers and only a small amount of fungal tissue is usually found in plants harboring endophytes. The second possibility was that we may have found the first described plant virus, replicating in the plant cells, in the family *Totiviridae*.

Virus families containing mycoviruses have been observed to include members infecting organisms other than fungi. A number of viruses in the family *Partitiviridae* are plant viruses and some members in the families *Partitiviridae and Totiviridae* also infect protozoa. However, there currently are no formally recognized plant viruses classified in the family *Totiviridae*. Some recent publications have mentioned totiviruses in plants (Covelli *et al.*, 2004; Kozlakidis *et al.*, 2006; Marais *et al.*, 2009) but no strong evidence has been provided to identify them as plant viruses. The dsRNAs associated with cherry chlorotic rusty spot and amasya cherry disease are suspected to be genome components of fungal viruses since fungal mycelium was visible by microscopic examination in the affected leaf areas (Alioto *et al.*, 2003). All attempts to culture or isolate fungi from plants failed, making their designation as fungal viruses uncertain. In another report, the dsRNA isolated from apparently healthy blackcurrant that showed similarity to the RdRp of the type member of the genus *Totivirus* was thought to be derived from fungi infecting *Ribes* (Cox *et al.*, 2000). In the case of blueberry fruit drop disease, the dsRNA isolated from symptomatic blueberries showed closest sequence similarity to a member of the genus *Totivirus*, suggesting it could be a plant virus but more work was needed to further identify the source of the virus in plants (Martin *et al.*, 2006). Thus, it remained to be determined whether the dsRNAs associated with these viruses represent genomes of plant or fungal viruses. The first objective of this study was therefore to assess in more detail the host of the discovered putative totivirus, suspected to be a plant virus, bearing in mind the possibility of the virus being a mycovirus.

Plant viruses in the family *Partitiviridae*, which also includes viruses in fungi, are one of the best studied plant persistent viruses, previously called cryptic viruses. The term “persistent virus” in plants is based on one of the four lifestyles a virus can adopt, namely persistent, acute,

chronic and endogenous (Roossinck, 2010). The characteristic properties of a persistent plant virus include lack of apparent disease symptoms induction, low viral concentration in the host, no horizontal transmission of the virus, no cell to cell movement due to the lack of movement protein and a near 100% seed transmission of the virus. Although, so far, all of the viruses belonging to this group have double-stranded (ds) RNA genomes, proposed to be particularly suitable genome for persistently intracellular viruses (Buck, 1986), this characteristic is not a criterion for persistence. The persistence of a virus tends to be highly host specific, possibly due to the need for a close coordination of the virus with host regulatory systems (Villarreal, 2005). Virus persistence appears to rely on host mechanisms for virus maintenance and in animals probably for competition and exclusion of other viral agents as well. The exact mechanism of virus persistence and maintenance and their consequences for the infected host are not very clear. It seems likely that persistent plant viruses exist in nature in a number far larger than determined. Two of the important characteristics of a persistent plant virus that promoted the idea that the putative totivirus may also be a persistent plant virus were: the putative virus did not express any apparent symptoms on infected host plants and the retrieved sequence information for the viral genome showed no obvious movement protein gene encoded by the virus. So, the second objective of the study was to test for the persistence of the virus in its host plants.

Here, we represent results of research that addressed two main questions. Firstly, whether the putative totivirus, designated as *Ruellia* asymptomatic virus 1, is a mycovirus of a fungal endophyte that is colonizing the host plants or a plant virus. Secondly, if the virus is a plant virus, is it possibly a persistent plant virus. Using *in situ* hybridization (ISH), the viral RNA was shown to localize inside the cells of the virus-infected plant stem sections, indicating it to be a plant virus. Seed transmission analysis on the progenies of virus-infected plant samples demonstrated vertical transmission of the virus, and hence persistence of the virus in its host plants. Another interesting question considered in the study was the likelihood of transmission of the virus between the plant and fungi.

METHODS AND MATERIALS

Plant material

A total of 33 wild *R. humilis* plants were collected without reference to symptoms from the TGPP of northeastern Oklahoma. The intact plant were harvested at the late flowering stage, along with their roots on August 22, 2009, and transplanted into pots on Aug. 23, 2009. The soil media was a 50%/50% (v/v) mixture of native top soil and Turface MVP (Profile Products, Buffalo Grove, IL, USA). After the plants recovered from their transplant shock, and new growth had emerged, the plants were watered with tap water every other day, and fed bimonthly with 1g/L Peter's Peat-Lite fertilizer (Scotts, Marysville, OH, USA). The environmental conditions in the air-conditioned greenhouse averaged to 24 °C, 50 % relative humidity, and 230 umole s^{-1} m^{-2} photoactive radiation with a photoperiod of 16 h. After a decline of the plant population was noted, a transplant was performed with the surviving population. All healthy plants were transplanted into new geranium pots. The soil media was 40/60% (v/v) mixture of Turface MVP and autoclaved Metro Mix 350 (Sun Gro, Bellevue, WA, USA).

Screening of plants for the presence and absence of the putative totivirus

All the primers used in this study are listed in Table 3. To screen plants for the presence of the putative totivirus, 100 mg of leaf tissue was harvested from each *R. humilis* sample collected from the prairie. The tissues were ground in liquid nitrogen and mixed with 200 µl of extraction buffer (0.1 M NaCl, 0.01 M Tris, pH 8, 0.1 mM EDTA, and 1% SDS). The mixture was extracted with 200 µl phenol:chloroform (1:1 [v/v]) twice and precipitated by addition of 3 M sodium acetate (25 µl) and 100% ethanol (500 µl). The nucleic acids were pelleted at 11,000 × *g* for 10 min and the pellets were washed in 500 µl of 70% ethanol. The final RNA pellets were resuspended in 50 µl of sterile H_2O. Reverse transcription, PCR amplification, cloning and

Table 3**:** List of all primer sequences used in the characterization of the putative totivirus.

Primer	Sequence 5' to 3'
Primer 1: Random dodecamer	CCTTCGGATCCTCCN$_{12}$
Primer 2: Linker	CCTTCGGATCCTCC
Primer 3: TotiFwd1	GGCAGTATCA
Primer 4: TotiRev1	GCTTGATCCCACC
Primer 5: TotiFwd2	ACAATATACAGAAYKGRAGGCAGTATCA
Primer 6: TotiRev2	ACAATAATGCTARRGCTTGATCCCACC
Primer 7: TotiFwd3	GGACTACATGGACCGAGGAAG
Primer 8: OligodC adapter	GACTCGAGTCGACATCGC$_{17}$
Primer 9: RACE adapter	GACTCGAGTCGACATCG
Primer 10: TotiRev3	CATGCTTGTGACTGCATTCCTC

sequencing were performed as previously described (Roossinck *et al.*, 2010). Briefly, RNA was converted to cDNA by reverse transcriptase using a random primer with a linker sequence at the 5'-end (Primer 1) followed by removal of any un-reacted template and primers by treatment with RNase A and further heating to 85°C. The samples were purified using Qiagen PCR purification columns (Qiagen, Valencia, CA, USA) and eluted in 0.1X EB buffer. This was followed by PCR amplification using the linker primer (Primer 2) which had various combinations of four nucleotide tag sequences attached to them for individual samples (Roossinck *et al.*, 2010). The amplified products were sequenced on a Roche 454 GS FLX sequencing machine.

Mechanical transmission of the virus

To test for the mechanical transmission of the virus, leaf tissues from two infected and two uninfected *R. humilis* plants were separately homogenized in 0.1 M phosphate buffer, pH 7.2. The homogenates were rubbed onto the carborundum-dusted leaves of two herbaceous species including *Nicotiana benthamiana* (NB) and *Chenopodium quinoa* (CQ). A total of 8 plants (4 NB and 4 CQ, inoculated with two totivirus-positive and two totivirus-negative homogenates) were kept in the greenhouse at 25°C, and observed daily over a period of 4 weeks for symptom expression. To test for the infectivity of these indicator plants, ds RNA, which is considered as a hallmark of RNA virus infection, was isolated from all 8 plants 14 dpi as described previously (Roossinck *et al.*, 2010). Briefly, nucleic acids were extracted with extraction buffer and phenol:chloroform (1:1 [v/v]) The aqueous phase was removed and adjusted to 16.5% ethanol and passed through a CF11 cellulose column using a low speed table-top centrifuge. Columns were washed with three column volumes of application buffer followed by dsRNA elution in elution buffer. The dsRNA was precipitated at -20 °C overnight by addition of 3 M sodium acetate and ice cold absolute ethanol followed by a centrifugation at 10, 000 g. The pellet was dissolved in NAE buffer (sodium acetate and EDTA) and again precipitated with absolute ethanol at -20 °C overnight. After centrifugation, pellets were successively washed with 70% and 100%

ethanol, the supernatant was discarded and tubes were air-dried followed by suspension of dsRNA in 50µl of 0.1 mM EDTA pH 8. Once purified, 10 µl of sample aliquots were analyzed on a 0.8% agarose gel stained with ethidium bromide.

Digoxigenin-labeled RNA probe synthesis

To synthesize a labeled RNA probe for *in situ* hybridization, the first step was the synthesis of a cDNA product of 340 bp from within the coat protein region of the RAV1 genome by RT-PCR. The viral dsRNA that served as a template for the RT reaction was isolated from a *R. humilis* during the earlier study, and was stored as -80 °C (Roossinck *et al.*, 2010). The same study also revealed a more or less complete sequence for this totivirus which was used in designing specific primers (mentioned below) for the second round of PCR. Reverse transcription, PCR amplification, cloning and sequencing were performed as previously described (Roossinck *et al.*, 2010). Briefly, RNA was converted to cDNA by reverse transcriptase using a random dodecamer primer with a linker sequence at the 5´-end (Primer 1) followed by removal of any un-reacted template and primers by treatment with RNase A and further heating to 85°C. The samples were purified using Qiagen PCR purification columns (Qiagen, Valencia, CA, USA) and eluted in 0.1X EB buffer. This was followed by PCR amplification using the linker primer (Primer 2). Samples were amplified in an Idaho Technologies Rapid Cycler II using 1.5 µl of the RT product in a 15 µl reaction mixture. The second round of PCR amplification was performed using RAV1-specific primers. The reaction contained a final concentration of 1 x buffer M (medium Mg++ buffer, Idaho Technologies), 0.5 µl 10mM dNTPs, 1µl of 20 µM TotiFwd1 (Primer 3), 1µl of 20 µM TotiRev1 (Primer 4), 1 µl of the randomly amplified product and 2 units of Taq polymerase (Invitrogen). Cycle parameters for the PCR amplification were as follows: 94 ºC for 1 min; 65 ºC for < 1 s; 72 ºC for 45 s, with a slope of 9, followed by 40 cycles of 94 ºC for < 1 s; 45 ºC for < 1 s; 72 ºC for 30 s, with a slope of 5, and a final 5 min at 72 ºC and 5 min at 37 ºC. The resulting amplicon was gel purified and ligated into pGEM-T Easy vector (Promega, Madison, WI, USA)

for transformation into *E. coli* DH5α. Colonies were selected based on colony PCR results (data not shown) and the plasmid prepared for DNA sequencing confirmed that the expected region had been amplified. Linearization of the plasmid was performed using manufacturer's protocol by digestion with *Nco*I (Promega) followed by *in vitro* transcription to generate Digoxigenin (Dig) labeled minus-sense RNA probe. One μg (5 μl) of the linearized plasmid, 2 μl each of 10mM ATP, CTP and GTP, 1 μl each of 10 mM UTP and Dig-labeled UTP (Roche Applied Science, Indianapolis, IN, USA), 1 μl (20 U) of RNAsin (Promega), 2 μl of 10X transcription reaction buffer (Ambion, Austin, TX, USA), 1 μl (20U) of SP6 RNA polymerase (Ambion) and 3 μl of H_2O were mixed and incubated at 37 ºC for 2 hours. The mixture was treated with RNase free DNase (Promega) followed by phenol: chloroform (1:1 [v/v]) purification as described above. The final RNA pellet was resuspended in 100 μl of water. The purified RNA band (~ 400 bp) was visualized on a 0.8% agarose gel stained with ethidium bromide.

Fixation, dehydration and embedding of plant stem tissues

Stem tissues from infected and uninfected plants were cut into small pieces and placed in glass vials containing a fixative solution of 4% paraformaldehyde (PFA). The vials were put in a vacuum chamber twice for 15 min each at 4 ºC with a 1 h shaking in between, also at 4 ºC. The vials containing the tissues and the fixative were then set for an overnight shaking for 14 h at 4 °C. The fixative was removed and tissue samples were treated with 1X PBS (Phosphate buffered saline) twice, each for 30 min at 4 ºC. Samples were then dehydrated in a graded series of ethanol solutions consisting of 30:70, 40:60, 50:50, 60:40, 70:30, 85:15 and 95:5 and 100:0 ethanol in water (v/v) at 4 °C. This was followed by a graded series of Histo-clear (National Diagnostics, Atlanta, GA) solutions consisting of 25:75 50:50, 75:25 and 100:0 Histo-clear in ethanol (v/v) at room temperature. The samples were then infiltrated with Paraplast® embedding medium (Sigma) by transferring them into a 1:4 (v/v) mixture of wax and Histo-clear for

overnight at room temperature, followed by a 1:2 (v/v) mixture for a few hours at 42 °C, and then to pure wax at 60 °C. This was followed by two wax changes each day for three days at the same temperature. Tissues were then embedded into blocks at room temperature and stored at 4 °C until sectioning.

Sectioning and tissue section pretreatment

The embedded tissues were sectioned with a microtome (Sakura Finetek, Torrance, CA, USA) to be 10 mm thick and transferred to Superfrost plus microscope slides (VWR International, West Chester, PA, USA). The sections were immobilized on the slides by adding a drop of water over the section on the slide and placing the slide on a slide-warmer set at 42°C for a few minutes (2-3 min). The extra water around the section was then wiped off the slide and the slide was left on the slide-warmer overnight. The slides were then stored at -80 °C until proceeding for the ISH. For the ISH pretreatment of tissue sections, the slides were treated with Histo-clear twice (10 min each), followed by a graded series of ethanol solutions consisting of 100:0, 95:5, 90:10, 80:20, 60:40 and 30:70 ethanol in water (v/v) for 2 min each. The slides were then rinsed with water (2 min) and 1X PBS (5 min), treated with 100 mM triethanolamine, pH 8.0 mixed with 600 µl acetic acid (in 200 ml triethanolamine) for 10 min. This was followed by two more rinses with 1X PBS for 5 min each and then a final wash with water for 2 min. All of these steps were performed at room temperature.

In situ hybridization for localization of plus-strand viral RNA in the infected stem tissues

Previously published protocols for RNA *in situ* hybridization (Coen *et al.*, 1990; Ding *et al.*, 1996) were modified and used here for localization of the plus-strand viral RNA. The sample slides were treated with 2X SSC (15 min) and proteinase K buffer (100 mM Tris-HCl pH 8, 50

mM EDTA) (5 min) at room temperature, followed by proteinase K treatment (1 µg/ml in proteinase K buffer) (30 min) at 37 °C, washing in 0.2% glycine (4 min) and 1X PBS (12 min) at room temperature. This was followed by a treatment with 4% PFA solution (20 min) and another rinse with 1X PBS (12min), both performed at room temperature. The slides were incubated in hybridization buffer for 2 h at 60 °C. The *in vitro* hybridization buffer contained salts (0.3 M NaCl, 0.01 M Tris-HCl, pH 6.8, 0.01 M Na_3PO_4, and 5 mM EDTA), 50% deionized formamide, 1.25 mg/mL tRNA, Denhardt's solution (0.002 g/L each Ficoll 400, polyvinylpyrrolidone, and BSA), and 12.5% dextran sulfate. The minus-sense Dig-labeled RNA probe was denatured in boiling water for 5 min with subsequent quenching on ice, and mixed with hybridization buffer at a concentration of 300 ng/ml hybridization buffer. One hundred microlitres of probe-hybridization buffer mix was applied to each slide section, and covered with a cover slip. The slides were incubated for hybridization at 60 °C for overnight. The sections were washed, first with 5X SSC (30 min) followed by a washing in 50% formamide in 2X SSC (1X SSC is 0.15 M NaCl and 0.017 M sodium citrate) (1 h) and then with 0.2X SSC (1 h), all steps carried out at 60 °C. All the treatments from here on were performed at room temperature. The slides were washed with maleate wash buffer, pH 7.5 (20 min), followed by treatments with 1% blocking reagent (Roche Applied Science) in maleate wash buffer (30 min), buffer 1 (0.1 M Tris-HCl, pH 7.5, and 0.15 M NaCl) (20 min), and 1% BSA in buffer 1 (30 min). This was followed by a 2 h treatment with anti-digoxigenin-alkaline phosphatase (DIG-AP) (Roche Applied Science), prepared in 1% BSA solution (in buffer 1) at 0.75 U/ml of solution. The slides were again rinsed with 1% BSA in buffer 1 (30 min) followed by buffer 2 (0.1 M Tris-HCl, pH 9.5, 0.1 M NaCl, and 0.05 M $MgCl_2$) for 16 min. Later, the slides were incubated with nitroblue tetrazolium (NBT)/5-bromo-4-chloro-3-indolyl phosphate (BCIP) solution (2 Tablets of NBT/BCIP [Sigma] and 0.0048 g tetramisole hydrochloride [Sigma] in 20 ml water) for 6 to 8 h, rinsed with TE buffer, and then examined under a light microscope.

Ruellia humilis controls for fungal staining

To examine the efficacy of the cotton blue stain for localization of fungal endophytes, we tested the dye on fungal colonized and non-colonized *R. humilis* stem sections. To obtain the fungal colonized and non-colonized *R. humilis* controls, surface sterilized seeds (washed in sterile water, 2% sodium hypochlorite for 1 min, 70% ethanol for 1 min, and again rinsed thoroughly in sterile water), harvested from a totivirus negative plant, were germinated (described in detail in a section below) after a vernalization period of 70 days at 4 °C. Three seedlings were colonized with a fungus (*Curvularia protuberta*, a class II endophyte) by incubating the roots and lower 1/3 portion of the stems in a solution of 0.035% agar containing 10^5 spores per ml, and one seedling without any colonization was used as a negative control. All of the four seedlings were put back in soil and grown for 30 days. Leaves were excised from all four samples (three colonized and one non-colonized) and again subjected to surface sterilization. To test for the successful fungal colonization of plants, small leaf pieces from individual plants were placed on 1X Potato dextrose agarose (PDA) media plates. Only one out of the three colonized plant leaves showed a characteristic fungal outgrowth of *Curvularia protuberata* on PDA plates. It was used as a positive control plant colonized by the fungus. No fungal outgrowth was observed from leaves of non-colonized plant, which was used as a non-colonized negative control. The fungal colonized and non-colonized plants were then processed for fixation, dehydration, embedding, and sectioning of tissues as described above.

Localization of fungal endophytes in stem cross-sections using cotton blue stain

The fungal colonized (positive control) and non-colonized (negative control) *R. humilis* sections as well as the serial sections of totivirus-infected plants, tested to be positive for the viral RNA by ISH, were all stained with cotton blue in a similar manner to test for the presence of fungal endophytes. The tissue section pretreatment before the cotton blue staining was performed on all

the sections as described in a previous section. After pretreatment of the slides, a drop of 70% ethanol was placed over the section followed by a drop of 0.05% lactophenol cotton blue (Hardy Diagnostics, Santa Maria, CA, USA). The section was allowed to stain for 30 min at room temperature followed by destaining of the excess stain by washing with 1X PBS for 5 min. A coverslip was gently placed over the section avoiding air bubbles and the slide was further examined under a light microscope.

Isolation of fungal endophytes from stem tissues of virus-infected plants followed by dsRNA extraction

Isolates of fungal endophytes were cultured from the two virus-infected plant stems on PDA media plates. Plant stems were surface sterilized as described above, and stem pieces were placed on 1X PDA media. Fungal colonization was observed on plates within a week, and were further grown for another week with agitation (100-150 rpm) at 25 ºC in 100 ml of liquid 1X Potato Dextrose (PD) medium supplemented with amplicillin, streptomycin and tetracycline at 50 µg/ml each. Mycelium mats were filtered through Miracloth (Calbiochem, San Diego, CA, USA), frozen at -80ºC for 30 min and lyophilized over night. Two hundred mg of lyophilized mycelium per isolate were ground in liquid nitrogen until completely pulverized. Enrichment of double-stranded RNA was performed as described previously (Roossinck *et al.*, 2010) and mentioned briefly in a previous section. After purification, 10 µl sample aliquots were analyzed on a 1.2% agarose gel stained with ethidium bromide.

Seed transmission analysis

Seeds were harvested from the uninfected and the two virus-infected *R. humilis* plants. Seed coats were removed and naked seeds were surface sterilized as described above. The seeds were vernalized at 4 °C in a moist petri dish for 70 days followed by sowing of seeds in sterile soil.

The moist stratified seeds were shallowly sowed into individual cell inserts filled with 40/60% (v/v) mixture of Turface MVP and autoclaved Metro Mix 350 (Sun Gro). The sowed seeds were placed under a mist system that had a bottom heat. The mist system was switched on for 16 s every hour for 13 h a day to try to keep the soil moist but not very wet. The bottom heat was provided by a heating pad set to 29 °C. The seedlings were transplanted into individual plastic square pots, and the soil media was a 40/60% (v/v) mixture of Turface and autoclaved Metro Mix 350. The plants were moved to a greenhouse, watered every other day, and fed bimonthly with 1g/L Peter's Peat-Lite fertilizer (Scotts). The environmental conditions in the greenhouse ranged from 22-33 °C, 50-97% relative humidity with a photoperiod of 16 hours. After 30 days, 100 mg of leaf tissue was harvested for RNA extraction from 16 progenies of infected plants (nine and seven from the infected parent plant R4 and R22, respectively). The same amount of tissue was used for the RNA extraction from three uninfected plant progenies and three positive control parent plants harvested from TGPP during the earlier study. The tissues were ground in liquid nitrogen and mixed with 400 µl of guanidine extraction buffer (8 M guanidine, 20 mM MES (pH 6.7), 20 mM EDTA, 50 mM mercaptoethanol). The mixture was centrifuged for 15 min at 11,000 × *g* and extracted with 150 µl phenol:chloroform (1:1 [v/v]) twice and precipitated by addition of 100% ethanol (0.75 vol) and 1 M acetic acid (0.25 vol). The nucleic acids were centrifuged at 11,000 × *g* for 10 min and the pellets were washed in 700 µl of 75% ethanol. The final RNA pellets were resuspended in 25 µl of sterile H_2O and then treated with RNase free DNase as per manufacturer's instructions (Promega) followed by phenol: chloroform (1:1 [v/v]) purification as described above. The RT reactions were carried out the same way as described above. Briefly, reverse transcription was performed with random dodecamer primer with a 5'-end linker (Primer 1) followed by RNase A treatment and purification of cDNA using Qiagen columns. The first round of PCR reaction contained 2.5 µl of 10X buffer (Invitrogen), 0.5 µl 10mM dNTPs, 1µl of 20 µM linker primer (Primer 2), 1 unit of Taq polymerase (Invitrogen) and 1 µl of the RT product in a total reaction mixture of 25 µl. Cycle parameters for the PCR amplification were as follows:

94 °C for 3 min followed by 39 cycles of 30 s at 94 °C, 30 s at 55 °C and 1 min at 72 °C with a final 10 min extension step at 72 °C. Specific PCR amplification was performed with the same specifications and conditions as above, using 1 µl of the randomly amplified product and 1 µl each of 20 µM totivirus specific primers (TotiFwd 2: Primer 5 and TotiRev 3: Primer 6). Synthesized PCR products were analyzed on a 0.8% agarose gel stained with ethidium bromide.

Determination of the precise 3'- and 5'-ends of the viral dsRNA

The dsRNA extracted from RAV1-infected *R. humilis* was stored at -80 °C, and served as the template for the reverse transcription reaction. To determine the 3'-end of the dsRNA molecule, RT reaction was performed as described above using 1 µl of 20 µM gene-specific forward primer (350 bp inwards to the known 3'-end) (TotiFwd 3: Primer 7) instead of the random primer (primer 1). This was followed by RNase A treatment and purification of cDNA using a Qiagen column. The synthesized cDNA was then tailed with dGTP. The tailing reaction comprised of 5 µl of RT product, 4 ul of 5X tailing buffer (Promega), 1 ul of 10 mM dGTP, 1 ul of Terminal deoxynucleotidyl transferase (TdT) (20 units) and 9 ul of water. The reaction mixture was incubated at 37 °C for 10 min, 95 °C for 10 min and then cooled on ice. The tailed cDNA was PCR amplified using an oligodC adapter primer (Primer 8) and the gene-specific forward primer. Total reaction mixtures of 25 µl comprised of 2.5 µl of 10X buffer (Invitrogen), 0.5 µl of 10 mM dNTPs, 1 µl of 20 µM Primer 7, 1 µl of 20 µM Primer 8,1 µl of tailing product, 0.5 µl of Taq (2 units) and 18.5 µl of water. Cycle parameters for the PCR amplification were as follows: 94 °C for 3 min followed by 39 cycles of 30 s at 94 °C, 30 s at 55 °C and 1 min at 72 °C with a final 10 min extension step at 72 °C. This was followed by a second round of PCR with the same specifications and conditions as above, using 1 µl each of 20 µM RACE adapter primer (Primer 9) and primer 7. Synthesized PCR products were analyzed on 0.8% agarose gel and an amplicon of ~400 bp was gel purified. The purified DNA product was ligated into pGEM-T Easy vector (Promega) and transformed into *E. coli* DH5α. Colony PCR was performed on five white

colonies as described for the second round of PCR using the adapter primer (primer 9) and the gene-specific forward primer (primer 7), and the amplified products were then submitted for sequencing. Attempts were made in a similar manner to determine the 5'-end of the molecule using a gene-specific reverse primer (TotiRev 3: Primer 10).

RESULTS

Screening of totivirus-infected plant samples

The wild *R. humilis* plants, sampled from the TGPP were transplanted and maintained in the greenhouse. Total RNA was extracted from all 33 plant samples, and subjected to reverse transcription using a random primer with a 5'-linker sequence followed by a PCR using just the linker sequence primer, resulting in the synthesis of dsDNA which was further run through 454 sequencing. For each sample, all contigs assembled by the 454 process pipeline were used as BLASTn and tBLASTx queries of the nr/nt nucleotide database and as BLASTx queries of the nr protein database. The search results obtained suggested the presence of the putative totivirus in two of the 33 plant samples, and the most closely related virus in Blast searches was Black raspberry virus F. The viral contigs identified from the infected samples showed a considerable level of nucleotide sequence identity with RAV1 and BRFV, identifying them as members of the BRFV like virus group. These two plant samples containing totivirus-like sequences were further processed for the localization of the viral RNA using *in situ* hybridization and for all other analysis.

Symptoms and mechanical transmission of the virus

No obvious viral symptoms were noticed in plants screened positive for totivirus presence, and they looked as healthy as uninfected plants (Fig. 9). Attempts to transmit the putative virus through mechanical transmission by rubbing the infected plant leaf homogenates onto the leaves of herbaceous plants including *N. benthamiana* and *C. quinoa* failed. These indicator plants were chosen since they are susceptible to infection by most of the known mechanically transmitted plant viruses. The inoculated plants did not develop any symptoms, nor did they appear to contain the virus when checked by performing a dsRNA diagnosis of the inoculated plants, 14 dpi. As

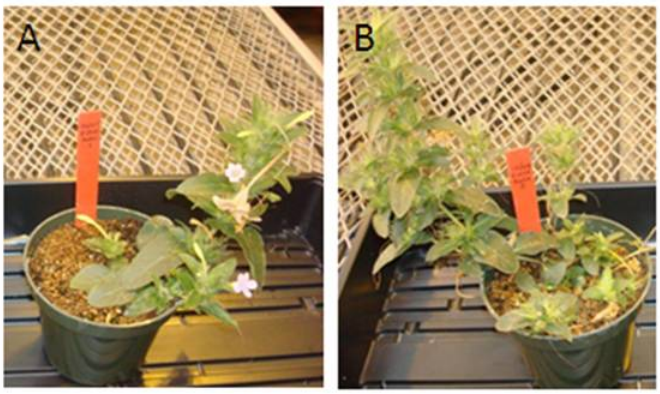

Figure 9: Virus-infected (A) and uninfected (B) *R. humilis* maintained in the greenhouse

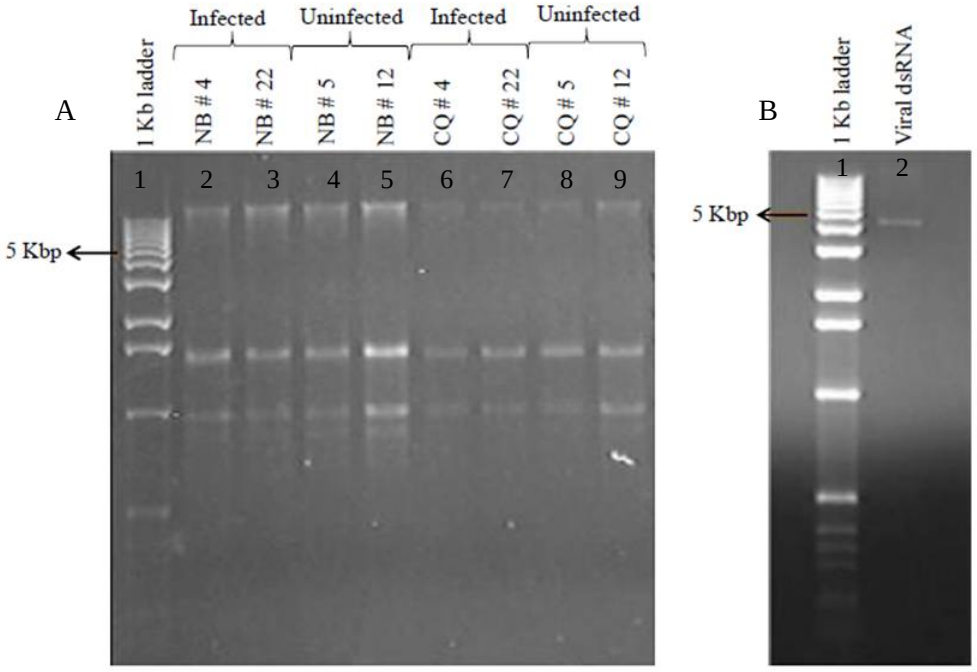

Figure 10: dsRNA extracted from the inoculated *N. benthamiana* (NB), *C. quinoa* (CQ), and putative-virus infected *R. humilis*.
Lanes 1 of **(A)** and **(B)** contain the 1 kb DNA ladder (Invitrogen).
(A) Lanes 2, 3, 6 and 7 contain dsRNA from *N. benthamiana* (NB # 4, NB # 22) and *C. quinoa* (CQ # 4, CQ # 22) inoculated with the two putative virus-infected *R. humilis* (R4 and R22) leaf homogenates. Lanes 4, 5, 8 and 9 contain dsRNA from *N. benthamiana* (NB # 5, NB # 22) and *C. quinoa* (CQ # 5, CQ # 12) inoculated with two uninfected *R. humilis* (R5 and R12) leaf homogenates.
(B) Lane 2 shows the dsRNA extracted during the earlier study from RAV1-infected *R. humilis*.

shown in Fig. 10A, no dsRNA band indicative of a totivirus RNA (~ 5 kbp) could be observed for any of the inoculated plants, suggesting a lack of mechanical transmission of the virus. Fig 10B shows a control dsRNA band, extracted from the RAV1 infected *R. humilis*, processed during the discovery of the virus. Before the screening of the totivirus positive plants, attempts for mechanical transmission were made with all 3 plant samples on *N. benthamiana* and *C. quinoa*. The inoculated plants did not show symptom expression, however inoculated plants were not processed for dsRNA diagnosis.

Localization of the viral RNA in the stem sections of infected plants by *in situ* hybridization

The uninfected (R5) and the two virus-infected (R4 and R22) plant stems were fixed, embedded and sectioned. The cross-sections were hybridized with the dig-labeled minus-strand RNA probe followed by immuno-histochemical staining. The distribution of the plus-strand viral RNA in the stem sections was observed under a light microscope. Positively infected cells exhibited a dark brown/purple reaction product. In infected stem sections, the viral RNA was clearly detected in most of the tissue cells including pith, cortical and epidermal cells (Fig. 11A-11F). In addition, the cambial zone was one of the strongly hybridized regions, but it was difficult to see the individual vascular cambium cells due to these cells growing intrusively into each other (Fig. 11B and 11E). Figure 12 shows a magnified view of portions shown in Fig. 11, focusing on pith, parenchyma, collenchyma and epidermal cells. As shown, the viral RNA appeared to localize inside these cells, indicating virus presence inside the plant cells (Fig. 12A-12H). The two virus-infected sample (R4 and R22) sections differed in signal intensity of hybridization, most likely reflecting a difference in the strength of virus infection in these samples. Support for this explanation came from the results of sequencing randomly amplified DNAs synthesized from the total RNA of these samples. The ratio of the number of viral reads to the total number of reads

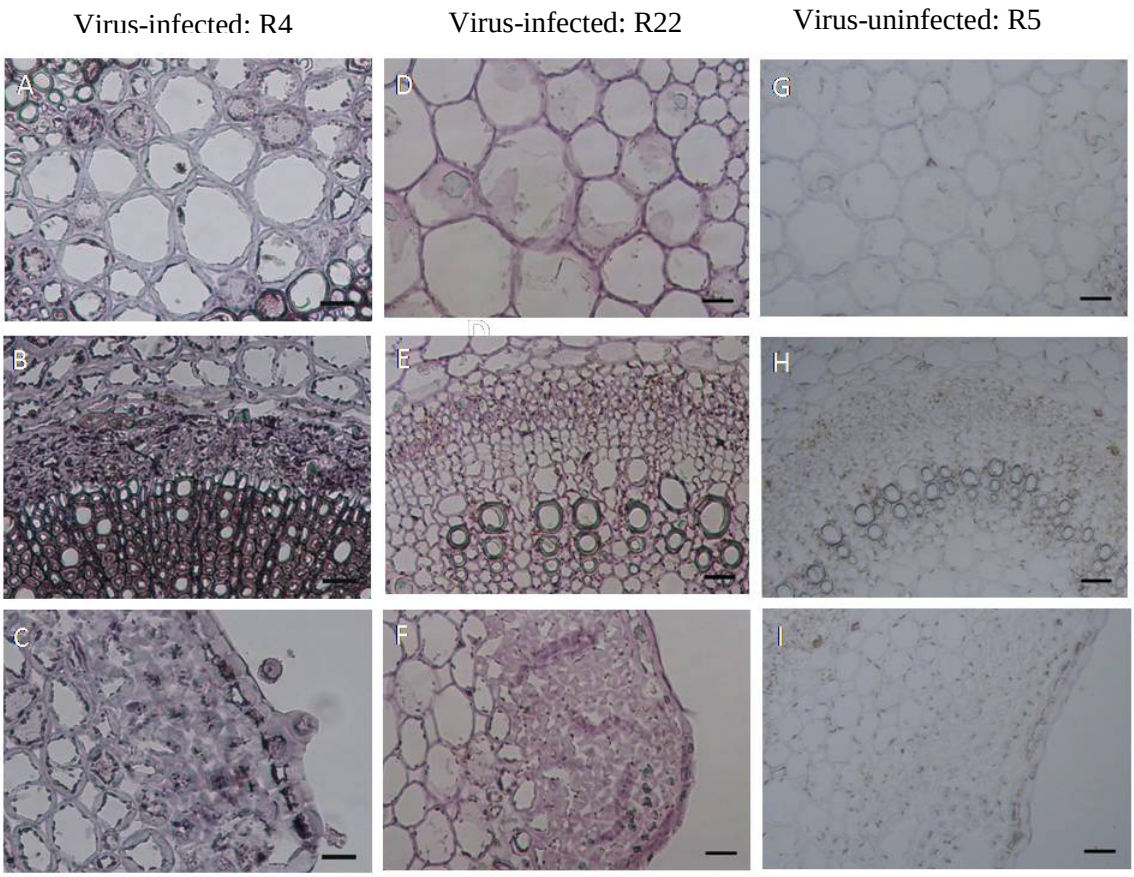

Figure 11: *In situ* hybridization for localization of plus-strand viral RNA in the stems of infected *R. humilis*. Positive cells exhibit a dark brown/purple reaction product.
(A), (D) and **(G)** Portion of the stem cross-section containing pith cells from the virus-infected [**(A)** and **(D)**] and uninfected samples **(G)**.
(B), (E) and **(H)** Portion of the stem cross-section containing vascular cells and cambial zone from the virus-infected [**(B)** and **(E)**] and uninfected samples **(H)**.
(C), (F) and **(I)** Portion of the stem cross-section containing cortical and epidermal cells from the virus-infected [**(C)** and **(F)**] and uninfected samples **(I)**. Bars = 0.03 mm.

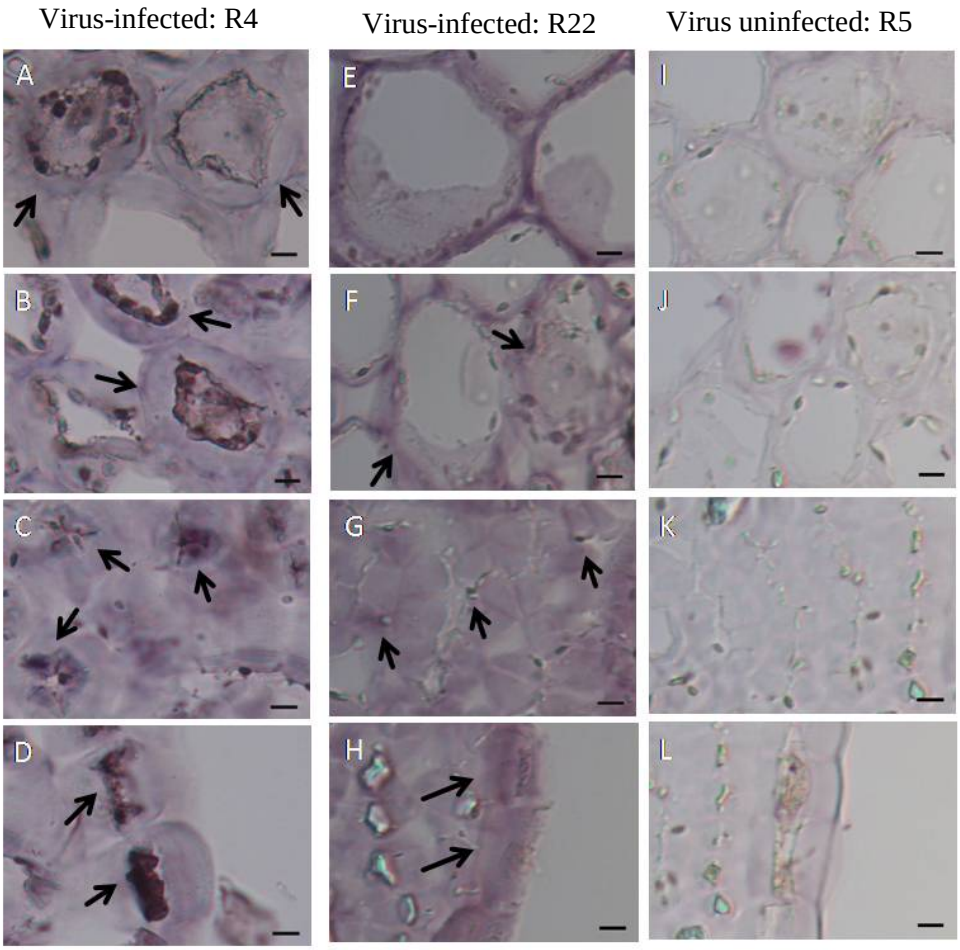

Figure 12: Closer view of the hybridized and stained sections to determine the site of viral RNA accumulation inside the infected plant stem tissues. The figure shows magnified view of the portions shown in figure 11. Positive cells (shown with arrows) exhibit a dark brown/purple reaction product.
(A), (E) and **(I)** Portion of the stem cross-section containing pith cells from the virus-infected [**(A)** and **(E)**] and uninfected samples **(I)**.
(B), (F) and **(J)** Portion of the stem cross-section containing parenchyma cells from the virus-infected [**(B)** and **(F)**] and uninfected samples **(J)**.
(C), (G) and **(K)** Portion of the stem cross-section containing collenchyma cells from the virus-infected [**(C)** and **(G)**] and uninfected samples **(K)**.
(D), (H) and **(L)** Portion of the stem cross-section containing epidermal cells from the virus-infected [**(D)** and **(H)**] and uninfected samples **(L)**. Bars = 0.01 mm

obtained for each sample was eight fold more for the sample showing stronger hybridization signal (R4) than for the sample with weaker signal (R22), suggesting a lower viral titer for the latter. In contrast to the infected samples, there was no such positive hybridization signal (except a minimal background) for viral RNA in non-infected stem sections (Fig. 11G-11I and 12I-12L). The results show localization of the viral RNA inside the plant cells, suggesting its identification as a plant virus.

Localization of fungal endophytes in the stem sections using cotton blue staining

Virus-free *R. humilis* seedlings were colonized with a fungal endophyte, *Curvularia protuberata*, at the two-leaf seedling stage. Successful colonization of the fungal isolate was confirmed by observing a characteristic outgrowth of *C. protuberata* on PDA media by plating surface-sterilized leaves from colonized plants (Fig. 13). Leaves from non-colonized plants did not show any such fungal outgrowth on the media plate. To test for fungal endophyte visualization using cotton blue dye, the stems from colonized and non-colonized plants were fixed, embedded, sectioned, stained with cotton blue dye and observed under a light microscope. As seen in figure 14A and 14B, the colonized plant stem sections showed stained structures (shown with arrows) of the shape characteristic of fungal hyphae in the intercellular spaces, where *C. protuberata* should most likely be located. No such hyphal structures were found in the intercellular spaces of the non-colonized plant stem sections (Fig. 14C and 14D). These results suggested a reliable employment of cotton blue stain for detection of fungal endophytes in colonized *R. humilis* samples.

The serial sections of virus-infected stem (R4 and R22) sections, shown to contain the virus inside the cells, were later stained with cotton blue dye to localize any fungal endophytes present inside the virus-infected cells. Figures 15 and 16 compare portions of stem sections stained for virus and fungus localization for plants R4 and R22, respectively. The search for

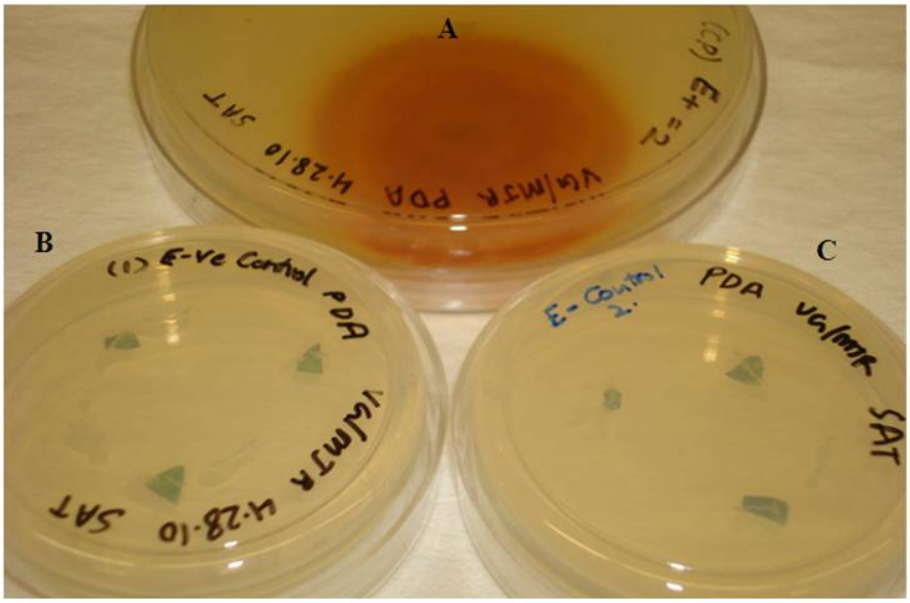

Figure 13: Leaves of fungal-colonized and non-colonized *R. humilis* plants plated on the PDA media. Successful fungal colonization of the plant is demonstrated in (A) by the characteristic outgrowth of *C. protuberata*, something which was not observed in non-colonized plants (B and C).

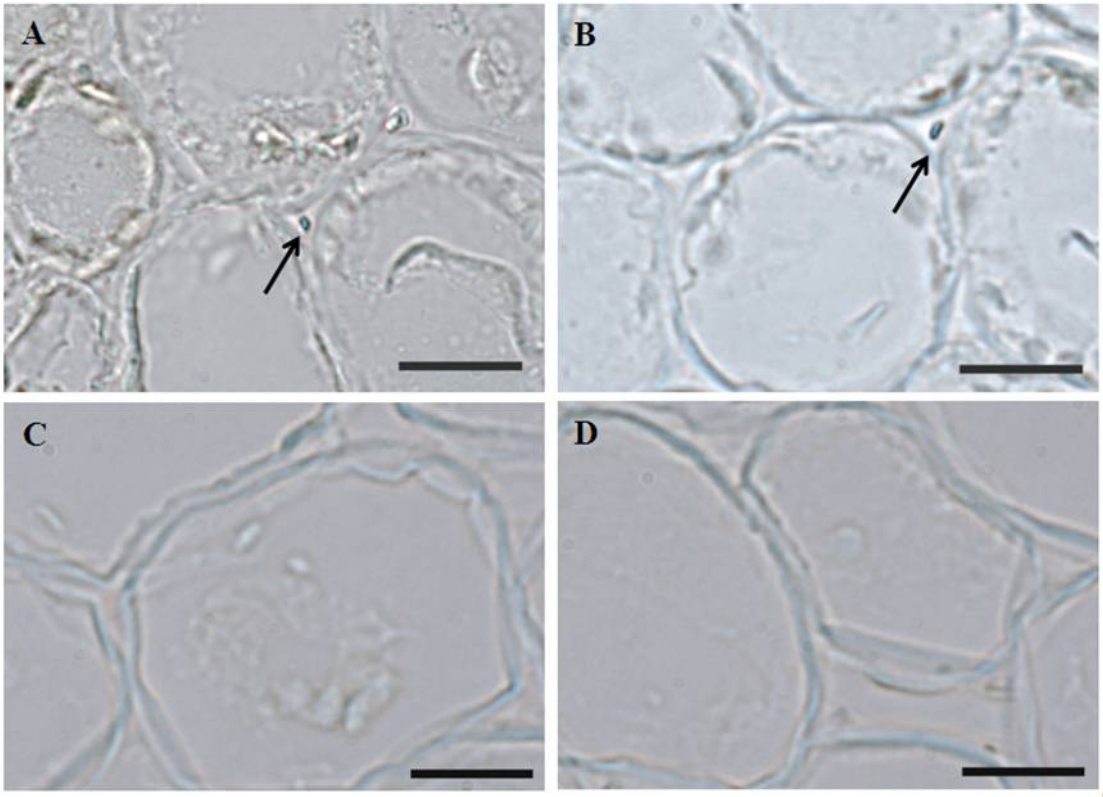

Figure 14: Localization of fungal endophytes in stem sections of colonized and non-colonized *R. humilis* using cotton blue staining.
(A) and **(B)** Portion of the stem cross-section from *C. protuberta* colonized *R. humilis* plant.
(C) and **(D)** Portion of the stem cross-section from a non-colonized *R. humilis* plant.
Arrows in **(A)** and **(B)** indicate stained structures characteristic of fungal hyphae observed in the intercellular spaces. Bars = 0.01 mm

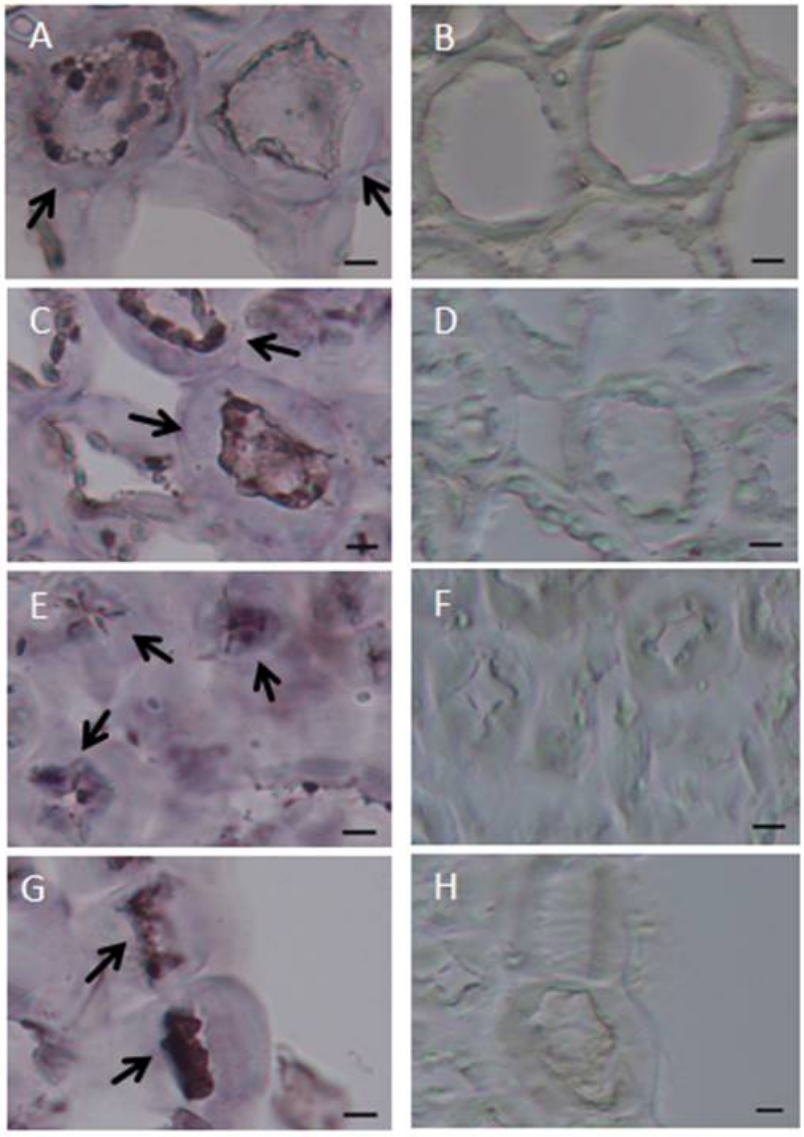

Figure 15: Localization of fungal endophytes by cotton blue staining in the serial stem sections of virus-infected plant R4, shown to contain the virus inside the stem tissue cells.
The micrographs on the right show portions of the section cut serially to the one shown on the left. **(A)**, **(C)**, **(E)** and **(G)** show portion of stem cross-sections containing pith, parenchyma, collenchyma and epidermal cells respectively, from the infected plant R4 after *in situ* hybridization. The same micrographs were shown in fig. 12A-12D.
(B), **(D)**, **(F)** and **(H)** show the cells in the same region, after cotton blue staining, as seen in the serial sections in micrographs on the left in order to look for fungal endophytes inside the virus-infected cells. Bars = 0.01 mm

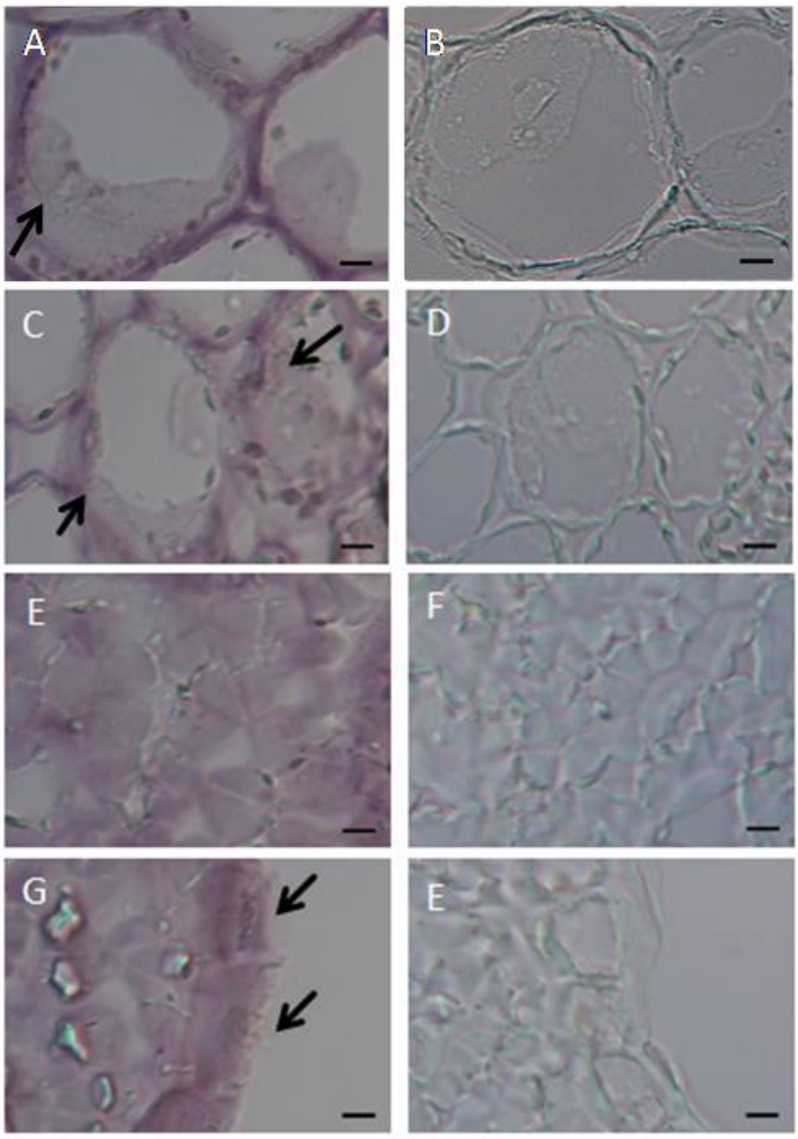

Figure 16: Localization of fungal endophytes by cotton blue staining in the serial stem sections of virus infected plant R22, shown to contain the virus inside the stem tissue cells.
The micrographs on the right show portions of the section cut serially to the one shown on the left. **(A)**, **(C)**, **(E)** and **(G)** show portion of stem cross-sections containing pith, parenchyma, collenchyma and epidermal cells respectively, from the infected plant R22 after *in situ* hybridization. The same micrographs were shown in fig. 12E-12H.
(B), **(D)**, **(F)** and **(H)** show the cells in the same region, after cotton blue staining, as seen in the serial sections in micrographs on the left in order to look for fungal endophytes inside the virus-infected cells. Bars = 0.01 mm

fungal endophytes found no structures inside the virus-infected cells characteristic of or similar to fungal hyphae. Thus, the results did not support an association of the virus with a fungus in these samples.

Isolation of fungal endophytes from stem tissues of infected plants followed by dsRNA extraction

Although no virus-fungus association was observed by cotton blue staining of serial sections of the virus-infected stem sections, the likelihood of interaction of the virus with a fungus could not be ruled out. In case of possible virus transmission between a fungus and the plant, fungal endophytes in other parts of the plant stem might harbor the virus. In order to test that, fungal isolates were cultured from different parts of the infected plants' stems on PDA plates, and further grown in liquid PD medium supplemented with antibiotics. Mycelium mats were filtered, lyophilized and processed for dsRNA extraction. Figures 17A and 17B show the dsRNA profiles of fungal isolates from the two virus-infected samples. A total of eight and three morphologically different fungal isolates could be isolated from plants R4 and R22, respectively. No dsRNA band indicative of a totivirus RNA (~ 5 kbp) could be observed from any of these fungal isolates. The results obtained here did not provide evidence for virus-fungus interaction but there still exists a possibility of the virus being harbored by endophytic fungi, present in parts of the plants that could not be tested during the analysis. It is possible that the cultured fungal isolates became cured of the virus, and thus did not show viral dsRNA presence.

Seed transmission analysis

The uninfected and the two virus-infected *R. humilis* plants were grown until fruiting, and seeds were harvested, surface sterilized, vernalized, germinated and grown in the greenhouse for a month. Surface sterilization of the seeds was performed before germination to eliminate the

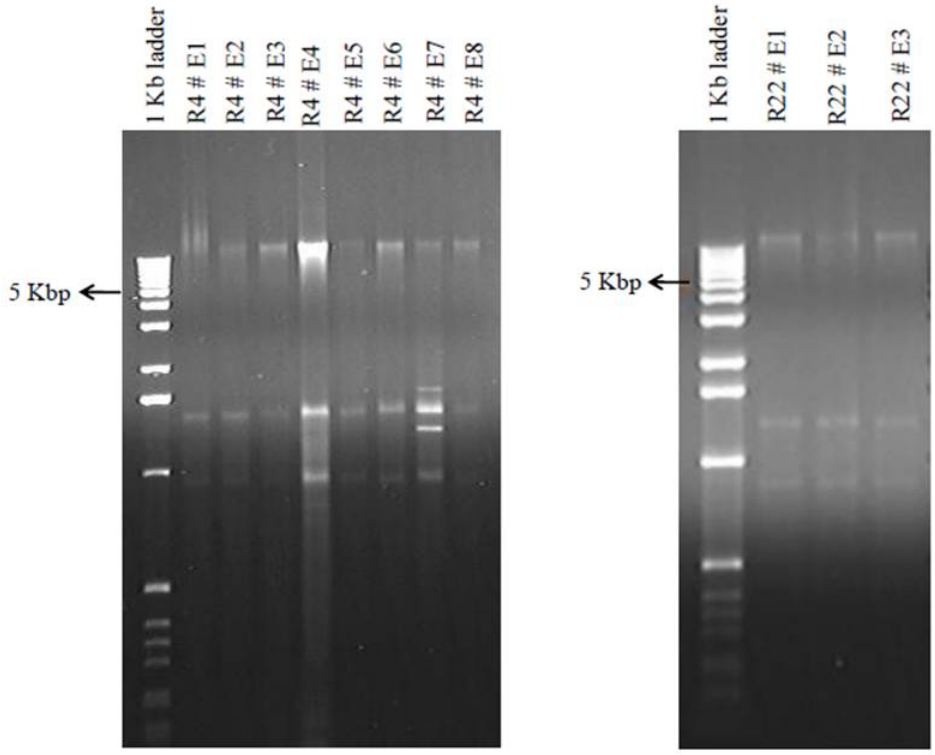

Figure 17: dsRNA extracted from fungal isolates of the two putative totivirus-infected *R. humilis* plants. **(A)** and **(B)** The first lane contains the 1 kb DNA ladder (Invitrogen).
(A) dsRNAs from eight different fungal isolates, R4 # E1 to R4 # E8, extracted from the virus-infected plant R4. **(B)** dsRNAs from three different fungal isolates R22 # E1 to R22 # E3, extracted from the virus-infected plant R22.

possibility of contamination due to a virus present on the outer surface of the seed. Seed transmission of the virus was evaluated by performing RT-PCR on the total RNA extracted from the test samples. A total of sixteen progenies (Pg) (nine and seven from the virus-infected parent plant R4 and R22, respectively) were tested. The test confirmed the presence of the putative virus in all sixteen progenies (PgR4-1 to PgR4-9 and PgR22-1 to PgR22-7). Figure 18 shows as an example, six of the sixteen tested progenies, three virus-infected positive control (PC-1, PC-2 and PC-3) and three uninfected progenies (NC-1, NC-2 and NC-3). The chosen positive controls were *R. humilis* samples collected from the TGPP during the earlier study, and were shown to be virus-positive through sequencing, while the three negative controls (NC) were progenies of the uninfected *R. humilis* (R5). The results provide evidence for 100% vertical transmission efficiency of the virus, indicating persistence of the virus throughout the life time of host.

Determination of the 3'-end of the viral genome

An almost complete sequence missing only the ends of the dsRNA genome of this putative virus had already been assembled. The 3'-end of the genome was retrieved using 5' RACE (Rapid amplification of cDNA ends) on the dsRNA template extracted from the WAPV1 infected plant. The dsRNA was reverse transcribed using a gene-specific forward primer followed by tailing of the cDNA with dGTP and generation of dsDNA by performing two rounds of PCRs using an oligo dC-adapter primer and gene-specific primer, and just the adapter sequence primer and the gene-specific primer. The amplified product of ~ 400 bp was cut out of the gel and cloned using pGEM-T Easy vector. The presence of the insert was confirmed by performing PCR on white colonies using the adapter sequence primer and the gene-specific primer. The amplified PCR products when sequenced with the gene-specific forward primer produced a sequence with the primer sequence at the 5'-end and the oligo G tail, added during the tailing step, in the 3'-end. The sequence in between the gene-specific primer and the oligo G tail consisted of a 43 nt long undetermined 3'-end of the molecule though the start of the G tail could not precisely be

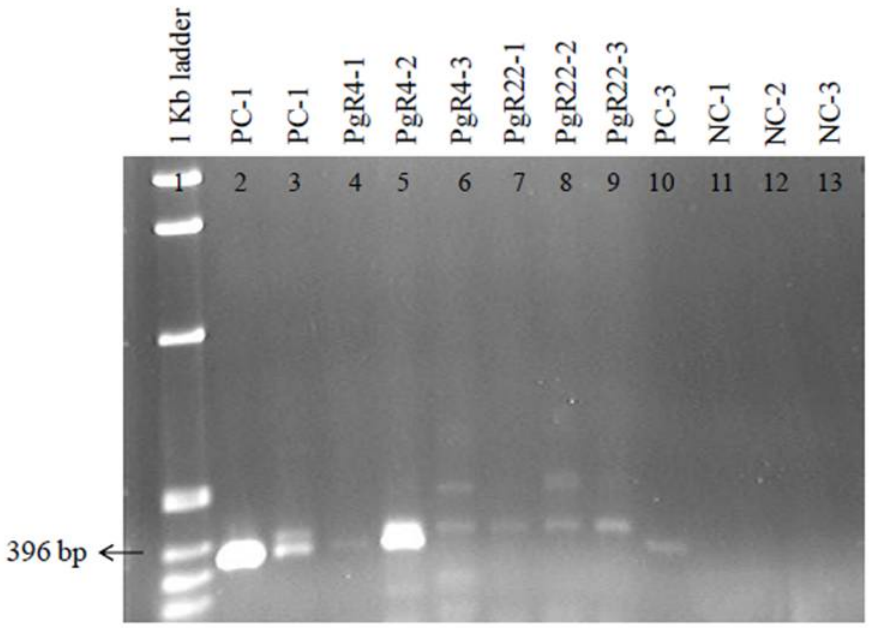

Figure 18: Seed transmission analysis for the putative totivirus. Reverse-transcription polymerase chain reaction (RT-PCR) identification of the putative totivirus RNA in leaf tissues collected from progenies of the virus-infected *R. humilis*. Lane1 contains the 1Kb ladder (Invitrogen). Lanes 2, 3 and 10 represent positive controls, and show RT-PCR amplification on RNA extracted from three virus-infected *R. humilis* plants. Lanes 4, 5, 6 and 7, 8, 9 show RT-PCR amplification on RNA extracted from three progenies (Pg), each of the virus- infected parent plant R4 (PgR4-1, PgR4-2 and PgR4-3) and R22 (PgR22-1, PgR22-2 and PgR22-3), respectively. Lanes 11, 12 and 13 represent three negative controls (NC1, NC2 and NC3), and show RT-PCR amplification on RNA extracted from three progenies of an uninfected *R. humilis*.

determined. Similar attempts to retrieve the precise 5'-end of the dsRNA genome failed. The available nucleotide sequence data for the RAV1 will soon be deposited in the GenBank database.

DISCUSSION

In the course of examining non-cultivated plants for signatures of viruses, a putative virus belonging to family *Totiviridae* was discovered. The sequence information retrieved for the virus showed similarity with Black raspberry virus F, a dsRNA totivirus. The study reported here on this putative virus, designated as *Ruellia* asymptomatic virus 1, was divided into two main objectives. The first objective was to determine the host of the virus, and its characterization as a fungal or a plant virus or one that can be transmitted between plants and fungi. Since the results of the first part of the study showed that the virus is a plant virus, the second objective was to test whether the virus is a persistent plant virus. The results of seed transmission tests supported persistence of the virus in its host plants.

The hypothesis concerning the first objective of the study was that the dsRNA detected in the infected samples represents the genomic dsRNA of a plant virus. This assumption was based on the observation that the apparent viral titer in terms of sequence reads obtained from individual plant samples was much higher than expected for a mycovirus in plants. In this study, an *in situ* hybridization approach was employed for determining the site of the viral RNA localization inside infected plant tissues. The plus sense viral RNA was detected in the stem tissues of infected plants using a dig-labeled RNA probe complementary to a region within the viral coat protein gene. The hybridization results for infected samples demonstrated the accumulation of viral RNA inside pith, cortical and epidermal cells. A clear difference in the hybridization signal intensities of the two virus-infected samples occurred, most likely due to the difference in the viral titers of the two plants. The reproducible absence of false hybridization signal for the viral RNA in uninfected plant stem sections made the interpretation of results simple and reliable. Fungal endophyte localization using cotton blue stain was successfully tested using *C. protuberta* colonized (positive control) and non-colonized (negative control) *R. humilis* stem sections. In order to test the association of the virus with a fungus, attempts were made to

visualize fungal endophytes in the serial sections of the virus-infected stems. The virus-infected cells were observed to contain no traces of fungal endophytes. Although the members of the genus *Totivirus* reported so far have primarily been found to infect fungi, protists and arthropods, the results clearly suggest that the viral RNA represents the genome of a plant virus replicating inside plant cells. The data presented here supports earlier literature that suggested dsRNA viruses recovered from plants, with similarity to fungal viruses, as possible plant viruses (Martin *et al.*, 2006).

Many fungal viruses, including totiviruses maintain only the genes that are essential for their survival, but make efficient use of host proteins (Ghabrial, 1998; Huang & Ghabrial, 1996). The virus in this study did not encode for any obvious putative movement protein. The movement protein gene of a plant virus could be employed or acquired from its plant host genes that encode proteins similar in function to viral movement proteins, assisting in RNA movement between the plant cells. Using similarity to plant movement proteins, the Lucas group (Lucas *et al.*, 1995) found an endogenous plant protein (KNOTTED 1) that helps transport mRNAs from cell to cell through plasmodesmata much like a viral movement protein (MP), suggesting origin of viral MP genes from plant host genes (Roossinck, 1997). This seems possible considering the extremely low level of sequence similarity in the movement proteins of viruses (Melcher, 2000) even with highly similar RdRp genes, suggesting that viral MPs have been acquired more than once and have diverse origins. This can also make us overlook a plausible movement protein due to its atypical MP gene sequence. Viruses may also use a helper virus, if available, or a viral coat protein for assistance in movement functions. However, persistent viruses in plants are primarily vertically transmitted, spreading throughout the plant only by cell division due to the lack of movement protein. These viruses could represent viruses which have either lost their movement functions or are precursors to new plant viruses before they attained their movement protein.

Hence, the second hypothesis of the study was that the virus is a persistent plant virus. This assumption was based on two typical plant persistent virus traits revealed by RAV1: 1) as

stated above, there was no obvious putative movement protein gene encoded by the dsRNA genome proposing the spread of the virus throughout the plant by cell division and 2) the plant samples shown to contain the viral dsRNA did not demonstrate any obvious disease symptoms. The other distinctive features of a persistent plant virus include lack of horizontal transmission of the virus and persistence of virus-infection through seed transmission of the virus. Attempts made to mechanically transmit the putative virus failed. Seed transmission of the virus was tested by performing RT-PCR on the total RNA extracted from the progenies of virus-infected parent plants. The virus was found to successfully transmit through seeds to the offspring of the infected plants, suggesting virus persistence in the host plants. Although there were no disease symptoms observed on the infected parent plants, the progenies of the presumably high titre virus-infected plant were significantly shorter than the uninfected plant progenies, while no such difference in height was observed in the progenies of the other infected parent plant. Due to the small sample size of the infected parent source, it is difficult to correlate the observed phenotypic effect with the presence of the virus. The effect could as well be due to genetic differences between the parental plants irrespective of the viral infection.

Some plant viruses are known to be related to fungal viruses, most notably the plant cryptic viruses, one of the most well studied persistent plant viruses. As observed in the case of RAV1, showing highest similarity to a fungal virus, some plant cryptic viruses also show higher similarity to fungal viruses than to other plant viruses of the same family (Roossinck, 2010; Strauss *et al.*, 2000; Veliceasa *et al.*, 2006). Because of such similarities, several authors have proposed the origin of plant cryptic viruses from fungal viruses or vice versa (Ghabrial, 1998; Osaki *et al.*, 2004; Roossinck, 1997, 2010), which means transfer of viruses across kingdom boundaries, and their subsequent adaptation. Though there are many known viruses that can propagate in both plant and animal hosts (e.g., members of the families *Reoviridae, Bunyaviridae* and *Rhabdoviridae)*, no viruses are known yet that can replicate in both fungi and another host in a different kingdom. This led us to wonder about the possible transmission of the putative

totivirus between the plant and an endophytic fungus, which could possibly act as a vector for the virus. In other words, it is possible that the virus might be spreading in plants, perhaps through fungal endophytes, a process which has been hypothesized but not observed before. This could also explain the spread of the virus in several different plant families since class II fungal endophytes are known to have a broad host range. In order to test the mycoviral character of the virus, fungal endophytes were isolated from different parts of virus-infected plants' stems for characterization of any dsRNAs (i.e. viruses) that they harbor. It was assumed that if fungal endophytes in the infected plants carry the virus, we should be able to culture the virus containing endophyte from these plants and isolate the virus from the fungus. Though the lack of isolation of dsRNA-containing fungus from infected plants failed to show an evident virus-fungus interaction, it does not rule out the possibility of the virus being associated with a fungus. It is also possible that the virus may have been transmitted sometime in the past. Hence, it might well be a fungal virus that has escaped its normal host during an endophytic association with the plant.

In conclusion, the study provides evidence for the first plant totivirus, discovered during a virus biodiversity survey of wild plants, representing a new group of viruses in the family *Totiviridae*. This putative plant totivirus has features typical of a persistent virus: the asymptomatic nature, the lack of horizontal transmission and the seed transmissibility of the virus. There are many questions that remain unanswered and will require a more detailed investigation in the subject. These include a better understanding of the origin of these plant persistent viruses, evolutionary relationships between fungal and plant viruses, exploration of possible inter-kingdom transmission of fungal viruses and plant persistent viruses, and the possible roles of these persistent plant viruses in our ecosystem. The depth of knowledge gained from searches of wild asymptomatic plants should pave the way for further investigations on such issues, in addition to expanding our understanding of the plant virus biodiversity.

CONCLUDING REMARKS

We live in uncertain times. Bioterrorism and biocrimes have become a concern. Transportation of people and goods has accelerated, leading to increased chances of cross-border movement of viruses and other microbes. Plant biosecurity has never been more important worldwide than it is today, as the potential employment of plant pathogens as agents of bioterrorism threatens food security (Rodoni, 2009). Plant biosecurity efforts are aimed at preventing the importation of undesirable microbes and viruses along with plant importations (Rodoni, 2009). Plant virus outbreaks can heavily affect both yield and quality of agricultural products, leading to significant economic damage and reduced public health. Prior surveys of virus biodiversity in the neighborhood of agricultural production fields will allow investigators to decide whether the infection is natural or possibly man-made. Wild plants constitute a reservoir of viruses which are generally considered not to harm their hosts but have indeed been shown, under some circumstances, to possess mutalistic relationships with their hosts. However, emerging infectious diseases of wild plants have long been recognized as a possible threat to economically important species, acting as a potential source of disease outbreaks. With global climate change, agricultural land use patterns can change, supplanting crops in one location with crops from another region. The new crops will be subjected to infection by a range of new (to the crop) indigenous viruses, or may spread their own viruses to indigenous plants (Webster *et al.*, 2007). Thus, information about the diversity, distribution and impact of plant viruses in natural plant communities is thus needed as the first step towards the design of effective disease management strategies. Exploitation of available genetic information of viruses in wild plant populations can also aid in preliminary risk assessment of virus-mediated gene flow from genetically modified organisms

and of recombination between virus-derived transgenes and endogenous viruses, prior to field release of transgenic crops.

Genomic approaches discussed in this thesis serve as indispensable tools for biodiversity surveys for signatures of plant viruses, and for rapid detection of viruses in case of viral outbreaks since they notably decrease the lag times between infection, detection and response. A remarkable example of this was demonstrated during the recent SARS outbreak, where the microarray hybridization patterns observed using the pan-viral microarray helped identify SARS as a novel coronavirus, demonstrating the power and utility of this approach. The microarray approach developed in the present work was based on the same concept that an oligonucleotide array could be created to detect and differentiate many plant viruses using sequence-independent amplification of target samples. A logical extension of the technology will be the design, development and validation of a comprehensive plant virus microarray for identifying previously described and discovering novel plant viruses. Ongoing research in Ulrich Melcher's laboratory includes a collaboration project to develop a universal plant virus microarray, consisting of oligonucleotide probes directed separately to sequences highly conserved between members of each virus and viroid genus, and sequences specific to particular virus and viroid species to represent all known taxonomic groupings of plant viruses. The array will be analogous to the one designed by Wang *et al.* (Wang *et al.*, 2002; Wang *et al.*, 2003) and would be of critical value in case of unexplained viral outbreaks, and regular biodiversity surveys as a multi-viral detection system. Microarray and the other approaches, discussed here, being used to explore biodiversity in single plants through cDNA or siRNA have also direct application to a general test for contaminating organisms, whether viral or other. Thus, they may become a routine component of plant quarantine operations.

The times demand that we understand that viruses are often not pathogens in their natural settings. Recent work clearly demonstrates that viruses associated with plants should not be regarded exclusively as pathogens, always having negative effects on the fitness of their hosts.

Most of the viral sequences discovered in the PVBE project came from asymptomatic, healthy-looking plant samples, supporting the idea that, if there is no selective advantage for the virus to cause disease, there might be a selective advantage for it to make its host healthy (Powledge, 1999). The biodiversity and abundance of viruses found in wild plants suggest an important role of these viruses in our ecosystem. These roles should not be under-rated just because they are under-researched. Most of the plant virology studies concentrate on disease-causing acute or chronic viruses. One of the largely neglected but widely prevalent groups of plant viruses include persistent viruses, one of which has been reported and studied as a part of this thesis work. The virus has been demonstrated as the first plant virus in the family *Totiviridae*, and is also found to be seed transmissible in the host plant. As persisting asymptomatic viruses, they are difficult to identify and survey, yet it is clear that most species harbor persistent viral agents. Such persistent viral infections can have significant effects on relative reproductive fitness of competing host populations. The impact of persistent viruses in wild plant ecosystem dynamics is an emergent area of research with many questions still to be unveiled. Some of these questions include notions about many vital roles that can be played by these viruses in plants, by establishing mutualistic symbioses, acting as epigenetic elements providing novel genes, or as sources for newly emerging viruses. Investigation of these possible functions and their mechanisms will require a thorough understanding of virus biodiversity and the influence of viruses on plant communities and, on the other hand, of plant community diversity on virus success. One of the foremost questions to study would be the likelihood of inter-kingdom transmission of fungal viruses and plant persistent viruses. In the past, the lack of information about both plant persistent viruses and fungal viruses had made it difficult to perform a comprehensive sequence data analysis. However, with the availability of more and more genomic data for these viruses, a thorough phylogenetic analysis can clarify the ecology of these viruses and their likely transmission in the recent or distant past. In conclusion, investigation of viruses in the wild plant population can serve as a

launching point for additional studies focused on demonstrating disease causality, and dissecting the possible mutualistic roles of plant viruses in natural settings.

BIBLIOGRAPHY

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M. & Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular Plant Pathology* **10**, 537-545.

Agindotan, B. & Perry, K. L. (2007). Macroarray detection of plant RNA viruses using randomly primed and amplified complementary DNAs from infected plants. *Phytopathology* **97**, 119-127.

Agindotan, B. & Perry, K. L. (2008). Macroarray detection of eleven potato-infecting viruses and Potato spindle tuber viroid. *Plant Disease* **92**, 730-740.

Al Rwahnih, M., Daubert, S., Golino, D. & Rowhani, A. (2009). Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* **387**, 395-401.

Alabi, O. J., Kumar, P. L. & Naidu, R. A. (2008). Multiplex PCR for the detection of *African cassava mosaic virus* and *East African cassava mosaic Cameroon virus* in cassava. *Journal of Virological Methods* **154**, 111-120.

Alioto, D., Zaccaria, F., Covelli, L., Serio, F. D., Ragozzino, A. & Milne, R. G. (2003). Light and electron microscope observations on chlorotic rusty spot, a disorder of cherry in Italy *Journal of Plant Pathology* **85**, 215-218.

Alkowni, R., Rowhani, A., Daubert, S. & Golino, D. (2004). Partial characterization of a new ampelovirus associated with grapevine leafroll disease. *Journal of Plant Pathology* **86**, 123-133.

Ambrose, H. E. & Clewley, J. P. (2006). Virus discovery by sequence-independent genome amplification. *Reviews in Medical Virology* **16**, 365-383.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R. & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers *Trends in Ecology & Evolution* **19**, 535-544

Arancibia, R. A., Valverde, R. A. & Can, F. (1995). Properties of a cryptic virus from pepper (Capsicum annuum). *Plant Pathology* **44**, 164-168.

Armour, S. L., Melcher, U., Pirone, T. P., Lyttle, D. J. & Essenberg, R. C. (1983). Helper component for aphid transmission encoded by region II of cauliflower mosaic virus DNA. *Virology* **129**, 25-30.

Azam, F. & Worden, A. Z. (2004). Microbes, molecules, and marine ecosystems. *Science* **303**, 1622-1624.

Boben, J., Kramberger, P., Petrovicˇ, N. a., Cankar, K., Peterka, M., Strancar, A. & Ravnikar, M. (2007). Detection and quantification of *Tomato mosaic virus* in irrigation waters. *European Journal of Plant Pathology* **118**, 59-71.

Boccardo, G. & Candresse, T. (2005). Complete sequence of the RNA2 of an isolate of *White clover cryptic virus 1*, type species of the genus *Alphacryptovirus*. *Archives of Virology* **150**, 403-405.

Boccardo, G., Lisa, V., Luisoni, E. & Milne, R. G. (1987). Cryptic Plant Viruses. In *Advances in Virus Research*, vol. 32, pp. 171-214. Edited by F. A. M. Karl Maramorosch & J. S. Aaron: Academic Press.

Boccardo, G., Milne, R. G., Luisoni, E., Lisa, V. & Accotto, G. P. (1985). Three seedborne cryptic viruses containing double stranded RNA isolated from white clover. *Virology* **147**, 29–40.

Bodaghi, S., Mathews, D. M. & Dodds, J. A. (2004). Natural incidence of mixed infections and experimental cross protection between two genotypes of *Tobacco mild green mosaic virus*. *Phytopathology* **94**, 1337-1341.

Bohlander, S. K., Espinosa, R., Le Beau, M. M., Rowley, J. D. & Díaz, M. O. (1992). A method for the rapid sequence-independent amplification of microdissected chromosomal material. *Genomics* **13**, 1322-1324.

Boonham, N., Tomlinson, J. & Mumford, R. (2007). Microarrays for rapid identification of plant viruses. *Annual Review of Phytopathology* **45**, 307-328.

Boonham, N., Walsh, K., Smith, P., Madagan, K., Graham, I. & Barker, I. (2003). Detection of potato viruses using microarray technology: towards a generic method for plant viral disease diagnosis. *Journal of Virological Methods* **108**, 181-187.

Bosque-Perez, N. A. (2000). Eight decades of *maize streak virus* research. *Virus Research* **71**, 107–121.

Bousalem, M., Durand, O., Scarcelli, N., Lebas, B. S. M., Kenyon, L., Marchand, J. L., Lefort, F. & Seal, S. E. (2009). Dilemmas caused by endogenous pararetroviruses regarding the taxonomy and diagnosis of yam (*Dioscorea* spp.) badnaviruses: analyses to support safe germplasm movement. *Archives of Virology* **154**, 297-314.

Branovic, K., Forcic, D., Ivancic, J., Strancar, A., Barut, M., Kosutic-Gulija, T., Zgorelec, R. & Mazuran, R. (2003). Application of short monolithic columns for improved detection of viruses. *Journal of Virological Methods* **110**, 163-171.

Breitbart, M. & Rohwer, F. (2005a). Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques* **39**, 729-736.

Breitbart, M. & Rohwer, F. (2005b). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology* **13**, 278-284.

Briese, T., Paweska, J. T., McMullan, L. K., Hutchison, S. K., Street, C., Palacios, G., Khristova, M. L., Weyer, J., Swanepoel, R. & other authors (2009). Genetic detection and characterization of Lujo Virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathogens* **5**, e1000455.

Brunt, A. A., Crabtree, K., Dallwitz, M. J., Gibbs, A. J., Watson, L. & Zurcher, E. J. (1996). Plant Viruses Online: Descriptions and Lists from the VIDE Database, Version: 20th August 1996 edn.

Buck, K. W. (1986). Fungal virology--an overview. In *Fungal Virology*, pp. 1-84. Edited by K. W. Buck. Boca Raton: CRC Press.

Bystricka, D., Lenz, O., Mraz, I., Piherova, L., Kmoch, S. & Sip, M. (2005). Oligonucleotide-based microarray: A new improvement in microarray detection of plant viruses. *Journal of Virological Methods* **128**, 176-182.

Canto, T., Aranda, M. A. & Fereres, A. (2009). Climate change effects on physiology and population processes of hosts and vectors that influence the spread of hemipteran-borne plant viruses. *Global Change Biology* **15**, 1884–1894.

Casagrande, R. (2000). Biological terrorism targeted at agriculture: The threat to US national security. *The Nonproliferation Review* **7**, 92-105.

Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M. & Moore, P. S. (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**, 1865-1869.

Chapman, N. M., Tracy, S., Gauntt, C. J. & Fortmueller, U. (1990). Molecular detection and identification of enteroviruses using enzymatic amplification and nucleic acid hybridization. *Journal of Clinical Microbiology* **28**, 843-850.

Chiu, C. Y., Urisman, A., Greenhow, T. L., Rouskin, S., Yagi, S., Schnurr, D., Wright, C., Drew, W. L., Wang, D. & other authors (2008). Utility of DNA microarrays for detection of viruses in acute respiratory tract infections in children. *The Journal of Pediatrics* **153**, 76-83.

Chou, C.-C., Chen, C.-H., Lee, T.-T. & Peck, K. (2004). Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Research* **32**, e99.

Ciuffo, M., Tavella, L., Pacifico, D., Masenga, V. & Turina, M. (2008). A member of a new *Tospovirus* species isolated in Italy from wild buckwheat (*Polygonum convolvulus*). *Archives of Virology* **153**, 2059-2068.

Coen, E. S., Romero, J., Doyle, S., Elliott, R., Murphy, G. & Carpenter, R. (1990). floricaula: A homeotic gene required for flower development in antirrhinum majus. *Cell* **63**, 1311-1322.

Covelli, L., Coutts, R. H. A., Serio, F. D., Citir, A., Acikgoz, S., Hernandez, C., Ragozzino, A. & Flores, R. (2004). Cherry chlorotic rusty spot and Amasya cherry diseases are associated with a complex pattern of mycoviral-like double-stranded RNAs. I. Characterization of a new species in the genus *Chrysovirus*. *Journal of General Virology* **85**, 3389-3397.

Cox, S., Mayo, M. A. & Jones, A. T. (2000). The occurrence of dsRNA species in apparently healthy and virus-infected Ribes cultivars, and evidence that one such species originates from a member of the virus family *Totiviridae*. *European Journal of Plant Pathology* **106**, 353-364.

Dale, J. & Gibbs, A. (1976). *Kennedya* yellow mosaic virus: another tymovirus. *Australian Journal of Biological Sciences* **29** 397–403.

Davalos, R. V., McGraw, G. J., Wallow, T. I., Morales, A. M., Krafcik, K. L., Fintschenko, Y., Cummings, E. B. & Simmons, B. A. (2008). Performance impact of dynamic surface coatings on polymeric insulator-based dielectrophoretic particle separators. *Analytical and Bioanalytical Chemistry* **390**, 847-855.

Department of Crop Sciences, U. o. I. (1991). In *Report on Plant Disease*.

Deyong, Z., Willingmann, P., Heinze, C., Adam, G., Pfunder, M., Frey, B. & Frey, J. E. (2005). Differentiation of Cucumber mosaic virus isolates by hybridization to oligonucleotides in a microarray format. *Journal of Virological Methods* **123**, 101-108.

Ding, S.-W. & Voinnet, O. (2007). Antiviral immunity directed by small RNAs. *Cell* **130**, 413-426.

Ding, X. S., Carter, S. A. & Nelson, R. S. (1996). Enhanced cytochemical detection of viral proteins and RNAs using double-sided labeling and light microscopy. *Biotechniques* **20**, 111-115.

Djikeng, A., Kuzmickas, R., Anderson, N. G. & Spiro, D. J. (2009). Metagenomic analysis of RNA viruses in a fresh water lake. *PloS One* **4**, e7264.

Djikeng, A., Halpin, R., Kuzmickas, R., Depasse, J., Feldblyum, J., Sengamalay, N., Afonso, C., Zhang, X., Anderson, N. G. & other authors (2008). Viral genome sequencing by random priming methods. *BMC Genomics* **9**, 5.

Dodds, J. A., Morris, T. J. & Jordan, R. L. (1984). Plant viral double-stranded RNA. *Annual Review of Phytopathology* **22**, 151-168.

Donaire, L., Wang, Y., Gonzalez-Ibeas, D., Mayer, K. F., Aranda, M. A. & Llave, C. (2009). Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology* **392**, 203-214.

Donehower, L. A., Bohannon, R. C., Ford, R. J. & Gibbs, R. A. (1990). The use of primers from highly conserved pol regions to identify uncharacterized retroviruses by the polymerase chain reaction. *Journal of Virological Methods* **28**, 33-46.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P. & other authors (2009). Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138.

Emrich, S. J., Lowe, M. & Delcher, A. L. (2003). PROBEmer: a web-based software tool for selecting optimal DNA oligos. *Nucleic Acids Research* **31**, 3746-3750.

Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S. & Hayashi, M. (2005). Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. *Nucleic Acids Research* **33**, e65.

Engel, E. A., Escobar, P. F., Rojas, L. A., Rivera, P. A., Fiore, N. & Valenzuela, P. D. T. (2010). A diagnostic oligonucleotide microarray for simultaneous detection of grapevine viruses. *Journal of Virological Methods* **163**, 445-451.

Fargette, D., Konate, G., Fauquet, C., Muller, E., Peterschmitt, M. & Thresh, J. M. (2006). Molecular ecology and emergence of tropical plant viruses. *Annual Review of Phytopathology* **44**, 235-260.

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A., (editors) (2005). *Virus Taxonomy*. London: Elsevier.

Feuer, R., Boone, J. D., Netski, D., Morzunov, S. P. & St. Jeor, S. C. (1999). Temporal and spatial analysis of Sin nombre virus quasispecies in naturally infected rodents. *Journal of Virology* **73**, 9544-9554.

Filée, J., Forterre, P. & Laurent, J. (2003). The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Research in Microbiology* **154**, 237-243.

Foissac, X., Svanella-Dumas, L., Gentit, P., Dulucq, M., Marais, A. & Candresse, T. (2005). Polyvalent degenerate oligonucleotides reverse transcription-polymerase chain reaction: a polyvalent detection and characterization tool for trichoviruses, capilloviruses, and foveaviruses. *Phytopathology* **95**, 617-625.

Fraile, A., Escriu, F., Aranda, M. A., Malpica, J. M., Gibbs, A. J. & Garcia-Arenal, F. (1997). A century of tobamovirus evolution in an Australian population of Nicotiana glauca. *Journal of Virology* **71**, 8316-8320.

Friess, N. & Maillet, J. (1996). Influence of cucumber mosaic virus infection on the intraspecific competitive ability and fitness of purslane (*Portulaca oleracea*). *New Phytologist* **132**, 103-111.

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541-548.

Fuller, C. W., Middendorf, L. R., Benner, S. A., Church, G. M., Harris, T., Huang, X., Jovanovich, S. B., Nelson, J. R., Schloss, J. A. & other authors (2009). The challenges of sequencing by synthesis. *Nature Biotechnology* **27**, 1013-1023.

Garrett, K. A., Dendy, S. P., Frank, E. E., Rouse, M. N. & Travers, S. E. (2006). Climate change effects on plant disease: Genomes to ecosystems. *Annual Review of Phytopathology* **44**, 489-509.

Ghabrial, S. A. (1998). Origin, adaptation and evolutionary pathways of fungal viruses. *Virus Genes* **16**, 119-131.

Ghabrial, S. A. & Suzuki, N. (2009). Viruses of plant pathogenic fungi. *Annual Review of Phytopathology* **47**, 353-384.

Gibbs, A. (1980). A plant virus that partially protects its wild legume host against herbivores. *Intervirology* **13**, 42-47.

Gibbs, A. (1987). Molecular evolution of viruses; 'trees', 'clocks' and 'modules'. *Journal of Cell Science - Supplement* **7**, 319-337.

Gibbs, A. J., Keese, P. L., Gibbs, M. J. & Garcia-Arenal, F. (1999). Plant virus evolution: past, present and future. In *Origin and Evolution of Viruses*, pp. 263-285. Edited by E. Domingo, R. Webster & J. Holland. San Diego: Academic Press.

Gibbs, A. J., Mackenzie, A. M., Wei, K. J. & Gibbs, M. J. (2008a). The potyviruses of Australia. *Archives of virology* **153**, 1411-1420.

Gibbs, A. J., Ohshima, K., Phillips, M. J. & Gibbs, M. J. (2008b). The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PLoS ONE* **3**, e2523.

Gibbs, M. J., Koga, R., Moriyama, H., Pfeiffer, P. & Fukuhara, T. (2000). Phylogenetic analysis of some large double-stranded RNA replicons from plants suggests they evolved from a defective single-stranded RNA virus. *Journal of General Virology* **81**, 227-233.

Grover, V., Pierce, M. L. & Melcher, U. (2007). Microarray hybridization for detection of plant viruses from natural settings. *Phytopathology* **97**, S43.

Grover, V., Pierce, M. L., Hoyt, P., Zhang, F. & Melcher, U. (2010). Oligonucleotide-based microarray for detection of plant viruses employing sequence-independent amplification of targets. *Journal of Virological Methods* **163**, 57-67.

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**, 25-36.

Hackett, J., Muthukumar, V., Wiley, G. B., Roe, B. A. & Melcher, U. (2009). Viruses in Oklahoma Euphorbia marginata. *Proceedings of the Oklahoma Academy of Sciences* **89**, 57-62.

Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., Dimeo, J. & other authors (2008). Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106-109.

Harrison, B. & Robinson, D. (1982). Genome reconstitution and nucleic acid hybridization as methods of identifying particle-deficient isolates of tobacco rattle virus in potato plants with stem-mottle disease. *Journal of Virological Methods* **5**, 255-265.

Harrison, B. D., Robinson, D. J., Mowat, W. P. & Duncan, G. H. (1983). Comparison of nucleic acid hybridisation and other tests for detecting tobacco rattle virus in narcissus plants and potato tubers. *Annals of Applied Biology* **102**, 331-338.

Harrison, B. E. (1981). Plant virus ecology: ingredients, interactions and environmental influences. *Annals of Applied Biology* **99**, 195-209.

Hassan, M., Sirlova, L., Jokes, M. & Vacke, J. (2009). Identification and characterization of a novel *Tritimovirus* species isolated from wild *Trisetum flavescens* L., family *Poaceae*. *Virus Genes* **39**, 146-152.

Ho, T., Rusholme Pilcher, R. L., Edwards, M.-L., Cooper, I., Dalmay, T. & Wang, H. (2008). Evidence for GC preference by monocot dicer-like proteins. *Biochemical and Biophysical Research Communications* **368**, 433-437.

Hohn, T., Richert-Pöggeler, K. R., Staginnus, C., Harper, G., Schwarzacher, T., Teo, C. H., Teycheney, P.-Y., Iskra-Caruana, M.-L. & Hull, R. (2008). Evolution of integrated plant viruses. In *Plant Virus Evolution*, pp. 53-82. Edited by M. J. Roossinck. Berlin: Springer.

Huang, S. & Ghabrial, S. A. (1996). Organization and expression of the double-stranded RNA genome of Helminthosporium victoriae 190S virus, a totivirus infecting a plant pathogenic filamentous fungus. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 12541-12546.

Hubert, C., Loy, A., Nickel, M., Arnosti, C., Baranyi, C., Bruchert, V., Ferdelman, T., Finster, K., Christensen, F. M. & other authors (2009). A constant flux of diverse thermophilic bacteria into the cold Arctic seabed. *Science* **325**, 1541-1544.

Ishihara, J., Pak, J. Y., Fukuhara, T. & Nitta, T. (1992). Association of particles that contain double-stranded RNAs with algal chloroplasts and mitochondria. *Planta* **187**, 475-482.

Jääskeläinen, A. J. & Maunula, L. (2006). Applicability of microarray technique for the detection of noro- and astroviruses. *Journal of Virological Methods* **136**, 210-216.

Jaspars, E. M. J. (1999). Single-stranded RNA viruses of plants reminiscent of early RNA endosymbionts. *Symbiosis* **26**, 327-345.

Jiu, M., Zhou, X.-P., Tong, L., Xu, J., Yang, X., Wan, F.-H. & Liu, S.-S. (2007). Vector-virus mutualism accelerates population increase of an invasive whitefly. *PLoS ONE* **2**, e182.

Kang, B.-C., Yeam, I. & Jahn, M. M. (2005). Genetics of plant virus resistance. *Annual Review of Phytopathology* **43**, 581-621.

Kassanis, B., White, R. F. & Woods, R. D. (1977). Beet cryptic virus. *Phytopathologische Zeitschrift* **90**, 350-360.

Kassanis, B., Russel, G. E. & White, R. F. (1978). Seed and pollen transmission of beet cryptic virus in sugar beet plants. *Phytopathologische Zeitschrift* **91**, 76-79.

Kawakami, K., Fuji, S. I. & Miyoshi, K. (2007). Endangered wild populations of endemic Calanthe orchids on an isolated Japanese island tested for viruses. *Australian Journal of Botany* **55**, 831-836.

Kimura, N., Oda, R., Inaki, Y. & Suzuki, O. (2004). Attachment of oligonucleotide probes to poly carbodiimide-coated glass for microarray applications. *Nucleic Acids Research* **32**, e68.

Kistler, A. L., Gancz, A., Clubb, S., Skewes-Cox, P., Fischer, K., Sorber, K., Chiu, C. Y., Lublin, A., Mechani, S. & other authors (2008). Recovery of divergent avian bornaviruses from cases of proventricular dilatation disease: identification of a candidate etiologic agent. *Virology Journal* **5**, 88.

Knierim, D. & Maiss, E. (2007). Application of Phi29 DNA polymerase in identification and full-length clone inoculation of tomato yellow leaf curl Thailand virus and tobacco leaf curl Thailand virus. *Archives of Virology* **152** 941-954.

Koonin, E. V., Senkevich, T. G. & Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biology Direct* **1**, 29.

Koonin, E. V., Wolf, Y. I., Nagasaki, K. & Dolja, V. V. (2008). The big bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nature Reviews Microbiology* **6**, 925-939.

Koonin, E. V., Choi, G. H., Nuss, D. L., Shapira, R. & Carrington, J. C. (1991). Evidence for common ancestry of a chestnut blight hypovirulence-associated double-stranded RNA and a group of positive-strand RNA plant viruses. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10647-10651.

Kozlakidis, Z., Brown, N., Jamal, A., Phoon, X. & Coutts, R. (2010). Incidence of endornaviruses in Phytophthora taxon douglasfir and Phytophthora ramorum. *Virus Genes* **40**, 130-134.

Kozlakidis, Z., Covelli, L., Di Serio, F., Citir, A., Acikgoz, S., Hernandez, C., Ragozzino, A., Flores, R. & Coutts, R. H. A. (2006). Molecular characterization of the largest mycoviral-like double-stranded RNAs associated with Amasya cherry disease, a disease of presumed fungal aetiology. *Journal of General Virology* **87**, 3113-3117.

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

Lane, L. C. (1986). Propagation and purification of RNA plant viruses. *Methods in Enzymology* **118**, 687-696.

Lane, L. C. (1992). A general method of detecting plant viruses. In *Plant diseases of viral, viroid, mycoplasma, and uncertain etiology*, pp. 1-15. Edited by K. Maramorosch. Boulder: Westview Press.

Lane, S., Evermann, J., Loge, F. & Call, D. (2004). Amplicon secondary structure prevents target hybridization to oligonucleotide microarrays. *Biosensors and Bioelectronics* **20**, 728-735.

Lartey, R. T., Voss, T. C. & Melcher, U. (1996). Tobamovirus evolution: Gene overlaps, recombination, and taxonomic implications. *Molecular Biology and Evolution* **13**, 1327-1338.

Lee, G. P., Min, B. E., Kim, C. S., Choi, S. H., Harn, C. H., Kim, S. U. & Ryu, K. H. (2003). Plant virus cDNA chip hybridization for detection and differentiation of four cucurbit-infecting Tobamoviruses. *Journal of Virological Methods* **110**, 19-24.

Letowski, J., Brousseau, R. & Masson, L. (2004). Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *Journal of Microbiological Methods* **57**, 269-278.

Li, F. & Stormo, G. D. (2001). Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics* **17**, 1067-1076.

Li, L., Tiam, Q., Du, Z., Duns, G. J. & Chen, J. (2009). A novel double stranded RNA virus detected in Primula malacoides is a plant-isolated partitivirus closely related to partivirus infecting fungal species. *Archives of Virology* **154**, 565-572.

Lima, W. F., Monia, B. P., Ecker, D. J. & Freier, S. M. (1992). Implication of RNA structure on antisense oligonucleotide hybridization kinetics. *Biochemistry* **31**, 12055-12061.

Linnen, J., Wages, J., Jr., Zhang-Keck, Z.-Y., Fry, K. E., Krawczynski, K. Z., Alter, H., Koonin, E., Gallagher, M., Alter, M. & other authors (1996). Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* **271**, 505-508.

Lisa, V., Boccardo, G. & Milne, R. G. (1981). Double-stranded ribonucleic acid from carnation cryptic virus. *Virology* **115**, 410-413.

Liu, W.-T., Guo, H. & Wu, J.-H. (2007). Effects of target length on the hybridization efficiency and specificity of rRNA-based oligonucleotide microarrays. *Applied and Environmental Microbiology* **73**, 73-82.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M. & other authors (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology* **14**, 1675-1680.

Lucas, W. J., Bouché-Pillon, S., Jackson, D. P., Nguyen, L., Baker, L., Ding, B. & Hake, S. (1995). Selective Trafficking of KNOTTED1 Homeodomain Protein and Its mRNA Through Plasmodesmata. *Science* **270**, 1980-1983.

Maldonado-Rodriguez, R., Espinosa-Lara, M., Calixto-Suarez, A., Beattie, W. & Beattie, K. (1999). Hybridization of glass-tethered oligonucleotide probes to target strands preannealed
with labeled auxiliary oligonucleotides. *Molecular Biotechnology* **11**, 1-12.

Malmstrom, C. M., Hughes, C. C., Newton, L. A. & Stoner, C. J. (2005a). Virus infection in remnant native bunchgrasses from invaded California grasslands. *New Phytologist* **168**, 217-230.

Malmstrom, C. M., McCullough, A. J., Johnson, H. A., Newton, L. A. & Borer, E. T. (2005b). Invasive annual grasses indirectly increase virus incidence in California native perennial bunchgrasses. *Oecologia* **145**, 153-164.

Marais, A., Faure, C., Couture, C., Svanella, L., Hullé, M., LeRomancer, M. & Candresse, T. (2009). Characterisation of plant virus populations by a metagenomic approach: Survey in french sub-antarctic islands. In *Understanding Emergence of Infectious*

Diseases: Focus on New Experimental and Theoretical Approaches to Virus Evolution. Roscoff, France: Conférences Jacques-Monod.

Margeridon, S., Carrouee-Durantel, S., Chemin, I., Barraud, L., Zoulim, F., Trepo, C. & Kay, A. (2008). Rolling circle amplification, a powerful tool for genetic and functional studies of complete hepatitis B virus genomes from low-level infections and for directly probing covalently closed circular DNA. *Antimicrobial Agents and Chemotherapy* **52**, 3068-3073.

Márquez, L. M., Redman, R. S., Rodriguez, R. J. & Roossinck, M. J. (2007). A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. *Science* **315**, 513-515.

Martelli, G. P., Adams, M. J., Kreuze, J. F. & Dolja, V. V. (2007). Family Flexiviridae: A case study in virion and genome plasticity. *Annual Review of Phytopathology* **45**, 73-100.

Martin, R. R., Pinkerton, J. N. & Kraus, J. (2009). The use of collagenase to improve the detection of plant viruses in vector nematodes by RT-PCR. *Journal of Virological Methods* **155**, 91-95.

Martin, R. R., Tzanetakis, I. E., Sweeney, M. & Wegener, L. (2006). A virus associated with blueberry fruit drop disease. *Acta Horticulturea* **715**, 497-501.

Maskell, L. C., Raybould, A. F., Cooper, J. I., Edwards, M. L. & Gray, A. J. (1999). Effects of turnip mosaic virus and turnip yellow mosaic virus on the survival, growth and reproduction of wild cabbage (*Brassica oleracea*). *Annals of Applied Biology* **135**, 401-407.

Matsui, S. M., Kim, J. P., Greenberg, H. B., Young, L. M., Smith, L. S., Lewis, T. L., Herrmann, J. E., Blacklow, N. R., Dupuis, K. & other authors (1993). Cloning and characterization of human astrovirus immunoreactive epitopes. *Journal of Virology* **67**, 1712-1715.

Mehlmann, M., Townsend, M., Stears, R., Kuchta, R. & Rowlen, K. (2005). Optimization of fragmentation conditions for microarray analysis of viral RNA. *Analytical Biochemistry* **347**, 316-323.

Melcher, U. (2000). The '30K' superfamily of viral movement proteins. *Journal of General Virology* **81**, 257-266.

Melcher, U., Muthukumar, V., Wiley, G. B., Min, B. E., Palmer, M. W., Verchot-Lubicz, J., R.S., N., Roe, B. A., Ali, A. & other authors (2008). Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. *Journal of Virological Methods* **152**, 49-55.

Menzel, W., Jelkmann, W. & Maiss, E. (2002). Detection of four apple viruses by multiplex RT-PCR assays with coamplification of plant mRNA as internal control. *Journal of Virological Methods* **99**, 81-92.

Mihindukulasuriya, K. A., Wu, G., St Leger, J., Nordhausen, R. W. & Wang, D. (2008). Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *Journal of Virology* **82**, 5084-5088.

Milgroom, M. G. & Cortesi, P. (2004). Biological control of chestnut blight with hypovirulence: A critical analysis. *Annual Review of Phytopathology* **42**, 311-338.

Muthukumar, V., Melcher, U., Pierce, M. L., Wiley, G. B., Roe, B. A., Palmer, M. W., Thapa, V., Ali, A. & Ding, T. (2009). Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Research* **141**, 169-173.

Nakatsukasa-Akune, M., Yamashita, K., Shimoda, Y., Uchiumi, T., Abe, M., Aoki, T., Kamizawa, A., Ayabe, S.-i., Higashi, S. & other authors (2005). Suppression of root nodule formation by artificial expression of the TrEnodDR1 (coat protein of *White clover cryptic virus 1*) gene in *Lotus japonicus*. *Molecular Plant-Microbe Interactions* **18**, 1069-1080.

Ndunguru, J., Taylor, N. J., Yadav, J., Aly, H., Legg, J. P., Aveling, T., Thompson, G. & Fauquet, C. M. (2005). Application of FTA technology for sampling, recovery and molecular characterization of viral pathogens and virus-derived transgenes from plant tissues. *Virology Journal* **2**, 45.

Nguyen, H. & Southern, E. (2000). Minimising the secondary structure of DNA targets by incorporation of a modified deoxynucleoside: implications for nucleic acid analysis by hybridisation. *Nucleic Acids Research* **28**, 3904–3909.

Nibert, M., Woods, K., Upton, S. & Ghabrial, S. (2009). *Cryspovirus*: a new genus of protozoan viruses in the family *Partitiviridae Archives of Virology* **154**, 1959-1965.

Odedara, O. O., Hughes, J. d. A., Odebode, A. C. & Odu, B. O. (2008). Multiple virus infections of lablab [*Lablab purpureus* (L.) Sweet] in Nigeria. *Journal of General Plant Pathology* **74**, 322-325.

Odedara, O. O., Hughes, J. d. A., Tarawali, S. A., Odebode, A. C. & Winter, S. (2007). Characterisation of a potyvirus from *Centrosema pubescens* Benth. *Tropical Science* **47**, 3-15.

Ooi, K., Ohshita, S., Ishii, I. & Yahara, T. (1997). Molecular phylogeny of geminivirus infecting wild plants in Japan. *Journal of Plant Research* **110**, 247-257.

Osaki, H., Nomura, K., Matsumoto, N. & Ohtsu, Y. (2004). Characterization of double-stranded RNA elements in the violet root rot fungus Helicobasidium mompa. *Mycological Research* **108**, 635-640.

Owor, B. E., Shepherd, D. N., Taylor, N. J., Edema, R., Monjane, A. L., Thomson, J. A., Martin, D. P. & Varsani, A. (2007). Successful application of FTA classic card technology and use of bacteriophage phi29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. *Journal of Virological Methods* **140**, 100-105.

Ozsolak, F., Platt, A. R., Jones, D. R., Reifenberger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M. & other authors (2009). Direct RNA sequencing. *Nature* **461**, 814-818.

Pagan, I., Alonso-Blanco, C. & Garcia-Arenal, F. (2008). Host responses in life-history traits and tolerance to virus infection in *Arabidopsis thaliana*. *PLoS Pathogens* **4**, e1000124.

Pagan, I., Alonso-Blanco, C. & Garcia-Arenal, F. (2009). Differential tolerance to direct and indirect density-dependent costs of viral infection in *Arabidopsis thaliana*. *PLoS Pathogens* **5**, e1000531.

Pasquini, G., Barba, M., Hadidi, A., Faggioli, F., Negri, R., Sobol, I., Tiberini, A., Caglayan, K., Mazyad, H. & other authors (2008). Oligonucleotide microarray-based detection and genotyping of *Plum pox virus*. *Journal of Virological Methods* **147**, 118-126.

Paximadis, M., Idris, A. M., Torres-Jerez, I., Villarreal, A., Rey, M. E. C. & Brown, J. K. (1999). Characterization of tobacco geminiviruses in the old and new world. *Archives of Virology* **144**, 703-717.

Peplies, J., Glockner, F. O. & Amann, R. (2003). Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes. *Applied and Environmental Microbiology* **69**, 1397-1407.

Peytavi, R., Tang, L.-Y., Raymond, F. R., Boissinot, K., Bissonnette, L., Boissinot, M., Picard, F. J., Huletsky, A., Ouellette, M. & other authors (2005). Correlation between microarray DNA hybridization efficiency and the position of short capture probe on the target nucleic acid *Biotechniques* **39**, 89-96.

Pfeiffer, P. (1998). Nucleotide sequence, genetic organization and expression strategy of the double-stranded RNA associated with the '447' cytoplasmic male sterility trait in Vicia faba. *Journal of General Virology* **79**, 2349-2358.

Polischuk, V., Budzanivska, I., Shevchenko, T. & Oliynik, S. (2007). Evidence for plant viruses in the region of Argentina Islands, Antarctica. *FEMS Microbiology Ecology* **59**, 409-417.

Powledge, F. (1999). Bay Biodiversity Awards - Marilyn Roossinck.

Pullen, M. E. (1968). Virus diseases in root crops sugar beet. In *The Incidence and Spread of Viruses Rothamsted Experimental Station Report for 1967*, pp. 124-125.

Pullen, M. E. (1969). Virus-like particles in sugar beet. In *Virus' Diseases of Grasses and Cereals Rothamsted Experimental Station Report for 1968* p. 127.

Qi, X., Bao, F. S. & Xie, Z. (2009). Small RNA deep sequencing reveals role for A*rabidopsis thaliana* RNA-dependent RNA polymerases in viral siRNA biogenesis. *PLoS ONE* **4**, e4971.

Rao, G. P. & Singh, M. (2008). Techniques in diagnosis of plant viruses. In *Techniques in diagnosis of plant viruses* (Plant Pathogens), pp. 1-47. Edited by G. P. Rao. New Delhi: New India Publishing Agency.

Rastgou, M., Habibi, M. K., Izadpanah, K., Masenga, V., Milne, R. G., Wolf, Y. I., Koonin, E. V. & Turina, M. (2009). Molecular characterization of the plant virus genus *Ourmiavirus* and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *Journal of General Virology* **90**, 2525-2535.

Raybould, A. F., Maskell, L. C., Edwards, M. L., Cooper, J. I. & Gray, A. J. (1999). The prevalence and spatial distribution of viruses in natural populations of *Brassica oleracea*. *New Phytologist* **141**, 265-275.

Rector, A., Bossart, G. D., Ghim, S.-J., Sundberg, J. P., Jenson, A. B. & Van Ranst, M. (2004). Characterization of a novel close-to-root papillomavirus from a Florida manatee by using multiply primed rolling-circle qmplification: *Trichechus manatus latirostris* Papillomavirus Type 1. *Journal of Virology* **78**, 12698-12702.

Rector, A., Tachezy, R., Van Doorslaer, K., MacNamara, T., Burk, R., Sundberg, J. & Van Ranst, M. (2005). Isolation and cloning of a papillomavirus from a North American porcupine by using multiply primed rolling-circle amplification: the *Erethizon dorsatum* papillomavirus type 1. *Virology* **331**, 449-456.

Redman, R. S., Sheehan, K. B., Stout, R. G., Rodriguez, R. J. & Henson, J. M. (2002). Thermotolerance generated by plant/fungal symbiosis. *Science* **298** 1581.

Rehtanz, M., Ghim, S.-J., Rector, A., Van Ranst, M., Fair, P. A., Bossart, G. D. & Jenson, A. B. (2006). Isolation and characterization of the first American bottlenose dolphin papillomavirus: *Tursiops truncatus* papillomavirus type 2. *Journal of General Virology* **87**, 3559-3565.

Renault, S., Stasiak, K., Federici, B. & Bigot, Y. (2005). Commensal and mutualistic relationships of reoviruses with their parasitoid wasp hosts. *Journal of Insect Physiology* **51**, 137-148.

Reyes, G. & Kim, J. (1991). Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Molecular and Cellular Probes* **5**, 473-481.

Robertson, N. L. (2005). A newly described plant disease complex involving two distinct viruses in a native Alaskan lily, *Streptopus amplexifolius*. *Canadian Journal of Botany* **83**, 1257-1267.

Robertson, N. L. (2007). Identification and characterization of a new virus in the genus *Potyvirus* from wild populations of *Angelica lucida* L. and *A. genuflexa* Nutt., family *Apiaceae*. *Archives of Virology* **152**, 1603-1611.

Rodoni, B. (2009). The role of plant biosecurity in preventing and controlling emerging plant virus disease epidemics. *Virus Research* **141**, 150-157.

Rodriguez, R. J., Redman, R. S. & Henson, J. M. (2004). Symbiotic lifestyle expression by fungal endophytes and the adaptation of plants to stress: unraveling the complexities of

intimacy. In *The Fungal Community: Its Organization and Role in the Ecosystem*, pp. 683–695. Edited by J. Dighton, J. F. W. Jr. & P. Oudemans. Boca Raton, FL: CRC Press.

Roossinck, M. J. (1997). Mechanisms of plant virus evolution. *Annual Review of Phytopathology* **35**, 191-209.

Roossinck, M. J. (2005). Symbiosis versus competition in plant virus evolution. *Nature Review Microbiology* **3**, 917-924.

Roossinck, M. J. (2010). Lifestyles of plant viruses. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 1899-1905.

Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., Chavarria, F., Shen, G. A. & Roe, B. A. (2010). Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Molecular Ecology* **19**, 81-88.

Rowhani, A., Biardi, L., Routh, G., Daubert, S. D. & Golino, D. A. (1998). Development of a sensitive colorimetric-PCR assay for detection of viruses in woody plants. *Plant Disease* **82**, 880-884.

Roye, M. E., McLaughlin, W. A., Nakhla, M. K. & Maxwell, D. P. (1997). Genetic diversity among geminiviruses associated with the weed species *Sida* spp., *Macroptilium lathyroides*, and *Wissadula amplissima* from Jamaica. *Plant Disease* **81**, 1251-1258.

Sabanadzovic, S. & Ghanem-Sabanadzovic, N. A. (2008). Molecular characterization and detection of a tripartite cryptic virus from rose. *Journal of Plant Pathology* **90**, 287-293.

Saldarelli, P., Rowhani, A., Routh, G., Minafra, A. & Digiaro, M. (1998). Use of degenerate primers in a RT-PCR assay for the identification and analysis of some filamentous viruses, with special reference to clostero- and vitiviruses of the grapevine. *European Journal of Plant Pathology* **104** 945-950.

Sampangi, R. K., Mohan, S. K. & Pappu, H. R. (2007). Identification of new alternative weed hosts for *Iris yellow spot virus* in the pacific northwest. *Plant Disease* **91**, 1683-1683.

Sanchez-Navarro, J., Aparicio, F., Herranz, M., Minafra, A., Myrta, A. & Pallas, V. (2005). Simultaneous detection and identification of eight stone fruit viruses by one-step RT-PCR. *European Journal of Plant Pathology* **111**, 77-84.

Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.

Scholthof, K.-B. G., Shaw, J. G. & Zaitlin, M. (1999). *Tobacco Mosaic Virus, One Hundred Years of Contributions to Virology*. St. Paul: APS Press.

Schubert, J., Habekuß, A., Kazmaier, K. & Jeske, H. (2007). Surveying cereal-infecting geminiviruses in Germany—Diagnostics and direct sequencing using rolling circle amplification. *Virus Research* **127**, 61-70

Seifarth, W., Spiess, B., Zeilfelder, U., Speth, C., Hehlmann, R. & Leib-Mösch, C. (2003). Assessment of retroviral activity using a universal retrovirus chip. *Journal of Virological Methods* **112**, 79-91.

Sengupta, S., Onodera, K., Lai, A. & Melcher, U. (2003). Molecular detection and identification of influenza viruses by oligonucleotide microarray hybridization. *Journal of Clinical Microbiology* **41**, 4542-4550.

Short, C. M. & Suttle, C. A. (2005). Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Applied and Environmental Microbiology* **71**, 480-486.

Simon-Buela, L. & Garcia-Arenal, F. (1999). Virus particles of cucumber green mottle mosaic tobamovirus move systemically in the phloem of infected cucumber plants. *Molecular Plant-Microbe Interactions* **12**, 112-118.

Sivalingam, P. N. & Varma, A. (2007). Non-tomato natural hosts of tomato infecting begomoviruses in north-western India. *Indian Journal of Virology* **18**, 20-27.

Skotnicki, M. L., Selkirk, P. M., Kitajima, E., McBride, T. P., Shaw, J. & Mackenzie, A. (2003). The first subantarctic plant virus report: Stilbocarpa mosaic bacilliform badnavirus (SMBV) from Macquarie Island. *Polar Biology* **26**, 1–7.

Southern, E., Mir, K. & Shchepinov, M. (1999). Molecular interactions on microarrays. *Nature Genetics* **21**, 5-9.

Staginnus, C., Gregor, W., Mette, M. F., Teo, C., Borroto-Fernandez, E., Machado, M. L., Matzke, M. & Schwarzacher, T. (2007). Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. *BMC Plant Biology* **7**, 24.

Stanley, W. M. (1935). Isolation of a crystalline protein possessing the properties of tobacco-maosaic virus. *Science* **81**, 644-645.

Stoye, J. P. (2006). Koala retrovirus: a genome invasion in real time. *Genome Biology* **7**, 241.

Strauss, E. E., Lakshman, D. K. & Tavantzis, S. M. (2000). Molecular characterization of the genome of a partitivirus from the basidiomycete Rhizoctonia solani. *Journal of General Virology* **81**, 549-555.

Sugiyama, S., Masuta, C., Sekiguchi, H., Uehara, T., Shimura, H. & Maruta, Y. (2008). A simple, sensitive, specific detection of mixed infection of multiple plant viruses using macroarray and microtube hybridization. *Journal of Virological Methods* **153**, 241-244.

Suttle, C. A. (2005). Viruses in the sea. *Nature* **437**, 356-361.

Suttle, C. A. (2007). Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology* **5**, 801-812.

Suzuki, A., Kobayashi, F., Abe, M., Uchiumi, T. & Higashi, S. (2001). Cloning and expression of a down-regulated gene (TrEnodDR1) of white clover responded by the nod genes derived from Rhizobium leguminosarum bv. trifolii strain 4S. *Gene* **266**, 77-84.

Szegö, A., Tóth, E. K., Potyondi, L. & Lukás, N. (2005). Detection of high molecular weight dsRNA persisting in *Dianthus* species. *Acta Biologica Szegediensis* **49**, 17-19.

Teycheneya, P.-Y., Acina, I., Lockhart, B. E. L. & Candresse, T. (2007). Detection of *Banana mild mosaic virus* and Banana virus X by polyvalent degenerate oligonucleotide RT-PCR (PDO-RT-PCR). *Journal of Virological Methods* **142**, 41-49.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

Thresh, J. M. (1982). Cropping practices and virus spread. *Annual Review of Phytopathology* **20**, 193-216.

Tillmann, H. L., Heiken, H., Knapik-Botor, A., Heringlake, S., Ockenga, J., Wilber, J. C., Goergen, B., Detmer, J., McMorrow, M., Stoll, M. & other authors (2001). Infection with GB virus C and reduced mortality among HIV-infected patients. *New England Journal of Medicine* **345**, 715-724.

Tzanetakis, I. E., Halgren, A., Mosier, N. & Martin, R. R. (2007). Identification and characterization of Raspberry mottle virus, a novel member of the *Closteroviridae*. *Virus Research* **127**, 26-33.

Urakawa, H., El Fantroussi, S., Smidt, H., Smoot, J. C., Tribou, E. H., Kelly, J. J., Noble, P. A. & Stahl, D. A. (2003). Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Applied and Environmental Microbiology* **69**, 2848-2856.

Valverde, R. & Gutierrez, D. (2007). Transmission of a dsRNA in bell pepper and evidence that it consists of the genome of an endornavirus. *Virus Genes* **35**, 399-403.

van der Hoek, L., Pyrc, K., Jebbink, M., Vermeulen-Oost, W., Berkhout, R., Wolthers, K., Wertheim-van Dillen, P., Kaandorp, J., Spaargaren, J. & other authors (2004). Identification of a new human coronavirus. *Nature Medicine* **10**, 368-373.

van Mölken, T. (2009). *Pros and cons of virus infections in plants, an ecological perspective*. PhD thesis, Radboud University, Nijmegen.

Veliceasa, D., Enünlü, N., Kós, P., Köster, S., Beuther, E., Morgun, B., Deshmukh, S. & Lukács, N. (2006). Searching for a new putative cryptic virus in *Pinus sylvestris* L *Virus Genes* **32**, 177-186.

Villarreal, L. P. (2005). *Viruses and the evolution of life*. Washington DC: ASM Press.

Villarreal, L. P., Defilippis, V. R. & Gottlieb, K. A. (2000). Acute and persistent viral life strategies and their relationship to emerging diseases. *Virology* **272**, 1-6.

Waigmann, E., Ueki, S., Trutnyeva, K. & Citovsky, V. (2004). The ins and outs of nondestructive cell-to-cell and systemic movement of plant viruses. *Critical Reviews in Plant Sciences* **23**, 195-250.

Wang, D., Coscoy, L., Zylberberg, M., Avila, P. C., Boushey, H. A., Ganem, D. & DeRisi, J. L. (2002). Microarray-based detection and genotyping of viral pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 15687-15692.

Wang, D., Urisman, A., Liu, Y.-T., Springer, M., Ksiazek, T. G., Erdman, D. D., Mardis, E. R., Hickenbotham, M., Magrini, V. & other authors (2003). Viral discovery and sequence recovery using DNA microarrays. *PLoS Biology* **1**, 257-260.

Webster, C. G., Wylie, S. J. & Jones, M. G. K. (2004). *Diagnosis of Plant Viral Pathogens*, vol. 86. Bangalore, INDE: Current Science Association.

Webster, C. G., Coutts, B. A., Jones, R. A. C., Jones, M. G. K. & Wylie, S. J. (2007). Virus impact at the interface of an ancient ecosystem and a recent agroecosystem: studies on three legume-infecting potyviruses in the southwest Australian floristic region. *Plant Pathology* **56**, 729–742.

Webster, R. G., Wright, S. M., Castrucci, M. R., Bean, W. J. & Kawaoka, Y. (1993). Influenza - a model of an emerging virus disease. *Intervirology* **35**, 16-25.

Wernersson, R., Junkcer, A. S. & Nielsen, H. B. (2007). Probe selection for DNA microarrays using OligoWiz. *Nature Protocols* **2**, 2677-2681.

Wheelis, M., Casagrande, R. & Madden, L. V. (2002). Biological attack on agriculture: low-tech, high-impact bioterrorism. *BioScience* **52**, 569-576.

Wilson, W. J., Strout, C. L., DeSantis, T. Z., Stilwell, J. L., Carrano, A. V. & Andersen, G. L. (2002). Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Molecular and Cellular Probes* **16**, 119-127.

Wong, C. W., Heng, C. L., Wan Yee, L., Soh, S. W., Kartasasmita, C. B., Simoes, E. A., Hibberd, M. L., Sung, W. K. & Miller, L. D. (2007). Optimization and clinical validation of a pathogen detection microarray. *Genome Biology* **8**, R93.

Wren, J. D., Roossinck, M. J., Nelson, R. S., Scheets, K., Palmer, M. W. & Melcher, U. (2006). Plant virus biodiversity and ecology. *PLoS Biology* **4**, e80.

Xie, W. S., Antoniw, J. F., White, R. F. & Jolliffe, T. H. (1994). Effects of beet cryptic virus infection on sugar beet in field trials. *Annals of Applied Biology* **124**, 451-459.

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

Yamaguchi, N., Seshimo, Y., Yoshimoto, E., Ahn, H. I., Ryu, K. H., Choi, J. K. & Masuta, C. (2005). Genetic mapping of the compatibility between a lily isolate of *Cucumber mosaic virus* and a satellite RNA. *Journal of General Virology* **86**, 2359-2369.

Yu, X., Li, B., Fu, Y., Jiang, D., Ghabrial, S. A., Li, G., Peng, Y., Xie, J., Cheng, J. & other authors (2010). A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences* **107**, 8387-8392.

Zhai, Y., Attoui, H., Mohd Jaafar, F., Wang, H., Chao, Y., Fan, S., Sun, Y., Liu, L., Mertens, P. & other authors (2010). Isolation and full-length sequence analysis of '*Armigeres subalbatus* totivirus', the first totivirus isolate from mosquitoes representing a proposed new genus (*Artivirus*) of the family *Totiviridae*. *Journal of General Virology*, vir.0.024794-024790.

Zhang, Y., Lartey, R. T., Hartson, S. D., Voss, T. C. & Melcher, U. (1999). Limitations to tobacco mosaic virus infection of turnip. *Archives of Virology* **144**, 957-971.

Zhu, Z., Chao, J., Yu, H. & Waggoner, A. S. (1994). Directly labeled DNA probes using fluorescent nucleotides with different length linkers. *Nucleic Acids Research* **22**, 3418-3422.

Zimin, A. V., Smith, D. R., Sutton, G. & Yorke, J. A. (2008). Assembly reconciliation. *Bioinformatics* **24**, 42 - 45.

Zitter, T. A. & Murphy, J. F. (2009). Cucumber mosaic virus. *The Plant Health Instructor*.

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

APPENDIX

Table 4: Sequences of DNA oligonucleotide non-target probes used as controls in the microarray study

Probe[a]	Target Species/Genus	Sequence 5' - 3'	Probe Length (nt)
10000831	*Cauliflower mosaic virus*	GTCTTCTTGGTGCAGCCAAAGGTAATCTAGCATACGGTAA GTTTATGTTTACTGTATACCCTAAGTTTGG	70
10000832	*Cauliflower mosaic virus*	CCAAACTTAGGGTATACAGTAAACATAAACTTACCGTATG CTAGATTACCTTTGGCTGCACCAAGAAGAC	70
10000833	*Cauliflower mosaic virus*	TACTTGAGAAAGTATGTTCAGAAAATCCATTAGATCCTAA CAAGACTAAGCAATGGATGAAAGCTTCTAT	70
10000834	*Cauliflower mosaic virus*	ATAGAAGCTTTCATCCATTGCTTAGTCTTGTTAGGATCTAA TGGATTTTCTGAACATACTTTCTCAAGTA	70
10000835	*Cauliflower mosaic virus*	GAAGCTAGAAGTACCGCTTAGGCAGGAGGCCGTTAGGGA AAAGATGCTAAGGCAGGGTTGGTTACGTTGA	70
10000836	*Cauliflower mosaic virus*	TCAACGTAACCAACCCTGCCTTAGCATCTTTTCCCTAACGG CCTCCTGCCTAAGCGGTACTTCTAGCTTC	70
10003779	*Turnip vein clearing virus*	GGCGTAGTGCGCACGATAGCGCATAGTGTTTTTCTCTCCA CTTAAATCGAAGAGATAAACTTACGGTGTA	70
10003780	*Turnip vein clearing virus*	TACACCGTAAGTTTATCTCTTCGATTTAAGTGGAGAGAAA AACACTATGCGCTATCGTGCGCACTACGCC	70
10003783	*Turnip vein clearing virus*	GCGAACCTCATGTGGAACTTCGAGGCCAAACTCTTCAGGA AGAAGTATGGTTACTTCTGTGGTCGTTATG	70
10003784	*Turnip vein clearing virus*	CATAACGACCACAGAAGTAACCATACTTCTTCCTGAAGAG TTTGGCCTCGAAGTTCCACATGAGGTTCGC	70
10003785	*Turnip vein clearing virus*	TATTCAAGACGAATATCCTGCATTGCAGACGATAGTCTAC CATTCGAAAAAGATCAATGCGATTTTCGGT	70
10003786	*Turnip vein clearing virus*	ACCGAAAATCGCATTGATCTTTTTCGAATGGTAGACTATC GTCTGCAATGCAGGATATTCGTCTTGAATA	70
10005353	*Turnip vein clearing virus*	GCGCCAGCTACATAGGCGTGGTGCACACGATAGTGCATAG TGTTTTTCTCTCCACTTAAATCGAAGAGAT	70
10005354	*Turnip vein clearing virus*	ATCTCTTCGATTTAAGTGGAGAGAAAAACACTATGCACTA TCGTGTGCACCACGCCTATGTAGCTGGCGC	70
10005355	*Turnip vein clearing virus*	AATACCCTGCATTGCAGACTATAGTCTATCATTCAAAAAA GATCAATGCAATTTTTGGTCCTATGTTTTC	70
10005356	*Turnip vein clearing virus*	GAAAACATAGGACCAAAAATTGCATTGATCTTTTTTGAAT GATAGACTATAGTCTGCAATGCAGGGTATT	70
10005357	*Turnip vein clearing virus*	AAGGAGTTTTTAGTTACTAGAGTTAATACGTGGTATTGTA AGTTTACTAGAGTCGATACTTTTACTCTTT	70
10005358	*Turnip vein clearing virus*	AAAGAGTAAAAGTATCGACTCTAGTAAACTTACAATACCA CGTATTAACTCTAGTAACTAAAAACTCCTT	70
10005361	*Turnip vein clearing virus*	CTCGACATTTCAAAGTATGACAAGTCACAAAACGAGTTCC ATTGCGCTGTTGAGTACAAAATATGGGAAA	70
10005362	*Turnip vein clearing virus*	TTTCCCATATTTTGTACTCAACAGCGCAATGGAACTCGTTT TGTGACTTGTCATACTTTGAAATGTCGAG	70
10006133	*Turnip vein clearing virus*	GCCGGCTAAATAGGCGTGGTGCACACGATAGTGCATAGTG TTTTCCTCTCCACTTAAATCGAAGAGGTAT	70
10006134	*Turnip vein clearing virus*	ATACCTCTTCGATTTAAGTGGAGAGGAAAACACTATGCAC TATCGTGTGCACCACGCCTATTTAGCCGGC	70
10006135	*Turnip vein clearing virus*	TATCCTGCTTTGCAGACGATAGTCTACCATTCAAAGAAGA TCAATGCAATTTTTGGTCCTATGTTTTCAG	70
10006136	*Turnip vein clearing virus*	CTGAAAACATAGGACCAAAAATTGCATTGATCTTCTTTGA ATGGTAGACTATCGTCTGCAAAGCAGGATA	70
10006141	*Turnip vein clearing virus*	TGCGGTCTCTTGAGCTGGAATATCTCATGATGCAAGTTCCT TTCGGATCTCTGACGTATGATATCGGCGG	70
10006142	*Turnip vein clearing virus*	CCGCCGATATCATACGTCAGAGATCCGAAAGGAACTTGCA TCATGAGATATTCCAGCTCAAGAGACCGCA	70
Caulimoa.3334	*Cauliflower mosaic virus*	AATGCAGGTGTTGGGTTTGT	20
Caulimoa.3774	*Cauliflower mosaic virus*	TATACCAGAAGAATWTTGG	19
Caulimoa.4272	*Cauliflower mosaic virus*	TGTTCAGAAAATCCTITAGATC	22

Caulimoa.4456	*Cauliflower mosaic virus*	CATATGTCICCAGCTTT	17
Caulimoa.4638A	*Cauliflower mosaic virus*	TTTGACTGTAAGAGIGGATTTTGGCAGGT	29
Caulimoa.4638B	*Cauliflower mosaic virus*	TTTGACTGTAAGTCIGGATTTTGGCAGGT	29
Caulimoa.4703	*Cauliflower mosaic virus*	GCWTTCACITGTCCACAAGGWCATTACCAATGGA	34
Caulimoa.5263	*Cauliflower mosaic virus*	ATTATTGAGACIGATGCATCTGA	23
Caulimoa.5389	*Cauliflower mosaic virus*	GAATTATCATAGTAATGAIAAGGA	24
Caulimoa.593	*Cauliflower mosaic virus*	TATGGIAAGTTTATSTTTACT	21
Caulimob.3683	*Cauliflower mosaic virus*	AAMAATATATIGAGAATTCATAGGA	25
Caulimob.4283	*Cauliflower mosaic virus*	AAATATTGCCCTAIAGGAAAGAA	23
Caulimob.4345A	*Cauliflower mosaic virus*	TTGCAAGTGTTGGATITGTCAIGAAGA	27
Caulimob.4345B	*Cauliflower mosaic virus*	GAAGATGGACATTATGCIAATGAATGTCCTAAGAAG	36
Caulimob.4587	*Cauliflower mosaic virus*	TATGTTGATACAGGTGCAICATTATGT	27
Caulimob.5247A	*Cauliflower mosaic virus*	AAAAGTCCICATTGTGCGCCAGC	23
Caulimob.5247B	*Cauliflower mosaic virus*	TTCTATGTAGAAAATCATAATGAGA	25
Caulimob.5305	*Cauliflower mosaic virus*	GTAATIAATTATAAAGCTITGAATGAAGC	29
Caulimob.5310	*Cauliflower mosaic virus*	TGAATGAAGCIACIATAGGA	20
Caulimob.5394	*Cauliflower mosaic virus*	TGGTTTTCAACITTAGATGTAAAATC	26
Caulimob.5487	*Cauliflower mosaic virus*	GAAACATTATIAGTGGAATGTTTTACCT	28
Caulimob.5509	*Cauliflower mosaic virus*	TTACCTITIGGGTTAAAGCAAGCTCCA	27
Caulimob.5532A	*Cauliflower mosaic virus*	GTTAAAGCAAGCTCCAGGTATTT	23
Caulimob.5532B	*Cauliflower mosaic virus*	TCAAGAITTTATGGATAGAA	20
Caulimob.5763	*Cauliflower mosaic virus*	GGAGAAATTAAACTICAACCATATGTTTT	29
Caulimob.5775	*Cauliflower mosaic virus*	CAACCAIATGTTTTGGAAAAATTGG	25
Caulimob.5828	*Cauliflower mosaic virus*	AAGCAATTACAGIGWTTTCTTGG	23
Caulimob.5876	*Cauliflower mosaic virus*	AAGGATTTTTTAAGGAIWTTGCAAAA	26
Caulimob.6373	*Cauliflower mosaic virus*	TGTTACAGGWTTTITATGATATAA	24
Caulimob.6408	*Cauliflower mosaic virus*	AATTATAATCAAGGACGGTTTGTT	24
Caulimob.6514	*Cauliflower mosaic virus*	GACACTCTAACCAGGGAATGGAAG	24
Caulimob.6886	*Cauliflower mosaic virus*	CTTCACAAGAGITATCCAACCAT	23
Caulimob.6901	*Cauliflower mosaic virus*	TCCAACCATIGAIGIAGCAAAGAAG	25
Potexa.2724	*Potexvirus*	GCATCAGCAGGCIAAGGATGAG	22
Potexa.3025	*Potexvirus*	TTTCTGAAGISICAGTGGGTIAAGAAGGT	29
Potexa.3124A	*Potexvirus*	GIACTATGGCSAGGTAC	17
Potexa.3124B	*Potexvirus*	GCACSATGGCGCGMTAC	17
Potexa.3124C	*Potexvirus*	GTACAATGGCCCGCTAC	17

Potexa.3441A	*Potexvirus*	GGTGAGGGICCMACTTTTGATGCTAACACAGAG	33
Potexa.3441B	*Potexvirus*	GGAGAGGGICCMACITTTGATGCTAATACTGAG	33
Potexa.3441C	*Potexvirus*	GGWGAGGGICCTACTTTTGATGCIAATACWGAG	33
Potexa.1737	*Potexvirus*	GGTGGITGIGGMAAGTC	17
Potexa.2299	*Potexvirus*	TATGCIGGITGTCAGGGIMTTAC	23
Potexb.2215	*Potexvirus*	GTMCTICCIACCAATGAGCT	20
Potexb.3114	*Potexvirus*	CATCAGCAGGCIAAGGATGA	20
Potexb.3411	*Potexvirus*	TTTCTIAAGTCICAGTGGGT	20
Potexb.3428	*Potexvirus*	TGGGTIAAGAAGGTGGAGAAGTT	23
Potexb.3505	*Potexvirus*	CTGTCATGCTTTITGGMACIATG	23
Potexb.3660	*Potexvirus*	GCCAATGACTITACIGCNGT	20
Potexb.3679	*Potexvirus*	TTTGATCAGTCICAGGATGG	20
Potexb.3825	*Potexvirus*	ACCGGIGAGGGICCCAC	17
Potexb.3855A	*Potexvirus*	CCCACITTTGATGCIAACACTGAGT	25
Potexb.3855B	*Potexvirus*	AACACTGAGTGTAATATIGCCCTTC	25
Potexb.544	*Potexvirus*	TTTATGGGIGATACTCTICATTT	23
Potexb.6016	*Potexvirus*	GCTGSITTTGACTTCTTTGATGGCGTG	27
Potexa.1737M (T)	*Ambrosia* aymptomatic virus1	CTTTTCCCACTGCCCCCTTTTTTTTTTTTTTTTTTTT	38
Potexa.1737P (T)	*Ambrosia* asymptomatic virus1	GGGGGCAGTGGGAAAAGTTTTTTTTTTTTTTTTTTT	37
Potexa.2299M (T)	*Ambrosia* asymptomatic virus1	GTGAGGCCCTGGCAGCCTGCGTATTTTTTTTTTTTTTTTTTTTT	44
Potexa.2299P (T)	*Ambrosia* asymptomatic virus1	TACGCAGGCTGCCAGGGCCTCACTTTTTTTTTTTTTTTTTTTTT	44
Potexb.2215 (T)	*Ambrosia* asymptomatic virus1	ATCCTTCCCACCAATGAGCTTTTTTTTTTTTTTTTTTTT	40
Potexb.3825 (T)	*Ambrosia* asymptomatic virus1	AGCGGTGAGGGGCCCACTTTTTTTTTTTTTTTTTTTT	38
Potexb.3855A (T)	*Ambrosia* asymptomatic virus1	CCCACCTTCGACGCAAACACCGAGTTTTTTTTTTTTTTTTTTTT	45
Potexb.3855B (T)	*Ambrosia* asymptomatic virus1	AACACCGAGTGTTCTATAGCCTATGTTTTTTTTTTTTTTTTTTTT	46
Potexb.544 (T)	*Ambrosia* asymptomatic virus1	TACATGTCGGATACCCTACATTTTTTTTTTTTTTTTTTTT	41
Potexb.6016 (T)	*Ambrosia* asymptomatic virus1	GCAGCTTTTGACTTCTTCCAGGGCGTCTTTTTTTTTTTTTTTTTTTT	48
TobamoII.1011	*Tobamovirus*	GTTTATGTAAAGGAGTTTATGGTTAAG	27
TobamoII.1023	*Tobamovirus*	GAGTTTATGGTTAAGCGTGIGATACTT	27
TobamoII.1023B	*Tobamovirus*	TTCCAGGATAAAGCCACGTTTTCTG	25

TobamoII.1198	*Tobamovirus*	AGCCACGTTTTCTGTGTGGTTTCCT	25
TobamoII.1390	*Tobamovirus*	GTGGIGGAATGTTCAGTC	18
TobamoII.1935	*Tobamovirus*	GTCATAGAGGCIATTTTTTCICA	23
TobamoII.2376	*Tobamovirus*	AGATTGAAGAAGACTATCACICCIGT	26
TobamoII.4086	*Tobamovirus*	GTAGATAATTTCTTTICTWCTTT	23
TobamoII.4339	*Tobamovirus*	TTGCAGACKATIGTTTATCAT	21
TobamoII.4369	*Tobamovirus*	GTIAATGCIGTTTTTGGTCC	20
TobamoII.4695	*Tobamovirus*	TATTATCAGIGGAAGTCTGG	20
TobamoII.4710	*Tobamovirus*	TCTGGIGATGTIACTACTTTTATAGG	26
TobamoII.4710B	*Tobamovirus*	GGTAATACTTTTATTATIGCIGCKTGTGT	29
TobamoII.4803	*Tobamovirus*	TTTTGTGGTGATGATTCGCTIATCTAT	27
TobamoII.4956	*Tobamovirus*	GGSTGTATTGTITATCCIGATCC	23
TobamoII.4986	*Tobamovirus*	TTAATTISTAAATTAGGTAATAAGAGT	27
TobamoII.5104	*Tobamovirus*	GATGATGCIATCCACGAGTWITTTCCTAG	29
TobamoII.5139	*Tobamovirus*	GGTAGCAGTTTTGTIATTAGTTCTG	25
TobamoII.5139B	*Tobamovirus*	TTTGTIATTAGTTCTGTGCAAGTATTT	27
TobamoII.6382	*Tobamovirus*	AATAGGGTTATTGAGGTTGAAAACCCTCTA	30
TobamoII.6452	*Tobamovirus*	AAGCGTAATGATGACGCGTCTACIGCTGC	29
TobamoII.79	*Tobamovirus*	ACAATATGGCAAACATTACACAACAAAT	28
TobamoI-III.1020	*Tobamovirus*	AAGGAGTTTITGGTIACTAG	20
TobamoI-III.1138	*Tobamovirus*	GCIATGGAIGATGCITGGIA	20
TobamoI-III.1454	*Tobamovirus*	GCIAGGTCIGAGTGGGATGT	20
TobamoI-III.2418	*Tobamovirus*	AGGTTCAGCAGATGIAGAACT	21
TobamoI-III.2822	*Tobamovirus*	GTTCCIGGITGTGGGAAGAC	20
TobamoI-III.350	*Tobamovirus*	AGTTGGAGTATCTSATGATGCAAGTTCC	28
TobamoI-III.4008	*Tobamovirus*	GTIGCIATGATTAAGAGAAAT	21
TobamoI-III.4340	*Tobamovirus*	TGCAGACIATWGTITATCATTCG	23
TobamoI-III.4557	*Tobamovirus*	CAGAATGAGTTTCATTGTGCWGTIGAGTAT	30
TobamoI-III.4626	*Tobamovirus*	GAGGTITGGAGACAGGGGCATAGGAAGAC	29
TobamoI-III.4693	*Tobamovirus*	TGTGGTATCAGAGGAAGAGTGGTGATGT	28
TobamoI-III.4878	*Tobamovirus*	ATGTGGAATTTTGAGGCIAAG	21
TobamoI-III.6368	*Tobamovirus*	CTTTTGATACTAGGAATAGGAT	22
TobamoI-III.6760	*Tobamovirus*	CCCTCCACTTAAATCGAAGGGTT	23
TobamoI-III.6760B	*Tobamovirus*	GGTTCCATTTAAATCGAAACCTG	23
TobamoI-III.6964	*Tobamovirus*	GAGGGGTTCGAATTCTCCC	19

Tymo.3202PD	*Tymovirus*	AACATGAAGAATGGTTTTGATGGC	24
Tymo.320MD (T)	*Tymovirus*	GCCATCAAAACCATTCTTCATGTTTTTTTTTTTTTTTTTTT	42
Tymo.5391PD	*Tymovirus*	ACTTATGATGACAATWCTGACTACAAC	27
Tymo.5391MD (T)	*Tymovirus*	GTTGTAGTCAGWATTGTCATCATAAGTTTTTTTTTTTTTTTTTTTT	46
Tymo.544PD	*Tymovirus*	CATGCACGATGCICTGATGTATT	23
Tymo.544MD (T)	*Tymovirus*	AATACATCAGIGCATCGTGCATGTTTTTTTTTTTTTTTTTTTT	43
Tymo.829PD	*Tymovirus*	TCCTGGAGTCTTGGGGCCCTC	21
Tymo.829MD	*Tymovirus*	GAGGGCCCCAAGACTCCAGGATTTTTTTTTTTTTTTTTTTT	41

70mer probes for *Turnip vein clearing virus* and *Cauliflower mosaic virus* are probes described by Wang et al., 2003; M-minus-sense probe, P-plus-sense probe, PD-plus-sense degenerate probe, MD-minus-sense degenerate probe, (T)-Tailed/Spacer probe.Probes not indicated as P or M are plus sense probes.

VITA

Veenita Grover Shah

Candidate for the Degree of

Doctor of Philosophy

Thesis: DEVELOPMENT OF OLIGONUCLEOTIDE-BASED MICROARRAYS FOR THE DETECTION OF PLANT VIRUSES AND THE FIRST CHARACTERIZATION OF A PLANT VIRUS BELONGING TO THE FAMILY *TOTIVIRIDAE*

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:
Bachelor of Science degree in Biochemistry from University of Delhi, Delhi, India in May 2003.
Master of Science degree in Biosciences from Jamia Milia Islamia University, Delhi, India in May 2005.
Completed the requirements for the Doctor of Philosophy in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in December 2010.

Experience:
Worked as an intern in the Institute of Genomics and Integrative Biology, Delhi, India in summer 2004.
Employed by Oklahoma State University, Department of Biochemistry and Molecular Biology as a graduate research assistant, 2005 to present.

Professional Memberships:
The American Association for the Advancement of Science
Oklahoma Academy of Sciences
American Phytopathological Society: May 2007-June 2008

Name: Veenita Grover Shah Date of Degree: December, 2010

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: DEVELOPMENT OF OLIGONUCLEOTIDE-BASED MICROARRAYS FOR THE DETECTION OF PLANT VIRUSES AND THE FIRST CHARACTERIZATION OF A PLANT VIRUS BELONGING TO THE FAMILY *TOTIVIRIDAE*

Pages in Study: 137 Candidate for the Degree of Doctor of Philosophy

Major Field: Biochemistry and Molecular Biology

Scope and Method of Study: Ongoing studies for discovery of viral species in non-cultivated plants utilize several genomic approaches for systematic unbiased searches for viruses related to known viruses. Some of these approaches and their suitability for investigations of viruses in wild plants are summarized here. These methods include particularly array hybridization to macroarrays and microarrays, and various megasequencing approaches. A part of this thesis work was aimed at developing a microarray based on a sequence-independent amplification method with the potential to detect a broad group of plant viruses including detection of new virus species, strains and variants. The other part of the work was to investigate a putative virus belonging to the family *Totiviridae*, discovered from the wild plant population of Tallgrass Prairie Preserve of Oklahoma and designated here as *Ruellia* asymptomatic virus 1. This study addressed two main questions. First, whether the putative virus is a mycovirus of a fungal endophyte that is colonizing the host plants or a plant virus or a virus that is transmitting between plants and fungi. Second, if the virus is a plant virus, is it possibly a persistent plant virus.

Findings and Conclusions: The microarray method developed and described in this study demonstrates a viable procedure for nucleic acid amplification and hybridization that should be effective in detecting most plant RNA or DNA viruses as long as the virus has some representative sequence information available. The system was validated in pure and mixed samples by detection of three *Tymovirus* species: *Asclepias* asymptomatic virus, *Kennedya yellow mosaic virus* and *Turnip yellow mosaic virus*. The method could detect sequence variants with 70-75% or higher sequence identity, indicating the possible utility of the approach for virus discovery. From the second part of the work, we show evidence for the first putative plant virus in the family *Totiviridae*. Using *in situ* hybridization, the viral RNA was observed to localize inside the cells of virus-infected plant stem sections, suggesting its identification as a plant virus. Seed transmission analysis performed on the progenies of the virus-infected plants demonstrated vertical transmission of the virus, and hence persistence of virus in its host plants. In conclusion, investigation of viruses in the wild plant population can serve as a launching point for additional studies focused on demonstrating disease causality, and dissecting the possible mutualistic roles of plant viruses in natural settings.

ADVISER'S APPROVAL: Dr. Ulrich K. Melcher