SEQUENCING AND CHARACTERIZATION OF

CANNA YELLOW MOTTLE VIRUS

By

DULANJANI SHANIKA WIJAYASEKARA

Bachelor of Science

University of Peradeniya

Sri Lanka

2013

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2016

SEQUENCING AND CHARACTERIZATION OF

CANNA YELLOW MOTTLE VIRUS

Thesis Approved:

Dr. Jeanmarie Verchot

Thesis Adviser

Dr. Robert Hunger

Dr. Hoabo Jiang

ACKNOWLEDGEMENTS

It is with great pleasure, that I express my sincere gratitude to my supervisor Dr. Jeanmarie Verchot, Department of Entomology and Plant Pathology, Oklahoma State University for her valuable guidance and supervision. She has given me the opportunity to learn not only to be a scientist but also to be a good person. I wish to mention with deep gratitude about her inspiration and encouragement given to me throughout this journey. I would also like to thank my committee members, Dr. Robert Hunger and Dr. Hoabo Jiang for their valuable suggestions and comments, your advice and helping hand through this journey was indeed a blessing for me. I also like to thank Dr. Peter Hoyt, Director of the OSU Array and Bioinformatics Core Facility and his crew for all the support given for my project. I also wish to acknowledge and thank Terry Colberg and Lisa Whitworth for the technical assistance and sample preparation for transmission electron microscopy. I also like to thank Dr. Phillip Mulder and all department members for all the encouragement and blessings given for me throughout my stay here. I express my thanks and appreciation to my mother Sunethra Liyandipitiya, father Sarath Wijayasekara, and my brother Lahiru Wijayasekara for the endless courage and support given to me. Jerry Clark, my husband I would like to thank you and your family for the love and motivation given to me. Lastly, but in no sense the least, I am thankful to all of my friends who made my stay at the university memorable.

Name: DULANJANI S WIJAYASEKARA

Date of Degree: JULY, 2016

Title of Study: SEQUENCING AND CHARACTERIZATION OF *CANNA YELLOW MOTTLE VIRUS*

Major Field: PLANT PATHOLOGY

Badnaviruses are pararetroviruses that exist as episomal viruses or as integrants in plant chromosomes by illegitimate recombination. Endogenous virus sequences can be a hidden reservoir of infection that emerge when plants are stressed, or a source of small RNAs that contribute to anti-viral defense. Advanced nucleic acid sequencing technology has enabled the discovery of complicated viromes existing in naturally infected plants and provided for new insights into virus synergistic interactions. Such approaches are vital to gaining insight into economically significant diseases. We report the entire *Canna yellow mottle badnavirus* (CaYMV) genome that exists in naturally infected hybrid cannas. We also identified two subviral DNAs related to CaYMV. One subviral DNA, named EVS-0036, encodes an intact reverse transcriptase and RNase H. Southern analysis identified episomal forms of CaYMV genomes. EVS-0036 transcripts are detected in canna plants naturally infected with *Canna yellow streak potyvirus* (CaYSV) or CaYMV. The results demonstrate the occurrence of endogenous subviral DNAs which, through evolution is related to badnaviruses, but preferentially occurs alongside an RNA virus. Because they are vegetatively transmitted, CaYMV, CaYSV and EVS-0036 share a long-term association within canna that represents a powerful model to explore the mutualistic-parasitic continuum that can involve virus-host and virus-virus interactions.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figures Page

CHAPTER II

CHAPTER IV

Figure Page

CHAPTER I

INTRODUCTION AND OBJECTIVES

Introduction

Canna Linnaeus is a new world genus, and the only member of the family *Cannaceae* belonging to the Order *Zingiberales*. The Order *Zingiberales* consists of eight families including: *Cannaceae*, *Musaceae*, *Costaceae*, *Heliconiaceae*, *Lowiaceae*, *Maranthaceae*, *Strelitziaceae* and *Zingiberaceae*. Members of this order include economically important crop plants grown primarily in tropical areas such as banana, ginger and canna. Members of this order include perennial herbaceous plants with rhizomes rich in aromatic, volatile oils, and many are used as spices or condiments (Specht et al., 2001). Based on previous phylogenetic studies, the family *Cannaceae* is known to be closely related to the family *Maranthaceae* (Prince, 2010). Cannas are herbaceous plants with fleshy horizontal rhizomes that consist of well-developed aerial stems bearing large foliage and varying colored flowers ranging from red, orange, purple, yellow or white (Mass-Van De Kamer and Mass, 2008). The genus *Canna* L. has ten to 50 species including a large *Canna indica* L. complex (Kress, 1990; Mass-Van De Kamer and Mass, 2008).

Cannas were domesticated either as a floricultural or agronomic crop. Cannas are native to Central and South America, which are centers for their diversification. *Canna indica L.* was reported to be one of the first plant species to be anthropogenically cultivated in Peru and Argentina (Prince, 2010) and the purpose of domestication was as a food starch (Gade, 1966) . It was speculated that these plants were subjected to early human dispersal from the New World (Kress, 1990). Canna was extensively hybridized during cultivation and these canna hybrids were naturalized in Asia and Africa and are now internationally traded as a floricultural crop. Canna is traded globally as a horticultural crop where large acreage of land in Oklahoma and southern US is dedicated to growing canna.

The canna horticulture industry is currently threatened by viral diseases. The canna industry in Oklahoma is mainly based on the production and sale of rhizomes (Rajakaruna et al., 2013). There are only five viruses reported to cause disease in canna that include *Canna yellow streak virus* (CaYSV; *Potyvirus*), *Bean yellow mosaic virus* (BYMV; *Potyvirus*), *Canna yellow mottle virus* (CaYMV; *Badnavirus*), *Cucumber mosaic virus* (CMV; *Cucumovirus*) and *Tomato aspermy virus* (TAV; *Cucumovirus*) (Rajakaruna et al., 2013).

Objectives

The focus of my research is on CaYMV. CaYMV is reported to be an emerging global pathogen and the genome sequence for this virus has never been reported (Borah et al., 2013). This project began with data suggesting that CaYMV is a member of the

Badnavirus group of plant viruses and belongs to the family *Caulimoviridae* (Geering et al., 2010).

My first objective was to sequence the full genome length of CaYMV from naturally infected canna grown in the greenhouse. While we were able to recover the entire CaYMV genome, we also identified subviral DNA elements related to CaYMV, which are likely integrated into the plant nuclear genome. In order to understand possible evolution of CaYMV and endogenous virus sequences (EVS), phylogenetic analysis was conducted out using other NCBI reported badnaviral species. Specific diagnostic primers were developed to distinguish between the CaYMV full genome and the EVS in canna. The initial objectives for this proposed research were:

1) Carry out next generation sequencing (NGS) to identify the full genome of CaYMV. Annotate the genome. Carry out phylogenetic analysis comparing CaYMV with other badnaviruses.

2) Develop PCR diagnostic primers to detect CaYMV in a population of infected cannas. Clone CaYMV from infected plants to confirm the NGS data.

3) Inoculate CaYMV to test plants to identify alternative hosts. Determine if CaYMV and CaYSV cause synergistic disease.

CHAPTER II

REVIEW OF LITERATURE

The *Caulimoviridae* family

Plant viruses with double stranded circular DNA genomes with one or two single stranded discontinuities belong to the family *Caulimoviridae*. Members of this family possess icosahedral or bacilliform-shaped virion particles. The average particle dimension for icosahedral particles are 45-50 nm and for bacilliform particles is 35-50 nm x 900 nm (Borah et al., 2013). Members of *Caulimoviridae* are also known as pararetroviruses as they contain a reverse transcription-based stage in their replication cycle (Geering et al., 2010).

Family *Caulimoviridae* is currently not assigned to any order or phylum. Based on the 2013 International Committee of Taxonomy of Viruses (ICTV) there are 103 total virus families consisting 2,827 species (Andrew et al., 2012). Among these families, there are 77 families that are not yet assigned to an order or phylum. *Caulimoviridae* contains seven genera which are categorized based on their genomic organization and virion shape. Five genera possessing icosahedral morphology include *Caulimovirus*

(type species *Cauliflower Mosaic Virus*; CaMV), *Soymovirus* (type species *Soybean Chlorotic Mottle Virus*; SbCMV), *Cavemovirus* (type species *Cassava Vein Mosaic Virus*; CVMV), *Petuvirus* (type species *Petunia Vein Clearing Virus*; PVCV) and *Solendovirus* (type species *Tobacco Vein Clearing Virus*; TVCV). There are two genera that possess bacilliform-shaped particle morphology that include *Badnavirus* (type species *Commelina Yellow Mosaic Virus*; CoYMV) and *Tungrovirus* (type species *Rice Tungro Bacilliform Virus*; RTBV).

The type species of the genus *Caulimovirus* is *Cauliflower Mosaic Virus* (CaMV) which was the first plant viral genus reported in 1968 to contain a double stranded DNA genome (Haas et al., 2002). CaMV particles are non-enveloped and icosahedral in shape. The circular DNA genome is approximately 8000 bp in size. Single stranded discontinuities are observed in the places where DNA strands usually overlap. Both positive and negative strands of this genome can retain these discontinuities and number of discontinuities can vary based on the CaMV strain (Haas et al., 2002). These discontinuities act as primer positions for negative and positive strand synthesis during the reverse transcription phase. All members of the genus *Caulimovirus* encode several open reading frames (ORFs) for single functional proteins and a larger ORF coding for a polyprotein that is proteolytically processed to produce mature products (Geering et al., 2010). ORF1 encodes the P1 protein which is the movement protein (MP). ORF2 and 3 encode proteins required for aphid transmission. ORF IV encodes a precursor of the coat protein (CP). ORF V encodes the polyprotein precursor which contains the viral protease (PR), reverse transcriptase (RT) and ribonuclease H (RH). ORF6 encodes a translation trans-activator (TAV). Other than ORF7 which lies between two intergenic regions of

the genome, all other ORFs overlap each other by several nucleotides (Haas et al., 2002). ORF7 protein is not a stable product and it is suggested that it may act as a nucleotide binding protein (Givord et al., 1988). CaMV 35S promoter is a strong constitutive promoter that is involved in transcriptional regulation of CaMV and is widely used in plant biotechnology as a promotor (Haas et al., 2002).

Genus *Badnavirus*

The genus *Badnavirus* consists of 25 species with average genome size of 7.5 kb (Hull, 2007). The particles are bacilliform shaped and average 30 nm in diameter and 130 nm in length (Bouhida et al., 1993; Zhang et al., 2015). A typical badnavirus genome encodes three ORFs as depicted in Fig. 1. ORF1 encodes a virion associated protein of unknown function (Bouhida et al., 1993; Zhuang et al., 2011). The product of ORF2 also has an unknown function (Bhat et al., 2013; Deeshma and Bhat, 2015). ORF3 encodes a polyprotein that is proteolytically processed by a viral encoded PR to produce mature proteins including the MP, CP, aspartic acid PR, RT, zinc finger binding protein (Zn finger) and RH (Zhuang et al., 2011). The global distribution of badnaviruses suggests that these viruses are emerging from tropical areas of the world. Most species were first reported in Africa, South America, Southeast Asia as well as the Pacific Islands (Borah et al., 2013). In the last decade, many new virus species with double stranded DNA genomes have been reported, which are tentative members of the genus *Badnavirus*. These viruses have potential for becoming a major threat to agriculture in the long term. It is reported that members of the genus *Badnavirus* share very low nucleotide sequence similarity between each other and also number of ORF can vary depending on the virus (Borah et al., 2013).

Compared to other badnaviral species, *Banana Streak Virus* (BSV) has been studied extensively as banana is considered as one of the most important agricultural crops in the world (Ploetz et al., 2015) and also one of the staple foods in African countries (Borah et al., 2013). Interestingly both banana and canna belong to the order *Zingiberales*, but banana belongs to a separate family, *Musaceae.* It is reported that most basal groups of *Badnavirus* infect monocotyledonous hosts (Bousalem et al., 2008). BSV was first reported from Ivory Coast in 1958 (Iskra-Caruana et al., 2014). BSV has posed a threat to banana production in many countries including Africa, Asia, Australia, America and Europe (Zhuang et al., 2011). Symptoms of banana streak disease include yellow and necrotic streaks on banana leaves, pseudo stem splitting and stem necrosis (Geering et al., 2000). BSV is transmitted by mealybugs or by vegetative propagation of banana plants (Iskra-Caruana et al., 2014; Kubiriba et al., 2001). BSV can occur as free viral particles infecting plant cell (episomal) or as endogenous BSV that is integrated into the banana genome. The endogenous presence of BSV has threatened the international trading of banana germplasm as they can successfully yield infectious episomal genome of BSV (Ndowora et al., 1999). The proposed mechanism of conversion of endogenous BSV into episomal BSV is most likely by homologous recombination. In the banana cultivar 'Obino', tissue culture is recognized as a factor that can induce production of episomal BSV from endogenous BSV (Ndowora et al., 1999). It is reported that episomal BSV isolates obtained from different *Musa* cultivars have significant sequence variation because of homologous recombination within the BSV genome upon excision from the plant genome (Geering et al., 2000; Ndowora et al., 1999).

Serological studies on sugarcane have shown that BSV is closely related to the *Sugarcane bacilliform virus* (SCBV) (Lockhart and Autrey, 1988). SCBV is a rare badnavirus that has the ability to infect multiple monocot hosts; sugarcane, banana and rice. In one study it was reported that mealybugs can transmit SCBV from sugarcane to banana (Bouhida et al., 1993). SCBV does not cause yield reduction in modern sugarcane cultivars. Endogenous sequences related to SCBV (approximately 500 bp in length) were identified in sugarcane samples isolated in Guadeloupe and suggest that endogenization of sequences may be common amongst badnaviruses (Muller et al., 2011).

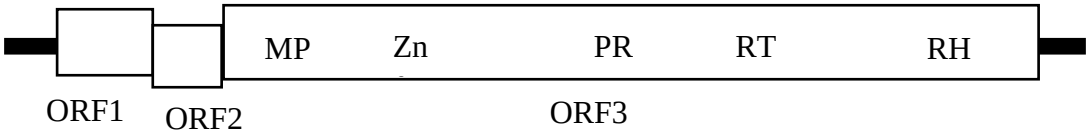

Fig. 1. Linear view of the circular genome organization of *Badnavirus* species representing three ORFs with different gene products. Open boxes identify each ORF. ORF3 is a polyprotein and the names for each region within the polyprotein are identified by the abbreviations: MP (movement protein), Zn finger (zinc finger protein), PR (aspartic protease protein), RT (reverse transcriptase), RH (RNase H protein). Black boxes at each end represent the noncoding region.

Endogenous Pararetroviruses

Mobile genetic elements, or transposable elements, are DNA sequences that can change their positions within the genome and were first predicted by Babara McClintock in 1948. These transposons were discovered in many organisms including bacteria, fungi and humans. Retrotransposons are categorized under class one transposable elements based on their ability to translocate through an RNA mediated copy and paste mechanism. A typical intact retrotransposon contains long terminal repeats and integrase (Hsu et al., 2016). Reverse transcribing elements can exist as mobile DNA or RNA elements. Essentially the DNA element containing long terminal repeats can insert into a chromosome location in nuclear DNA with the assistance of an integrase enzyme. An active retrotransposon will produce RNA that is full length and can be reverse transcribed into another DNA copy that can insert into another location in the genome using the same integrase activity. Alternatively, the DNA itself can be excised from its location in the genome and then reinsert into another location providing two mechanisms of mobility throughout the plant genome.

Pararetroviruses contain DNA genomes while retroviruses contain RNA genomes (Bousalem et al., 2008). The existence of pararetroviruses is witnessed in both the plant and animal kingdoms. In plants, family *Caulimoviridae*, and in animals there is only one family that is reported to have pararetroviruses such as family *Hepadnaviridae* (Bousalem et al., 2008; Hohn et al., 1997). Seven genera of the family *Caulimoviridae* including badnaviruses are recognized as plant pararetroviruses (Staginnus and Richert-Pöggeler, 2006). Paralleling the animal retroviruses, in plants this reverse transcribed DNA copy of the pararetrovirus genome can be integrated to the plant host genome

which are then called endogenous pararetroviruses (EPRV) (Geering et al., 2010). The underlying mechanism for genome insertion of EPRVS is not known since most pararetroviruses do not encode an integrase (Ndowora et al., 1999). However, some reports suggest that integration of DNA can occur via illegitimate recombination (Jakowitsch et al., 1999) or repairing double stranded breakage of host chromosome using viral DNA (Ndowora et al., 1999). Most members of EPRV from family *Caulimoviridae* are known to be inactive in host plant genome but some sequences are able to cause infection when host plant in stress condition (Geering et al., 2010).

Integrated BSV isolates in the nuclear genome of *Musa balbisiana Colla* species were introduced to cultivated banana species via interspecific hybridization with *Musa acuminata*. These integration events are more recent and occur frequently, and result in high sequence variation between BSV isolates through production of episomal BSV from these integrated sequences (Gayral and Iskra-Caruana, 2009; Gayral et al., 2008; Geering et al., 2001). The global trade of banana is threatened as unprecedented vegetative dissemination of these endogenous BSV elements can cause epidemics of BSV infection in cultivated banana (Dahal et al., 2000). Similar threats were reported from other badnaviral species as well, such as *Sugarcane bacilliform virus* (SCBV), *Dahlia mosaic virus* (DMV) D10, *Pineapple bacilliform virus* (PBV), *Fig Badnavirus* (FBV) and *Yam badnavirus* (YBV) (Bousalem et al., 2009; Laney et al., 2012; Pahalawatta et al., 2008; Viswanathan et al., 1996). For the most part, EPRV pose a major threat in many agricultural crops because they can be transported vertically through the germ line to subsequent generations and can remain silent until conditions of environmental stress cause them to become active and cause disease (Pahalawatta et al., 2008).

Pararetrovirual replication

CaMV is a pararetrovirus and replication of CaMV occurs from the DNA genome that is transcribed in to RNA and then reverse transcribed in to DNA. This reverse transcription phase is unique to DNA containing plant viruses. The major phases of pararetroviral replication include transcription of viral DNA in the host nucleus and reverse transcription of these major transcripts in the cytoplasm. The viral DNA converts in to a supercoiled molecule in the host nucleus associating with histones and sealing its single stranded discontinuities which is then called a minichromosome (Pfeiffer and Hohn, 1983). This DNA is then transcribed in to terminally redundant RNA by using of host RNA polymerase. This step creates repeated sequences at the end of the RNA and is called pregenomic RNA (Hohn et al., 1997). In the host cytoplasm this pregenomic RNA can be involved in viral protein production or viral replication via reverse transcription. The tRNA methionine (tRNA ^{met}) act as a primer for initiation of reverse transcription (Borah et al., 2013).

Pararetroviral replication is involved with regulatory sequences. It is reported that in BSV intergenic regions in between ORFs contain polyadenylation signal and transcription start site that is considered as a promotor region of the genome. The pararetroviral promoter sequences are functionally known to bind plant transcriptional factors to facilitate viral transcription and replication. The BSV promotor sequence allow production of terminally redundant pregenomic RNA using host DNA-dependent RNA polymerase (Remans et al., 2005). The CaMV 35S promoter sequence is extensively studied and known as a constitutive promoter to transcribe whole CaMV genome (Haas et al., 2002). Studies on sequences downstream of a transcription start site of

pararetroviruses have shown that these sequences are involved in gene expression efficiency (Pauli et al., 2004). It is reported that downstream regulator sequences of *Rice tungro bacilliform virus* interact with nuclear binding proteins in rice protoplasts (He et al., 2002). Also functional promoter elements that are potentially involved in binding of plant transcriptional factors were identified in the BSV Cavendish strain. These regulatory sequences play a major role in replication and transcription of viral proteins (Remans et al., 2005). The presence of a poly A signal allows transcription termination and maturation of mRNA (Kazmi et al., 2015). Another regulatory sequence is the TATA box which is known to be involved in regulation of transcription. The function of a TATA box in badnavirus intergenic region could be for the initiation of transcription of pregenomic RNA, but extensive studies on these regulatory sequences have not been conducted (Zhang et al., 2015). A model system for badnaviral replication is presented in Fig. 2.

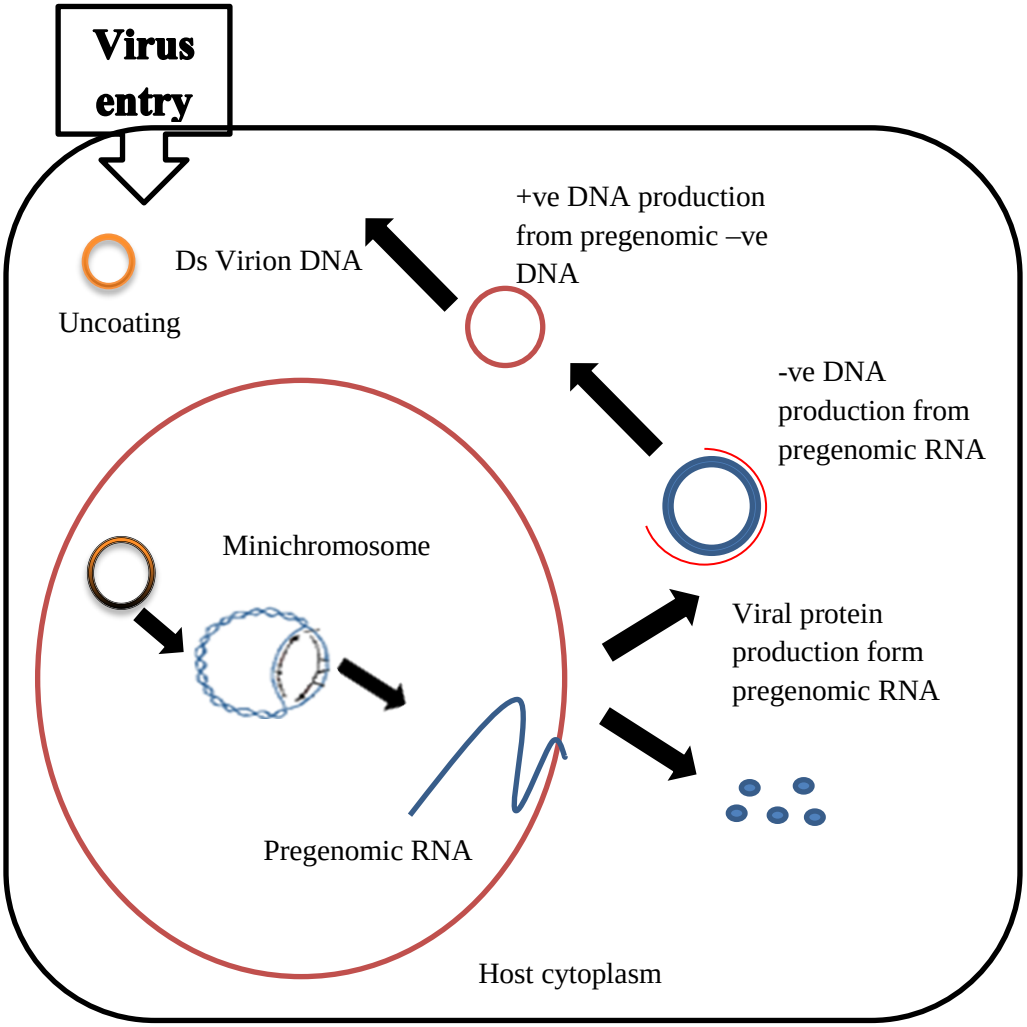

Fig. 2. Schematic diagram for badnaviral replication (Pfeiffer and Hohn, 1983)

Canna Yellow Mottle Virus

Among 25 species belonging to the genus Badnavirus, *Canna yellow mottle virus* (CaYMV) is considered one of the greatest threats to canna industry (Chauhan et al., 2015a; Rajakaruna et al., 2013). CaYMV was first reported in Japan in 1978 and North America in 1988 (Momol et al., 2004). More recently in 2008, this virus was reported in canna plants in Italy and Netherland where plants were observed to have CaYMV infection with or without symptoms (Marino et al., 2008). Disease symptoms can vary depending on their occurrence in red burgundy leaf canna varieties and green leaf canna varieties (Chauhan et al., 2015a). Symptoms caused by this virus in green leaf canna varieties include yellowing and mottling with veinal necrosis. In later stages, severe necrosis associated with death of the whole leaf can occur (Momol et al., 2004). Currently CaYMV diagnostics is based on primers developed by (Momol et al., 2004) for a genomic fragment of 565 bp reported in Genbank, which we report as an endogenous viral element associated with CaYMV (EVS-0036) (Momol et al., 2004). These diagnostic primers, CaYMV-3 and CaYMV-4 (Table 1) are widely used by researchers and diagnostic laboratories which, we learned through this proposed research, only detect CaYMV associated EVS while they fail to detect the entire CaYMV genome.

Family Potyviridae

Research by other members of our laboratory suggested that CaYMV infection often occurs in mixed infections with potyviruses. Therefore, for the purposes of this study it is important to briefly explain these viruses and our research involved careful diagnosis of plants for CaYMV as well as potyviruses.

The *Potyviridae* family of plant viruses is considered to be one of the largest positive single stranded RNA containing viral families (Oana et al., 2009). The *Potyviridae* family consists of seven genera: *Potyvirus* (type species; PVY), *Rymovirus* (type species; *Rygrass mosaic virus*), *Bymovirus* (type species; *Barley yellow mosaic virus*) *Macluravirus* (type species; *Maclura mosaic virus*), *Ipomovirus* (type species; *Sweet potato mild mottle virus*), *Brambyvirus* (type species; *Blackberry virus Y*) and *Tritimovirus* (type species; *Wheat streak mosaic virus*) (Oana et al., 2009).

The genus *Potyvirus* consist of a few economically important viral species, including *Potato virus Y* (PVY), *Bean yellow mosaic virus* (BYMV), and *Canna yellow streak virus* (CaYSV). Members of this genus are known to be transmitted by aphids (Andrew et al., 2012). Viral members of this genus have monopartite genomes with a particle modal length 650-900 nm. The single positive RNA genome is approximately 7.5 to 9 kbp in length. The only genus that consists of a bipartite genome is *Barley yellow mosaic virus* (*Bymovirus*), which contains two positive single strand RNA molecules.

Genus Potyvirus

Genus Potyvirus contains nearly 200 species with PVY as the type species (Spetz et al., 2003). Morphologically, these virions are non-enveloped flexuous filaments ranging from 680-900 nm in length and 11-13 nm in width. Members of the genus Potyvirus are reported to be the most destructive disease causing agents in agronomic and horticultural crops (Grabowski, 2010). In 2007 CaYSV was reported to cause foliar disease in canna. This virus is more closely related to *Johnson grass mosaic virus* (Monger et al., 2007). The positive strand RNA genome of potyviruses contains a single ORF with 5’ and 3’ untranslated regions (UTR). The 5’UTR is linked to a VPg protein

and adjacent to the 3’ UTR is a polyadenylated tail (PolyA) (Monger et al., 2007). The coding region of CaYSV genome is cleaved into ten mature gene products including P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP using three virus encoded proteinases: P1, NIa-Pro and HC-Pro (Monger et al., 2007). Typical genomic organization of CaYSV is presented in Fig. 3. CaYSV infection of canna is reported to be one of the major destructive threats to the canna industry. CaYSV symptoms include severe necrotic lesions, yellow streaking, mottling and discoloration in green leaf cultivars of canna (Monger et al., 2007).

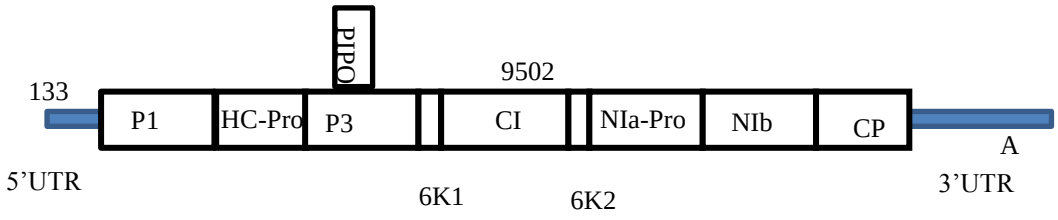

Fig. 3. Diagrammatic representation of the genomic organization of CaYSV indicating the ssRNA as a line. The open box represents the single ORF translated into a large polyprotein. Individual mature proteins are named inside the boxes. Cleavage sites recognized by virus encoded proteinases (Pro) are indicated by vertical lines through the large open box. Blue boxes at each end indicate the UTRs which contribute to regulation of translation and replication. Abbrev: P1- P1 protein, HC-Pro- Aphid transmission helper component, PiPo, internal protein, P3- P3 protein, 6K1/6K2- 6 kDA proteins, CI- cytoplasmic inclusion, NIa/NIb- nuclear inclusions a or b and CP- capsid protein (Monger et al., 2010)

CHAPTER III

EXPERIMENTAL DESIGN AND METHODOLOGY

Plant growth

Healthy and naturally infected canna rhizomes (cultivar 'Striped Beauty', 'Pretoria', 'Australia') were grown in the greenhouse for 8 weeks at 30ºC and 14 h light using SUNSHINE® LC1 growing mix. For mechanical inoculation experiments seeds of alternative host plants including *Phaseolus coccineus* L. var *Tenderstar* (Eden Brothers, Asheville NC), P. *vulgaris* L. var. *Roma* II, *Lunine albus* L. (Purcell Mountain Farms, Moyie Springs, Idaho), *Nicotiana benthamiana* and *Canna indica* 'Verdi' were grown in a growth chamber for 21 d at 23ºC with 12 h day light using special mix custom blend soil (Berger®, Saint-Modeste, QC Canada).

Extraction of nucleic acids for various applications

Extraction of total nucleic acid and RNA using the Maxwell 16 system for PCR and RT-PCR

Two leaf punches collectively weighing 70 mg were collected in pre-sterilized XXtuff® reinforced micro vials containing 2.0 mm zirconia beads (Biospec Products, Bartlesville OK) and frozen immediately in liquid nitrogen. Preserved tissues were stored for a short period of time at -80 ºC. Adequate grinding is critical to obtain a high yield of genetic material. Samples were taken from the -80 ºC freezer and placed on ice for extraction using the Maxwell® 16 LEV SimplyRNA Tissue Kit (Promega Corp, Madison WI, USA). Chilled 1-thioglycerol/homogenization solution (200 µl) provided in the kit was added to each sample. Tubes were loaded into the BeadBug® microtube homogenizer (Benchmark Scientific, Edison, NJ) and grinding was carried out at 16000 g for 2 min. Ground samples were centrifuged at 16000 g for 5 min. Then 200 µl of the supernatant was transferred to an Eppendorf tube and 200 µl of lysis buffer was added. Samples were vortexed, and then 400 µl of the lysate is transferred to the Maxwell® 16 LEV cartridge for processing in the Maxwell® 16 Instrument. To recover total nucleic acids (TNA), the DNAse I solution provided by the manufacturer was left out of the protocol. Elution was carried out using pre-chilled nuclease-free water. To recover RNA, the manufacturer protocol was followed. The NanoDrop ND-1000 (Thermo Fisher Scientific, Waltham, MA, USA) was used to check the quality and yield of TNA or RNA with an expectation of 260/280 value of 1.9 OD. Expected yield for extracted RNA is in between 5-10 µg per sample. For each experiment depending on required DNA/RNA yield different extraction methods were used. The sample quality was analyzed by 1% agarose gel electrophoresis using 2% ethidium bromide staining.

Total RNA extraction for northern blot analysis

Northern blot analysis was conducted to detect the presence of viral transcripts in infected canna plants. Total RNA extraction was conducted using Spectrum™ Plant Total RNA kit (Sigma-Aldrich Corp, St. Louis, MO, USA). Leaf tissue was ground using liquid nitrogen and 100 mg of grounded tissue sample was transferred into a 1.5 ml tube. Then 500 µl of lysis buffer with C_2H_6OS was added; samples were vortexed for 30 sec, incubated at 56℃ for 5 min, and centrifuged at 13000 g for 6.5 mins. The lysate was carefully separated and added to the blue filtration column provided with the kit. After centrifugation at 13000 g for 1.5 minus, the filtrate was mixed with 500 µl of binding solution and transferred to a red RNA binding column provided with the kit. After centrifugation for 1.5 min at 13000 g, the flow through was discarded and then the wash solution was added twice followed by centrifugation at 13000 g for 1.5 min. Each time the flow through was discarded. Elution buffer was used to recover RNA as instructed. For notherns, high yield (~10 µg) of RNA is required per sample. The expected RNA yield for this method is 20–60 µg. The RNA integrity was assessed based on absorbance ratio of 1.8 to 2.0 at 260/280 nm using NanoDrop ND-1000 Spectrophotometer (Nanodrop Technologies, Rockland, USA). RNA was stored at -80℃.

Genomic DNA extraction for Southern blot analysis

Plant DNA was extracted from 15 g of infected canna using the CTAB method (Ausubel et al., 1988). Leaves were ground in liquid nitrogen, 69 ml of warm extraction buffer [100 mM Tris-Cl, 50 mM EDTA, 500 mM NaCl, 1.25% SDS, 8.3 mM NaOH and $NaHSO_3$ and $NaS_2CN(C_2H_5)_2$ to 0.38%] was added, and samples were incubated at 65℃

for 20 min. Then 21.3 ml of 5M $C_2H_3KO_2$ was added and the mixture was inverted 5 times before centrifugation at 9000 g for 20 min (Beckman JLA 16250). The supernatant was filtered through Mira cloth and separated into equal volumes in 30 ml corex tubes. Nucleic acids were precipitated using 0.7 vol of isopropanol at -20℃ for 10 min and then dissolved in 6 ml of TE (pH 8.0) at 65℃. RNA was eliminated using 30 µl of 4 mg/ml RNase A at 37℃ for 10 min. Genomic DNA was extracted by mixing the solution with 6 ml of 2x CTAB buffer [200 mM Tris-Cl pH 8, 50 mM EDTA, 2M NaCl and 2% $CH_3(CH_2)_{15}N(Br)(CH_3)_3$] at 65°C for 15 min and then extracting with 24: 1 (v/v) chloroform: iso-amyl alcohol. The aqueous phase was transferred to a fresh tube and DNA was precipitated using 0.9 vol isopropanol. After incubation of 30 min at -20ºC, sample was centrifuged at 9000 g for 5 min. The resulting pellet was washed with 70% ethanol for 20 min at room temperature and then centrifuged at 5000 g for 5 min. The pelleted DNA was dissolved in 100 µl of TE buffer (pH 8.0). For Southern blot analysis sample DNA yield should be in between 10-15 µg. Expected yield using CTAB method is 15 µg. The purity of genomic DNA was assessed based on absorbance ratio of 1.8 to 2.0 at 260/280 nm using NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Rockland, USA) and verified by resolving in 1% agarose gel stained with 2% ethidium bromide.

Single, duplex and multiplex RT-PCR and PCR using plant extracted nucleic acids and plasmid preparation

Reverse transcription

Reverse transcription was employed throughout the study including, optimization of duplex/multiplex RT-PCR assays, cloning of CaYMV and endogenous viral sequences and screening of naturally infected canna plant populations for viral infection. In each experiment reverse transcription was carried out using the High Capacity DNA Reverse Transcription kit (Thermo Fisher Scientific, Waltham, MA, USA), random primers and 500 ng TNA or RNA, according to the manufacturer's protocol. The final cDNA was determined using the NanoDrop ND-1000 spectrophotometer.

Singleplex, duplex and multiplex RT-PCR

Singleplex or duplex PCR reactions were performed to detect BYMV, CaYMV-like sequences and CaYSV using approximately 2 µg of cDNA and GoTaq Flexi DNA polymerase (Promega Corp, Madison WI, USA) with 5 µM primers listed in Table 1. The resulting amplicons of 382, 565 and 695 bp correspond to BYMV, CaYMV-like sequences and CaYSV genomes, respectively (Chauhan et al., 2015a; Rajakaruna et al., 2013). The thermal cycling conditions for singleplex RT-PCR include one cycle at 95℃ for 5 min; 25 cycles at 95°C for 1 min, annealing temperature was determined based on each primer set (for Canna F1/R1 50°C, CaYMV-3/4 52℃ and BYMV F/R 52℃) for 45 sec, 72°C for 1 min, followed by a final extension at 72°C for 7 min. For duplex RT-PCR thermal cycling conditions include one cycle at 95℃ for 5 min; 25 cycles at 95°C for 1 min, 50°C for 45 sec, 72°C for 1 min, followed by a final extension at 72°C for 7 min. Sample reactions were 20 µl. For optimization of multiplex RT-PCR detecting CaYSV, endogenous CaYMV-like sequences and BYMV, temperature-gradient PCR was conducted using 50 to 55.8℃ and optimization of $MgCl_2$ was conducted using varying $MgCl_2$ concentrations (2.0, 2.5, 3.0, 3.5, 4.0, 4.5 and 5.0 µM).

The final thermal cycling conditions for multiplex RT-PCR were one cycle at 95°C for 2 min; 35 cycles at 95°C for 45 sec, 50°C for 45 sec, 72°C for 45 sec, followed by a final extension at 72°C for 7 min.

The control plasmid named pYM0036P1 contains a fragment of the EVS-0036 sequence (CaYMV-like sequence) inserted into the p-GEMT Easy vector. The pYSOK7672 contains a 1,827-bp fragment of the CaYSV genome (OK isolate) which extends from nucleotide position 7,672 to 9,502, inserted into the pCR2.1 vector. The pBYMV382 which contains a 382-bp fragment of the BYMV coat protein (nucleotide position 8,949 to 9,329) inserted into pGEM-T Easy vector. These plasmids were used as controls for RT-PCR experiments throughout this study using the same thermal cycling conditions. For each experiment 1 ng of plasmid diluted in nuclease free water was used (Chauhan et al., 2015b; Rajakaruna et al., 2013). PCR reaction products were observed alongside the RT-PCR reaction products using 2% agarose gel electrophoresis and 2% ethidium bromide stain.

Table 1. PCR primers used for diagnostic detection of viruses to identify naturally infected plants for purification and sequencing of CaYMV infected plants for purification and sequencing of CaYMV

Virus	Primer	Amplicon size (bp)	Primer sequences (5'-3')	Melting Temperature (°C)	References
CaYSV	Canna-F1	695	GAT CTA ACG CAA CAT ATT CGA AAG	F:52.2	(Monger et al., 2007)
	Canna-R2		GAT CAT CCA TCA TTA CCC AAT AC	R:51.0	
Endogenous CaYMV	CaYMV-3F	565	GAC TTC CTG GGT GCA ACA AT	F:55.6	(Momol et al., 2004)
	CaYMV-4R		TCT GTG CAA TCT TGG CGT AG	R:55.2	
BYMV	BYMV-F	382	GTG AAT GGA CAA TGA TGG ATG	F:51.6	(Chauhan et al., 2015b)
	BYMV-R		CAC CAT TGA CAT CTC CTG CTG	R:55.8	

Virion DNA purification for next generation sequencing

Virion DNA extraction A:

For this method we employed the virion DNA extraction protocol adopted by Zhuang et al., (2011). Leaves from naturally-infected canna (cultivar 'Striped Beauty') were harvested from greenhouse grown plants for this procedure. Plants were previously identified by another laboratory member, who also harvested the leaves for virion DNA extraction and stored these leaves in bags in the -80°C freezer. These plants were identified by PCR using CaYMV-3 and CaYMV-4 primers (Table 1), which were the only available diagnostic primers at the time this study began (Momol et al., 2004). Approximately 80-100 g of leaf tissue was ground in 4 volumes of grinding buffer at [containing 0.5% (w/v) Na_2SO_3, 2% (w/v) PVP-40, and 10 mM EDTA·Na_2] 4°C and filtered through three layers of gauze to remove plant debris. The supernatant was equally distributed among six rotor bottles and centrifuged at 7500 g for 20 min at 4°C (Beckman JA-14). The supernatant was recovered and centrifuged for a second time at the same speed for 20 min. The final supernatant 256 ml was transferred to a fresh 1 L sterile beaker. Then 1.5% (w/v) NaCl and 7.0% PEG (w/v) was added to the supernatant. This suspension was kept static overnight at 4°C and then centrifuged at 7500 g for 20 min at 4°C (Beckman JA-14). Then the pellet was recovered and re-suspended in 15 ml of suspension buffer [10 mMol/L sodium phosphate buffer pH7.2, 0.9% (w/v) NaCl, Triton X-100], and transferred to a 50 ml sterile beaker with a stir bar. The suspension was stirred for 1 hour at 4°C, then transferred to centrifuge tubes, and centrifuged at 7500 g for 20 min at 4°C (Beckman JA-14). The final supernatant was recovered and loaded onto a 20% sucrose in 10 mM NaH_2PO_4 (pH 7.2) cushion, which was prepared in thin

wall, 38.5 ml ultracentrifuge tubes. These tubes were loaded into a Beckman SW32 Ti and centrifuged at 60,000 g for 3 h at 4°C. Three fractions were recovered from these tubes: a green pellet, the middle yellow layer, and a white layer at the top of the cushion. These fractions were stored separately at 4°C overnight.

Viral DNA was extracted from the above recovered fractions using the GeneJET plasmid purification kit (Thermo Fisher Scientific, Waltham, MA, USA). One ml of re-suspension buffer and lysis solution was added and mixed thoroughly by inverting tubes six times until the solution became viscous and slightly clear. Then one ml of neutralization solution was added and mixed by inverting the tube 6 times. Then samples were centrifuged at 12000 g for 5 min in a Beckman JA 17. The supernatant was transferred to the spin column provided in the kit and centrifuged for 1 min. Then 500 µl of wash solution was added to the column and centrifuged for 1 min and the flow through was discarded. The spin column was transferred to a new collection tube and 50 µl of nuclease free water was added and incubated at room temperature for 2 min and centrifuged at 13000 rpm for 2 min. Then each sample was treated with 12 units of DNase I and 100 mg/ml proteinase K with 1% SDS. Tubes were incubated at 37°C for 30 min and then nucleic acids were extracted with phenol:chloroform: isomylalcohol. DNA was precipitated with 2.5 vol ethanol and 1/10 vol 3M NaOAC pH 5.5 by centrifuging at 4000 rpm for 5 min. The pellet was washed with 70% ethanol and centrifuged at 9000 rpm for 20 min. Each pellet was dissolved in 1 ml of TE buffer pH 8.0. Viral DNA integrity was assessed using 1% agarose gel electrophoresis and 2% ethidium bromide stain. The yellow fraction produced the best DNA product; other fractions seemed empty

of DNA and therefore DNA extracted from this fraction was sent for sequencing at OSU Biochemistry and Molecular Biology Array and Bioinformatics Core Facility.

Virion DNA extraction B:

Virions were isolated using a modified protocol adapted from (Covey et al., 1998). First we conducted multiplex RT-PCR on a population of naturally infected canna ('Striped Beauty') in our greenhouse to identify plants that were likely infected with CaYMV using primers in Table 1. Plants that tested positive with the CaYMV-3/4 primer set were selected for virion DNA extraction. Fresh canna leaves (100g) were harvested and ground in a homogenizer at 4°C using 200 ml grinding buffer [0.5 M NaH_2PO_4 , 0.5M Na_2HPO_4 (pH 7.2) and 0.5% (w/v) Na_2SO_3]. The homogenate was placed in a beaker with 12 g of urea and 50 ml 10% Triton X-100 and kept static at 4ºC overnight. The homogenate was centrifuged at 4°C at 4000 g for 10 min and then filtered through 4 layers of cheesecloth. The final volume of the homogenate was measured, equally divided into 16.5 ml polyallomer centrifuge tubes, and centrifuged for 2 ½ hours in a Beckman Ti70 at 40,000 g. The pellets from each tube were recovered and resuspended in 1 ml of ddH_2O and centrifuged again at 8000 g for 10 min (Beckman JA-17). The supernatants were transferred to 38.5 ml thin wall tubes and re-centrifuged at 136,000 g for 2 h at 4 °C (Beckman SW32 Ti). The resulted brown pellet and upper flocculent white layers from each tube were pooled. These fractions contained intact virions and likely contaminating cellular DNA. To eliminate contaminating cellular nucleic acids, the fractions were treated with DNAse I for 10 min at 37℃. Then the virions were treated with 40 µl of 2 µg/ µl proteinase K at 37 °C for 15 min to remove contaminating DNAse as well as virion coat proteins to expose the circular virion DNA.

Both reactions were stopped by adding 4 µl of 0.5M EDTA pH 5.5. The virion DNA was extracted with phenol:chloroform:isomylalchol, and both fractions were then pooled and precipitated with isopropanol. The pellet was washed with 70% ethanol, centrifuged at 9000 g for 20 min (Beckman JA-17) and then dissolved in 1 ml of TE buffer pH 8.0 (Covey et al., 1998). To verify the viral DNA integrity, agarose gel electrophoresis was carried out using 1% agarose gel stained with 2% ethidium bromide. The samples were sent for sequencing at OSU Biochemistry and Molecular Biology Array and Bioinformatics Core Facility.

Next generation sequencing (NGS) and bioinformatics analysis of data

Purified virion DNA was sequenced using a Roche 454-Junior™ Genomic Sequencer by Dr. Hong at the OSU Array and Bioinformatics Core Facility. The first sequencing data set obtained from virion DNA isolated using protocol A created 157765 raw reads while virion DNA isolated from protocol B resulted 188,626 raw reads with a median read length of 441 bp. Roche Genome Assembly software version 2.7 was used by Dr. Peter Hoyt, Director of the OSU Array and Bioinformatics Core Facility, to assemble 13,269 contigs. He also used FastQC and Timmomatic software to clean row reads and to eliminate contaminating vector and primer sequences. He used Newlber v2.7 software to reduce the contigs and reads to 8,243 cleaned contigs. NCBI MegaBlast was used to identify contigs with viral origin. Dr Hoyt provided us with the clean contigs used to further study and characterize viral sequences. This research focused on three contigs: CaYMV, EVS-0036, and EVS-00206 contigs.

For each badnaviral contig that was recovered from a MegaBlast search, primers were developed using Primer3 software for cloning and sequencing purposes. The insert

sequences were trimmed for vector contamination using NCBI vector screen program. The trimmed sequences were uploaded to SDSC Biology Workbench and analyzed by aligning with each other as well as with the candidate individual contigs. Multiple cloned sequenced were aligned using MEGA6 software and ClustalW. The Expasy Translation tool was used to predict translation products for CaYMV, EVS-0036, and EVS-00206 and these products were compared with other badnavirus proteins using the Uniport Knowledgebase Protein Blast analysis tool.

The full genome sequences of 30 badnaviral species were obtained from NCBI database (Table 2). All phylogenetic analyses were carried out using MEGA6 software. Multiple sequence alignments and several maximum likelihood trees were generated were using Muscle in MEGA6.

Table 2. Badnavirus species used in phylogenetic analysis

Name	Acronym	GenBank
Banana Streak CA virus	BSCAV	KJ013511
Bougainvillea spectabilis chlorotic vein-banding virus	BSCVB	EU034539
Banana streak GF virus	BSGFV	KJ013507
Banana streak IM virus	BSIMV	HQ593112
Banana streak MY virus	BSMYV	KJ013509
Banana streak OL virus	BSOLV	KJ013506
Banana streak VN virus	BSVNV	KJ013510
Canna yellow mottle virus	CaYMV	pending
Commelina yellow mottle virus	ComYMV	X52938
Cycad leaf necrosis virus	CyLNV	EU853709
Cacao swollen shoot virus	CSSV	L14546
Cassava vein mosaic virus	CVMV	U59751
Dioscorea bacilliform virus	DiBV	DQ822073
Canna yellow mottle virus fragment in NCBI	CaYMV	KJ561635
Dracaena mottle virus	DrMV	DQ473478
Fig badnavirus 1	FBV-1	NC_017830
Fig badnavirus 1 isolate Arkansas	FBV-1Ar	JF411989
Gooseberry vein banding virus	GVBaV	NC_018105
Grapevine vein-clearing virus	GVCV	JF301669
Kalanchoe top-spotting virus	KTSV	AY180137
Lucky bamboo bacilliform virus	LBBV	NC_009568
Pineapple bacilliform comosus virus	PBComV	GU121676
Pelargonium vein banding virus	PVBV	GQ428155
Petunia vein clearing virus	PVCV	U95208
Piper yellow mottle virus	PYMoV	KC808712
Rice tungro bacilliform virus	RTBV	X57924
Rubus yellow net virus	RYNV	KF241951
Sugarcane bacilliform IM virus	SCBIMV	NC_003031
Sugarcane bacilliform Mor virus	SCBMorV	M89923
Sweetpotato badnavirus B	SPBV-B	FJ560944
Sweet potato caulimo-like virus	SPCLV	HQ694978
Taro bacilliform virus	TaBV	AF357836

Transmission electron microscopy

Transmission electron microscopy was used to visualize badnavirus and potyvirus virions in leaf extracts from naturally infected canna ('Striped Beauty') or virus inoculated *P.vulgaris* L. var. *Roma* II. One g of leaf was ground in phosphate buffer (0.5M Na_2HPO_4, pH7.2) and then centrifuged in 5000 rpm for 5 min in a Beckman JA-17 centrifuge. The supernatant was recovered, diluted four fold and then mounted onto formvar-coated grids and incubated for one minute at 25 °C. Then phosphotungstic acid (PTA) stain was added and incubated for 30 second at 25°C. Excess stain was removed by using a sterile Whatman® filter paper. Grids were viewed using a JEOL JEM-2100 transmission electron microscope system with an EDAX Genesis 2000 EDS system (JEOL Ltd., Tokyo, Japan) located at the Oklahoma State University Microscopy Laboratory by Terry Colberg and Lisa Whitworth.

PCR and RT-PCR amplification and cloning of CaYMV and badnavirus related sequences in virion preparation A and B

Virion preparation A produced a short contig which we named endogenous virus sequence (EVS)-0036. We developed five badnavirus primers, (DWF1, 50R, 350R, CaYMV-DF and CaYMV-DR) (Table 3). These primers were used for de novo PCR amplification of badnavirus sequences to expand the 5' end of the EVS-0036. We assumed that EVS-0036 was a fragment of an entire viral genome which we were not successful in recovering, perhaps due to the DNAse treatment in virion preparation A. These primers were developed as degenerate badnavirus primers that should recognize related upstream badnavirus sequences. These badnavirus primers were developed by

alignment of the *Sugarcane bacilliform virus* (SCBV; FJ439817.1) and *Banana streak virus* (BSV; NC008018.1) genome sequences and the short 0036 contig (Table 3). The virion preparation B was conducted 6 months later than virion preparation A and produced the entire CaYMV genome and two shorter elements that we cloned. The shorter elements are named EVS-0036 and EVS-00206. Table 3 contains the PCR primers used for RT-PCR amplification of EVS-0036, CaYMV and EVS-00206 in these sets of experiments.

PCR and RT-PCR was carried out using TNA or RNA isolated using the Maxwell 16 LEV RNA Extraction Kit. All PCR reactions were conducted using 5µM each of forward and reverse primer with 35 cycles of PCR amplification. Cycling conditions include denaturation at 95°C for 60 sec, annealing at 50°C for 45 sec to 1 min, extension at 72°C for 1-2 min with a final extension at 72°C for 7-10 min (GoTaq, Promega Corp, Madison WI, USA).

Cloning and plasmid preparation

Each PCR products were gel purified using Wizard SV gel and PCR Clean-Up System (Promega Corp, Madison WI, USA). The quality of the gel-purified PCR products were assessed based on the absorbance ratio of 1.8 to 2.0 at 260/280 nm using Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies, Rockland, USA). Then these gel purified PCR products were used for ligation into pGEM-T Easy vector (Promega Corp, Madison WI, USA) and transformed into JM109 High Efficiency Competent Cells (Promega Corp, Madison WI, USA). Transformed cells were plated on LB-agar media containing 100µg/ml ampicillin, 50µg/ml X-gal, and 50µg/ml IPTG. The plates were incubated at 37 °C for 16 to 18 hours. Five to ten white colonies containing

ligated CaYMV DNA were selected, re-streaked onto fresh plates and grown overnight at 37°C in liquid LB media containing 100µg/ml ampicillin. The plasmids prepared for this study are presented inTable 4.

Plasmid DNA extraction was conducted using a Wizard MiniPrep kit (Thermo Fisher Scientific, Waltham, MA, USA). Plasmids were subjected to restriction digestion using E*coR*I to verify successful ligation. (Promega Corp, Madison WI, USA). Around 500 ng of plasmids were treated with E*coR*I and incubated at 37 C for 2 h. The enzymatic reaction was stopped by heating at 85 °C for 5 min and then 1% agarose gel electrophoresis was used to verify that digested products were of the correct sizes. Candidate plasmids were sent for sequencing at the OSU Biochemistry and Molecular Biology Array and Bioinformatics Core Facility. Then these sequences were analyzed using SDSC biology work bench.

Table 3. Primers developed for PCR/RT-PCR and cloning of sequence based on contigs developed from NGS of virion preparation B.

Primer	Primer sequence	Amplicon size (bp)	Sequence position	Plasmid name
degenerate Primers for extension of 0036 sequence				-
DWF1	5'MTT TGA AAG RTG GGA R3'	1762	-	-
50R	5' CCC CGA TCC AGA TCT TTC C3'			
DWF1	5' MTT TGA AAG RTG GGA R3'	2063	-	-
350R	5' GAA CAG CTG GGA GTC CAC3'			
CaYMV-DF	5'GCA GAG AGG CAG CAT ATC AA3'	1990	-	-
CaYMV-DR	5'AAT ACT GCT CCC CGA TCC AG3'			
cloning of EVS-0036 sequence				
301F	5'ATGGACGAGGAGAAGGGCTA 3'	1296	372-1705	pYM0036P1
1621R	5'GGATGGCTTGTTCTGTGCAA 3'			
1081F	5'TAGACTTCCTGGGTGCAACAA3'	1298	1147-2477	pYM0036P2
1930R	5'AGTGTATGACTCGATGCCTAGC3'			
0036F	5'AAGCTAAATGAAACTAAGGGA3'	585	1-585	-
0036R	5'GCTTCCGCATGTCTTCTTTC3'			
0036-86F	5'CATGTGATACGCCACCTAAA3'	316	2012-2517	-
0036-86R	5'GAACTGGAACTCTTACAAACTCACA 3'			
cloning of CaYMV genome				
0028 1F	5'GTGAGAGAATCGGAGGAGGG 3'	1366	2880-4221	pYM0028Zn
0028 1R	5'GATAATCTCATGCTGTGCTAAAG 3'			
0028 2F	5'CAGCAAGATAGGACGTTCATTC 3'	1270	1844-3093	pYM0028OR
0028 2R	5'CGTGATCTTCTCCATCTCTC 3'			F3b
0028 3F	5'ACAGCAACAGACGATCCAGA 3'	1269	844-2089	pYM0028OR
0028 3R	5'CTCCAGAAATCCTTAATTGTCATC 3'			F1UF2
0028 4F	5'CGAGGTTGAAGGTATGTCTT 3'	1058	11-1052	pYM0028OR
0028 4R	5'CTTCTCAAGCCTGTGGAT 3'			FUF1
0028 5F	5'CATCCCAGAACAAGCCTTCG 3'	1039	6063-108	pYM0028OR
0028 5R	5'GACATTCGACTTATCAACCAA 3'			F3c
0028 6F	5'CAAACAGCAAGCACAGGACA 3'	1313	4983-6227	pYM0028OR
0028 6R	5'GGCCTCCTTGTCAGAACG 3'			F3a5'
0028 7F	5'CGAGAAGACAAGAGCGGG 3'	961	4249-5202	pYM0028OR
0028 7R	5'CCATGGCTACCTGATGAAAAC3'			F3a3'
cloning of EVS-00206 sequence				
00206F	5'CCATGTTTTTGCGGTTTTCC 3'	752	88-820	-
00206R	5'AAGCTAAATGAAACTAAGGGA 3'			
00206F1	5'CTTTCTCGCTTTGCCTTTCTT3'	897	534-1411	-
00206R1	5'CGACTGCAATGCTTGTTGAA3'			

Table 4. Plasmids containing viral sequences used in this study

Plasmid name	Insertion size (bp)	Virus/sequence
pYM0036P1*	1296	EVS-0036
pYM0036P2	1298	
pYM0028Zn	1364	
pYM0028ORF3b	1279	
pYM0028ORF1UF2	1269	
pYM0028ORFUF1	1058	CaYMV
pYM0028ORF3c	1032	
pYM0028ORF3a5'	1313	
pYM0028ORF3a3'	979	
pYM00206	752	EVS-00206
pBYMV382*	382	BYMV (Chauhan et al., 2015c)
pYSOK7672*	1827	CaYSV (Chauhan et al., 2015c)

Note- For duplex RT-PCR pYM0036P1 and pYSOK7672 was used. For multiplex RT-PCR pYM0036P1, pBYMV382 and pYSOK7672 was adopted.

Mechanical inoculation

Mechanical inoculation of *P. vulgaris L.* var. 'Roma II', *C. indica*, *N. benthamiana* and *P. coccineus L*. var 'Tenderstar' was carried out using naturally infected canna leaf sap. Inoculum sources were identified and segregated by Mr. Austin Gimondo using multiplex RT-PCR to characterize plants that are infected with BYMV, CaYSV, and EVS-0036 and simplex PCR to detect CaYMV. Plants were segregated as inoculum sources for these experiments. One g of infected canna leaf tissue was ground using 5 vol. of phosphate buffer (pH7.2) and centrifuged at 5000 g for 10 mins to remove plant debris. Depending on the experiment canna leaf tissue was chosen such as CaYMV only, CaYSV and EVS-0036 only. For mixed inoculation experiments, equal volumes of CaYSV only sap was mixed with EVS-0036 sap. Two leaves per test plant were dusted with carborundum. Prepared plant sap was applied on to the test plant leaf using a cotton swab by gently rubbing in circular motion pattern. Negative controls were treated as same as test plants using only dd H_2O. These plants were kept in an growth chamber for 4 weeks for symptom development and symptoms were recorded.

Southern and northern blot analysis

Probe Preparation

The North2South Biotin Random Prime Labeling kit (Thermo Fisher Scientific, Waltham, MA, USA) was used for probe preparation. Plasmids pYM0036P1, pYM0028ORF3a3'and pYSOK7672A containing CaYMV genome fragment, EVS-0036 genome fragment and CaYSV genome fragment were digested with E*coR*I and the relevant genome fragments were gel purified. Then 100 ng of gel purified DNA was

diluted to the final volume of 24 µl and denatured by boiling for 5 min followed by quick freezing in a dry ice ethanol bath for 5 min. The denatured DNA was thawed on ice and briefly centrifuged to collect the contents at the bottom of the tube. Then 10 µl of 5x dNTP mixes, 5µl of 10x reaction buffer and 1µl of Klenow fragment was added to a final volume of 50 µl. The contents of the tube was mixed by briefly vortexing and then incubated at 37℃ for 60 min. Finally, the enzyme was inactivated by adding 2µl of 500 mM Na-EDTA (pH 5.5) and the probe was ethanol precipitated to remove unincorporated nucleotides. The probe was often stored in pre-hybridation buffer at 4°C for up to one month. To test the probe labeling efficiency, a positively charged nylon membrane was carefully taken and soaked in 1x TE buffer. A dilution series of the probe and the positive control was done by dissolving 2µl of probe in 10µl of 1x TE buffer, preparing 7 dilutions. The membrane was blotted on clean filter paper to remove excess TE buffer and 2 µl of the dilution series of the labeling reaction was spotted on the membrane. Then this membrane was subjected for UV crosslinking.

Visualization of immobilized nucleic acids

To detect immobilized nucleic acids, a chemiluminescent nucleic acid detection module kit (Thermo Fisher Scientific, Waltham, MA, USA) was used. The blocking buffer and 4x wash buffer was warmed in a 42°C water bath. Then 16 ml of blocking buffer was added to the membrane and incubated with shaking for 15 min. Then blocking buffer was decanted and 16 ml of conjugate blocking buffer (prepared by mixing 50 µl of stabilized streptavidin-horseradish peroxidase conjugate with 16 ml of blocking buffer) was added and then incubated for 15 min with gentle shaking. The membrane was transferred to a new bottle and 20 ml of 1x wash buffer was added and incubated for 5

min with gentle shaking. This step was repeated for four more times. The membrane was transferred to a new bottle and 30 ml of substrate equilibration buffer was added and incubated for 5 min with gentle shaking. The membrane was removed from the substrate equilibration buffer and placed in a clean container with substrate working solution. The membrane was incubated for 5 min in the dark, then wrapped in a clean plastic wrap and exposed to UV light for observation (FluorChem™ E System).

Restriction digestion of genomic DNA

For Southern blotting, 10 µg of genomic DNA was restricted digested using three different restriction enzymes such as: E*co*RI, B*am*H1 and H*ind* III (Ausubel et al., 1988). Genomic DNA from a healthy control plant (Tangelo variety) was used as a negative control. The following conditions for restriction digestion were used: 15 ul of 10x Restriction Digestion buffer, 1.5 µl 10 mg/ml RNaseA, 1.5 µl 10 mg/ml BSA, 5 µl restriction enzyme, and water to a final volume of 150 µl. The reaction was incubated at 37ºC for 1 hour. An additional 50 units of enzyme was added and incubated for another hour. Ethanol precipitation was carried out and DNA was resuspended in 50 µl TE.

Gel electrophoresis and blotting

Ten µg of undigested and digested genomic DNA isolated from variety 'Striped Beauty' and 'Pretoria' was resolved in a 1% agarose gel for approximately 4 hours at 35 volts (Ausubel et al., 1988). For northern blotting desaturating RNA gel was prepared. Before preparation of RNA gel tools were socked in 0.5 M NaOH for 20 min and rinsed with DEPC treated water. Then 1% agarose/formaldehyde gel was prepared and allowed to cool in a 60℃ water bath and ten ml of 10x MOPS/EDTA buffer (MOPS 41.8 g/1,

NaOAc 4.1 g/ l, 0.5 M EDTA 20 ml/ l) and 5.1 ml of 37% formaldehyde was added. Gel was allowed to solidify and approximately ten µg of RNA was loaded to each well. The RNA gel electrophoresis was carried out for 4 hours at 100 volts using 1x MOPS/EDTA as the buffer. A photograph of the gel was taken and the dimension was measured. Then a positively charged Hybond N+ membrane was carefully cut to the gel dimensions and soaked in 2x SSC buffer. The gel was placed in a slightly larger container with 250 ml of denaturation solution [NaOH 20 g/ l, NaCl 58.4 g/ l and add ddH_2O up to 1 l] and incubated for 15 min with shaking. This step was repeated, denaturation solution was discarded, and the gel was washed with deionized water. Then 250 ml of neutralization solution was added and incubated with shaking for 30 min. This step was repeated. The gel blot chamber was filled with 10x SSC buffer and a glass support was placed on top of the chamber. A wick of 3 MM blotting paper was immersed in 10x SSC and placed on top of the glass support in a manner that two ends of the wick completely touch the buffer in the gel blot chamber. Then the gel was placed on top of the wick upside down being sure there were no air bubbles trapped in between. Then Hybond N+ membrane was placed on top of the gel and two 10x SSC soaked blotting papers were paced on top of the membrane. A stack of paper towels were placed on top of the blotting papers. A 50 g weight was placed on top of the blot and the setup was kept overnight at 25℃ on the bench top.

Washing and hybridization of the membrane

The blot was disassembled and the membrane was inverted and placed in 2x SSC for 2 min (Ausubel et al., 1988). Then DNA on the membrane was cross-linked using a UV cross linker for 120 seconds. The membrane was then transferred to a glass roller

bottle with 10 ml of pre-pre-Hybridization solution (25x SSC 0.8ml, 10% SDS 1 ml and ddH_2O 18.2 ml) and incubated at 60˚C with 1 hour shaking. Then pre hybridization solution [formamide 5ml, 25x SSC 2 ml, 2M Na-P buffer (pH 6.8) 0.25 ml, Denhardt's 100x 0.5 ml, 10% SDS 0.2 ml, salmon sperm DNA 10 mg/ml 200 µl and ddH_2O 1.45 ml] was used to replace the prior solution and the membrane was incubated at 42°C for 4 hours. Then the pre-hybridization solution was discarded and the hybridization solution (formamide 2.5 ml, 25x SSC 1ml, 2M Na-P buffer pH 6.8 0.05 ml, Denhardt's 100x 0.05 ml, 10% SDS 0.05 ml, salmon sperm DNA 10 mg/ml 50 µl) was added to the membrane with the probe and incubated overnight at 42° C. For detection of the immobilized nucleic acids, the chemiluminescent nucleic acid detection module was used (Thermo Fisher Scientific, Waltham, MA, USA). Then the membrane was incubated for 5 min in the dark, then wrapped in a clean plastic wrap and exposed to UV chamber for observation (FluorChem™ E System).

CHAPTER IV

RESULTS AND FINDINGS

Development of a diagnostic PCR test for CaYMV

The full genome of CaYMV has never been reported before, however a small fragment of 565 bp has been reported in Genbank and diagnostic primers named CaYMV-3 and CaYMV-4 have been published for use in diagnosis (Momol et al., 2004) (Table 1). Our laboratory used these primers to develop a multiplex diagnostic PCR for detection of three viruses infecting cannas: CaYSV, BYMV and CaYMV. These three viruses are a significant threat to the global trade of canna (Rajakaruna et al., 2013). In 2016 we determined that the CaYMV-3 and CaYMV-4 primers detect endogenous sequences named EVS-0036 and EVS-00206, and does not detect the entire CaYMV genome and we are now working to correct the record. For this chapter, we will refer to the sequences detected by CaYMV-3 and CaYMV-4 as 'endogenous CaYMV sequences'. The goal was to develop duplex RT-PCR to detect 'endogenous CaYMV' sequences and CaYSV together in infected canna plants. Since CaYMV has a DNA genome but replicates through an RNA intermediate, and CaYSV has an RNA genome, we postulated that by isolating total nucleic acids (TNA) and carrying out RT-PCR, we

could detect both CaYMV and CaYSV in a duplex PCR reaction, which could be used for diagnostics in the greenhouse. This would benefit screening for clean germplasm needed for the industry. PCR conditions were optimized including annealing temperature and magnesium concentration in order to increase the sensitivity of the assay.

As a first step of developing diagnostic PCR assay, positive controls for each virus were generated. Viral sequences were obtained from greenhouse-grown canna plants that were known positives for CaYSV and endogenous CaYMV and these sequences were cloned into pGEM-T and confirmed via sequencing. Plasmid positive control for endogenous CaYMV contains 1296 bp section (Fig.4). This plasmid was named as pYM0036P1. Two other plasmids that contain fragments of BYMV and CaYSV genome were developed by other members of the lab and these were named as pBYMV382 and pYSOK7672 (Table 4).

As the second step, PCR primers were tested for singleplex PCR using plasmids and TNA samples to identify the best PCR conditions (these primers are listed in Table 1). High -yield TNA was obtained from greenhouse grown canna cultivar 'Striped Beauty' using 70 mg of starting tissue. The average TNA obtained from fifty 'Striped Beauty' leaves was 150.15 ng/µl with 260/280 ratio values of 1.98 to 1.74 indicating purity of extracted TNA. Regarding TNA integrity, the Maxwell kit maintained intact RNA as inferred from the 2:1 proportion of 28S to 18S ribosomal RNA (rRNA). Visual confirmation of RNA and DNA was done using an agarose gel (Fig.5). One significant DNA species appeared at the top of the gel, just above the ladder and the DNA appeared

to be intact and the proportion of genomic DNA relative to RNA across samples was consistent.

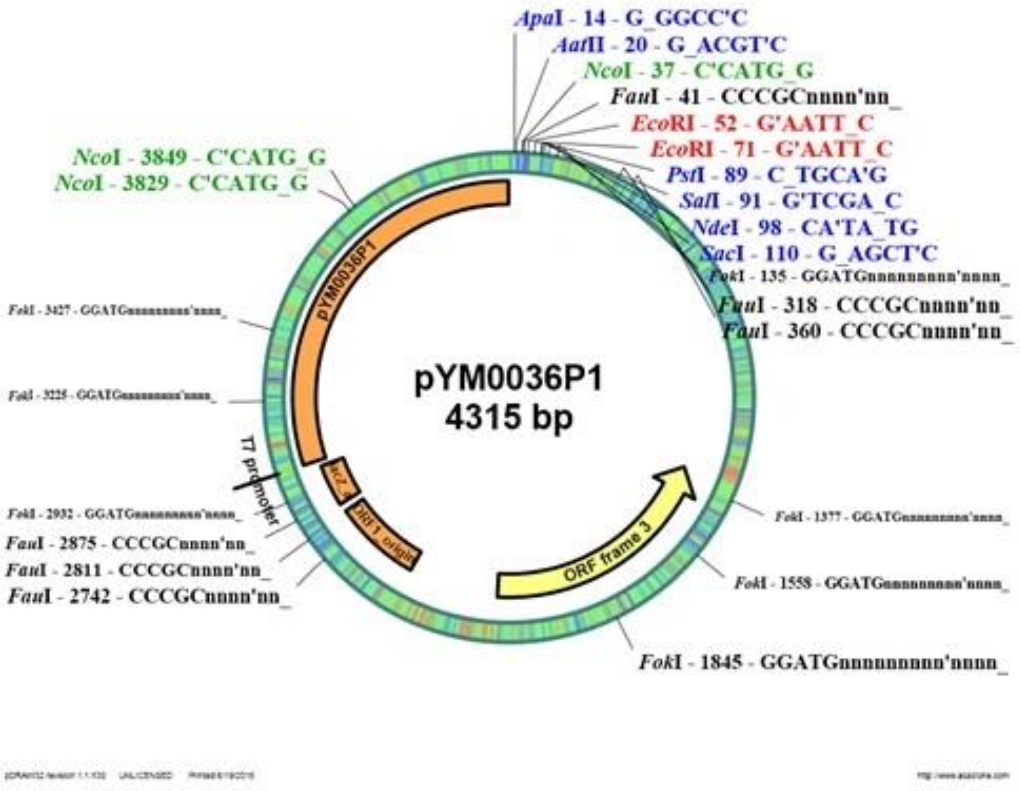

Fig. 4. Diagrammatic representation of plasmid (pYM0036P1) containing fragment of 0036 sequence inserted to p-GEMT Easy vector. Restriction sties and promoters of p-GEMT Easy vector is also indicated

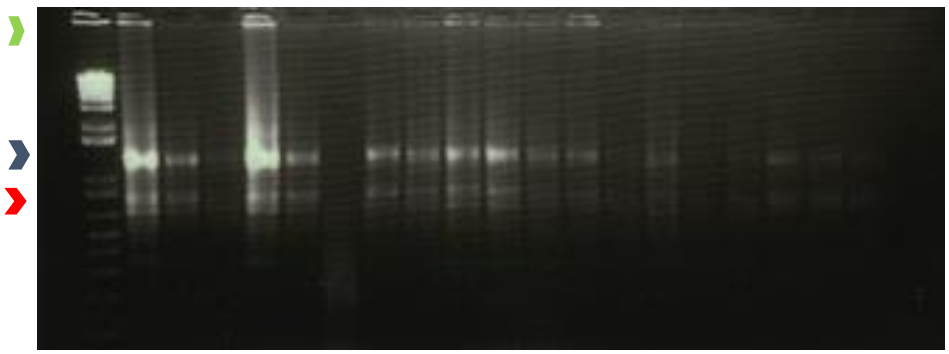

Fig. 5. Gel picture representing optimization of Duplex RT-PCR, 2% agarose gel indicating nucleic acid integrity following extraction with Maxwell kit. Green arrow head identifies plant genomic DNA, blue and red arrow heads identify 28S and 18S RNA respectively.

Duplex RT-PCR screening of 50 'Striped Beauty' plants

While working to develop duplex detection of CaYSV and endogenous CaYMV sequences, others in the laboratory were working to develop duplex and multiplex diagnostics to detect various other combinations of viruses. Information was shared among all who were working on these assays. For detection of CaYSV and CaYMV, primers were mixed in 0.5 µM final concentration with 50 °C annealing temperature. The standard RT-PCR conditions used for singleplex reactions also worked for duplex reactions. Fig. 6 presents an example gel picture where we screened fifty 'Striped Beauty' plants using duplex RT-PCR to detect CaYSV and endogenous CaYMV in greenhouse grown plants. As positive controls, 1 ng of plasmids including pYSOK7672 and pYM0036P1 diluted in nuclease free water was used. Fifteen out of 50 plants produced a 565-bp PCR product, suggesting that these plants contain the endogenous CaYMV sequences (Fig.6). One plant produced a 695-bp PCR product indicating that it was infected with CaYSV. Twenty eight samples produced two PCR products of 695 and 565 bp, which indicate these were co-infected with CaYSV and endogenous CaYMV. Finally, six samples which did not produce any PCR product, which indicates that these plants are free of endogenous CaYMV and CaYSV infection (Fig.6). Based on these data, we confirmed that we successfully detected of CaYSV and endogenous CaYMV simultaneously in an RT-PCR assay.

One drawback of this assay is we cannot confirm which phase of CaYMV is detected. As CaYMV is a badnavirus, which has a DNA genome, a full length RNA replicative intermediate, as well as transcripts, it is not clear if the RT-PCR reaction succeeded in amplifying viral DNA, viral RNA or both. However, for diagnostics in

screening plants to segregate clean germplasm, this method is adequately sensitive to be applied in a greenhouse program.

Multiplex RT-PCR detecting CaYMV, CaYSV and BYMV

This advance was combined with work by Ravendra Chauhan who's primary goal was to develop a multiplex RT-PCR to detect CaYSV, BYMV and CaYMV. Alongside his efforts to test the sensitivity of the assay, I conducted experiments to optimize the annealing temperature for the multiplex RT-PCR using temperature gradient PCR. One ng of mixture of plasmids (pYM0036P1, pBYMV382 and pYSOK7672) diluted in nuclease free water. Table 1 indicates the optimum annealing temperatures for three primers. Positive plasmid mixture was prepared by diluting 1 ng of each plasmid in nuclease free water. Direct gradient PCR was conducted with each reaction containing 7.5% DMSO, 2.5 mM $MgCl_2$ with 35 reaction cycles and a melting temperature range from 50 – 56 °C. Based on our resulting gel (Fig. 7a) the temperature range we selected (50 – 55.8℃) didn't indicate profound difference in PCR outcomes. At the temperature of 56℃, there was no PCR product for all three primers which leads to the conclusion that at this temperature there could be no primer annealing to the template (Fig.7a). Based on our data we selected 52℃ as the optimal annealing temperature for our multiplex RT-PCR. These outcomes are published in (Chauhan et al., 2015b).

The next step was to optimize the $MgCl_2$ concentrations for the reaction as this is a co-factor and a catalyzer in PCR reactions. The GoTaq buffer contains 1.5 mM $MgCl_2$, and additional $MgCl_2$ was added, raising the concentrations to 2.0, 2.5, 3.0, 3.5, 4.0 and 5.0 µM.

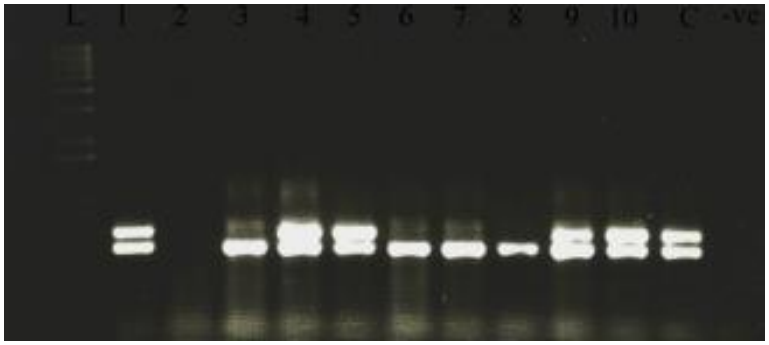

Fig. 6. Duplex RT-PCR detection of CaYSV and CaYMV in greenhouse grown canna cultivar 'Striped Beauty'. "L" is the 1-kb Plus ladder and "C" is the control duplex RT-PCR carried out using mixture of pYSOK7672 and pYM036P1. Samples 1,4,5,9 and 10 tested positive for both endogenous CaYMV and CaYSV. Samples 3, 6, 7 and 8 were tested positive for endogenous CaYMV only. Sample 2 was clean from both endogenous CaYMV and CaYSV infection.

Direct PCRs were conducted using 1 ng of mixture of plasmids diluted in nuclease free water. Our results indicated that a minimum of 2.5 µM $MgCl_2$ is needed to detect all three bands in equal proportion, and that continuing to raise the $MgCl_2$ concentration did not hamper the reaction outcomes (Fig.7b). Based on these outcomes we chose 2.5 µM concentration of $MgCl_2$ as optimal for multiplex RT-PCR. Based on this developed multiplex RT-PCR assay we were able to detect two Potyviruses and one Pararetrovirus that infect canna together. These data are published in a journal article, "A reliable and rapid multiplex RT-PCR assay for detection of two Potyviruses and a Pararetrovirus infecting cannas" (Chauhan et al., 2015c). With having a successful assay to detect CaYMV, our next goal was to elucidate the whole genome of CaYMV.

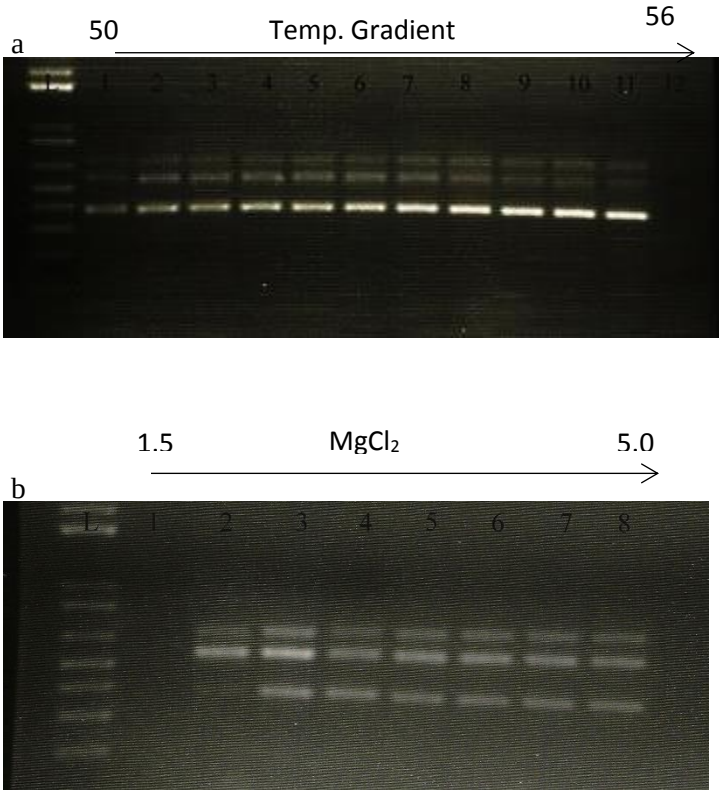

Fig. 7. Optimization of multiplex RT-PCR, (a) Multiplex RT-PCR was carried out using 1 ng of each plasmid template and 0.2/0.9/0.9 μM BYMV/CaYMV/CaYSV primers. "L" is the 1-kb Plus ladder and lane 1, 50.0℃; lane 2, 50.2℃; lane 3, 50.6℃; lane 4, 51.2℃; lane 5, 51.9℃; lane 6, 52.6℃; lane 7, 53.4℃; lane 8, 54.1℃; lane 9, 54.8℃; lane 10, 55.4℃; lane 11, 55.8℃ and lane 12, 56℃. (b) Lanes 1 to 8 contain PCR products generated using 1.5, 2, 2.5, 3, 3.5, 4, 4.5 and 5.0 mM $MgCl_2$ respectively.

Molecular characterization of an endogenous viral sequence

DNA was isolated from virion preparation A and sent for sequencing to obtain the genome sequence of CaYMV. Dr. Peter Hoyt carried out the initial bioinformatics work to build contigs and they sent us the sequence of an element that he named 0036. This sequence is 2542 bp in length and contains the translational start and stop codons. Using Expassy software to obtain a best likely translation product and then BLASTp, we determined that 0036 produces a protein of 717 amino acids which contains badnavirus-like RT/RNAse H domains. This protein has high sequence similarity to an unnamed protein of SCBV with 75% identity, 1102 max score and 95% query cover (YP_003284237.1). Also 66% identity and 96% query cover was shown to *Pineapple bacilliform comosus virus* polyprotein with a 993 max score (ADN96018.1). NCBI BLASTn revealed that the nucleotide sequence of 0036 shares 75% identity with an 1153 max score to SCBV (FJ439817.1) and 99% identity with 996 max score to other reported partial sequences of CaYMV that have been reported in NCBI (see below Table 5). Based on these NCBI BlASTp and BLASTn results we concluded that the 0036 sequence is a badnaviral sequence. Our next step was to obtain the full badnaviral genome sequence by expanding the 0036 genome through PCR and sequencing confirmation.

Table 5. NCBI BLASTp and BLASTn results for EVS-0036 sequence

Search	Best hit	NCBI accession no	Max score	%Query cover	% identity
BLASTp					
	SCBV Unnamed protein	YP_003284237.1	1102	95	75
	PBCoV polyprotein	ADN96018.1	993	96	66
	BSV polyprotein	YP_224289.1	944	94	66
BLASTn					
	CaYMV polyprotein partial CDs	KP836341.1	1016	22	99
	BSV complete genome	KT895259.1	303	52	73
	SCBV complete genome	JN377537.1	106	12	73

Attempt to expand the 0036 sequence to build a complete badnavirus genome

We hypothesized that 0036 was a fragment of the CaYMV genome and for some unexplained reason we were not able to obtain a complete genome sequence from the virion preparation. We considered that the use of DNase I to degrade cellular DNA, prior to extraction of virion DNA, may have degraded portions of the viral genome as an unfortunate consequence. Therefore, to obtain the full viral genome we devised a strategy using PCR to extend the 0036 genome in the 5' direction. Since 0036 showed high similarity to NCBI reported badnaviruses, we developed two reverse primers called 50R and 350R which sit down 116 and 408 bp, respectively from the 5' end of the 0036 sequence (Table 3). We developed one forward generic PCR primer called DW1F (Table 3) based on alignments with BSV (NC008018.1) and SCBV (FJ439817.1), which would lie approximately 1000-2000 bp upstream of the 50R primer a conserved RT sequence of the polyprotein. PCR was carried out according to previously described materials and methods and we obtained bands that we then gel purified and sent for sequencing. The sequences that we received from the OSU Recombinant DNA/Protein Core Facility failed to align with the 0036 sequence, SCBV or BSV (Fig. 8). NCBI blast results of these obtained sequences indicated plant origin instead of badnaviral sequences. Sequences obtained from DWF1 & 50R primer products indicated 99 query cover to *Musa accuminata* sub species *malaccencs* inactive leucine rich repeat protein kinase (XM 009393546.1). Since out first attempt to PCR walkthrough of 0036 sequence failed we developed another primer set based on SCBV and 0036 alignments (CaYMV-DR & CaYMV-DF) (Table 3). Direct PCR was conducted to obtain expected PCR product of the above primers. Similar to our first attempt we successfully cloned and sequenced the

PCR products and again the resulting sequences were of plant origin and had no relationship to 0036 or any badnavirus sequence. These multiple attempts to extend the CaYMV genome using PCR were unsuccessful. One explanation is that the short 0036 sequence is a partial badnaviral genome sequence that is either episomal or is integrated into the canna plant genome. Based on these outcomes, we named the sequence endogenous viral sequence 0036 (EVS-0036). Since EVS-0036 was a partial genome we developed EVS-0036 specific primers (301F/1621R, 1081F/1930R, 0036F/R and 0036-86F/R) to amplify the entire 2542 bp sequence and 301F/1621R, 1081F/1930R PCR products were successfully cloned in to the pGEM-T Easy vector (Tables 3,4). Further evidence that EVS-0036 may be incorporated into the nuclear plant DNA was obtained by Southern analysis (see next section). Because this method of virion purification (virion DNA extraction A) failed to produce the complete CaYMV genome sequence, we employed an alternative protocol for badnavirus purification and NGS sequencing in which we successfully obtained the CaYMV genome, and again recovered the EVS-0036 contig as well as another short sequence which we name EVS-00206.

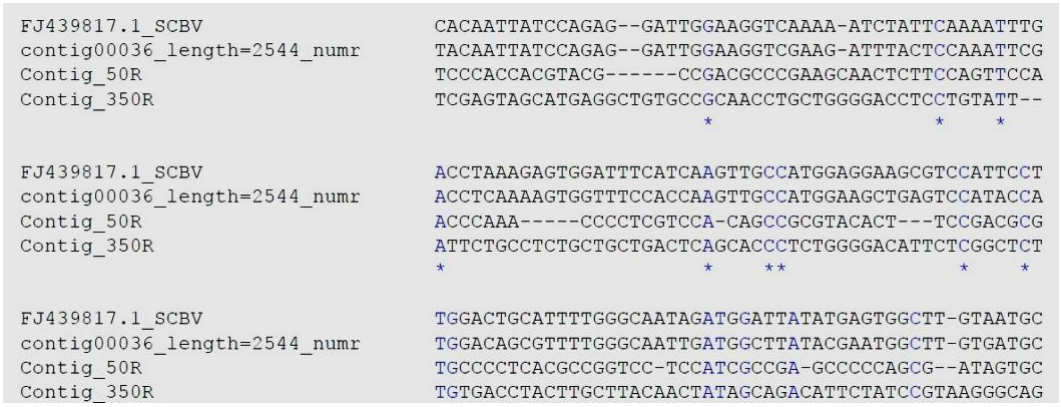

Fig. 8. Multiple sequence alignment of newly obtained sequences with 0036 and SCBV full genome. Contigs 50 and 350 R resembles the obtained sequences from developed primers (DWF1 and 50/350R) for 0036 sequence expansion. Stars represent matching nucleotides. These sequences dose not match with either 0036 or SCBV.

Sequencing and molecular characterization of intact CaYMV genome

CaYMV virions were observed in canna leaf sap using transmission electron microscopy. We identified two particles measuring 60-63 nm or 124-133 nm in length (Fig. 9a). The existence of two modal length particles suggests that CaYMV genome concatemers may be encapsidated, as has been reported for the badnavirus SCBV (Geijskes et al., 2004).

DNA sequencing of crude virion preparations from infected canna plants produced a total of 153,488 cleaned individual sequence reads with a mean read length of 385 bp. Contig assembly (see Materials and Methods) produced the consensus full length, circular 6966 bp CaYMV genome at 19.6-fold coverage, which was then cloned and sequenced to verify (Fig. 9b,c).

Molecular characterization of this recovered CaYMV genome was carried out. Numbering of CaYMV genome was carried out similar to other pararetroviruses, starting with putative tRNA ^{met} binding site (Xu et al., 2011). The sequence of the first 15 nucleotides of CaYMV tRNA^{met} consists of TGGTATCAGAGCGAG, which is highly conserved with other reported badnaviruses (Laney et al., 2012). A potential TATA box with the sequence of TATATAA is located upstream of tRNA^{met} binding site at nucleotide positions, between 6752-6758. There is a potential polyadenylation signal (AATAAA) located between nucleotide positions 6868-6873 of the CaYMV genome (Fig. 9c).

Annotation of CaYMV complete genome indicates the presence of two smaller ORFs and one large OFR3 encoding a polyprotein. Even though many badnaviral species are reported to bear four to six ORFs some badnaviral species including *Gooseberry vein*

banding associated virus, SBCV and *Banana streak IM virus* are reported to have only 3 ORFs with a large polyprotein (Bouhida et al., 1993; Franck et al., 1980; Geering et al., 2011; Xu et al., 2011).

The CaYMV ORF1 is positioned in between 564-1085 nucleotides and encodes a putative protein of 173 amino acids. BLASTp analysis of ORF 1 indicates high sequence similarity to *Cycad leaf necrotic virus* ORF 1 protein (B5AK48) with a score of 167 and SCBV ORF 1 protein (D0QEI7) with score of 161. Function of ORF1 protein product is unknown. The ORF2 commences at nucleotide position 1085 and concludes at nucleotide position 1486 showing high sequence similarity to SCBV (D0QEI7) and *Banana Streak Mysore virus* (A0A075DDT9) ORF2 proteins with scores of 125 and 112 respectively. The ORF2 protein product is unknown. We observed that translation stop and start codons for OFR1 and ORF2, respectively, reside in common TAATGA element. A similar phenomenon was observed in a BSV isolate from *Musa acuminata* (BSAcYNV) from China (Zhuang et al., 2011)(Fig. 9c).

The large ORF3 of CaYMV encodes a putative polyprotein of 1684 amino acids. BLASTp analysis of this polyprotein with other badnaviral polyproteins indicates high sequence similarity to *Fig badnaviral polyprotein* (A0A0F7RQB1) with a score of 1154 and SCBV polyprotein (H8Y6I4) with a score of 1147. The predicted amino acid sequence of ORF3 polyprotein contains several characteristic features of badnaviruses including aspartic protease, zinc finger motif and reverse transcriptase and RNase H. The ORF3 coding region extends from nucleotide position 1491 to 6545 of the genome. The zinc finger domain is encoded between nucleotide positions 2520-3862. It is reported that this zinc finger can serve as a RNA binding motif throughout retroviral replication, and it

is necessary for the viability of CaMV; however, the exact function of zinc finger is yet to be discovered (Hohn and Rothnie, 2013). The aspartic protease domain of CaYMV is encoded between nucleotide positions 3862-5090 and acts as a structural enzyme required for the maturation of the polyprotein. Most importantly RT and RNase H motifs of CaYMV, which encoded between nucleotide positions 5097 and 6545, are considered as the work horse of pararetroviral replication and known to be highly conserved among pararetroviruses (Hohn and Rothnie, 2013).

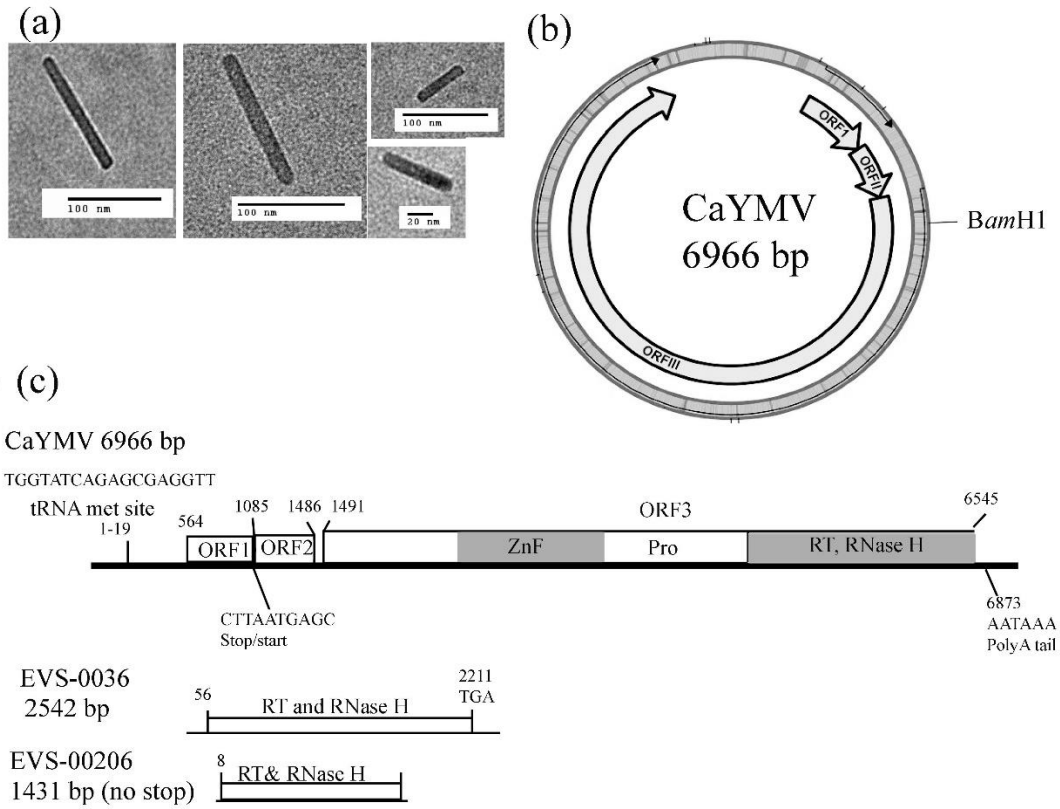
(a)
100 nm
100 nm
100 nm
20 nm
(b)
CaYMV
6966 bp
ORF1
ORF2
ORFIII
BamH1
(c)
CaYMV 6966 bp
TGGTATCAGAGCGAGGTT
tRNA met site
1-19
564
1085
1486
1491
ORF3
6545
ORF1
ORF2
ZnF
Pro
RT, RNase H
CTTAATGAGC
Stop/start
6873
AATAAA
PolyA tail
EVS-0036
2542 bp
56
RT and RNase H
2211
TGA
EVS-00206
1431 bp (no stop)
8
RT& RNase H

Fig. 9. Properties of CaYMV and its associated subviral elements.

(a). CaYMV virion particles under transmission electron microscopy. Bacilliform shaped CaYMV virion particles with two different sizes (60-63 nm and 124-133 nm). Scale bars represent 100 nm at the bottom of each image.

(b, c) Circular and linear illustrations of the CaYMV genome. The genome length is 6966 bp. The open boxes represent ORFs1, 2 and 3. Nucleotide position 1 is defined based on $tRNA^{met}$ binding site. The nucleotide positions for start and stop of each ORF is identified above the boxes in the linear diagram. The ORF3 polyprotein with relevant positions for the zinc finger binding domain (ZnF), the aspartic acid protease (Pro) domain, the reverse transcriptase (RT) and RNAse H domains are given. Two additional diagrams represent EVS-0036 and EVS-00206 with nucleotide positions for the translation start and stop codons are shown (no stop codon for EVS-00206).

Identification of CaYMV associated subviral elements and defective genomes through NGS

From our second virion DNA purification, we found two CaYMV-associated subviral elements and two defective CaYMV genomes. We were able to identify EVS-0036 again in this virion preparation alone side with another mutant, subviral element EVS-00206, which is 1431 bp in length (Fig. 9c). The EVS-00206 has shorter 5' and 3' ends than EVS-0036 and lacks the translational stop codon.

The two defective CaYMV genomes were named as Gap5RC and Gap6RC. Efforts to PCR and cloning confirmation of these two genomes were unsuccessful (Fig. 10a). Molecular annotations of these genomes were carried out. Both sequences were shorter than the full CaYMV genome and named as Gap5RC, which was 6978 bp in length, and Gap6RC consisting of only 5458 bp. Both of these defective genomes consist of fragmented polyprotein creating five to six ORFs. We were able to detect several nucleotide insertions that disrupt the integrity of the ORF3 polyprotein in these defective genomes. GAP5RC, ORF3A, 3B, 3C1 and 3C2 regions have legitimate start and stop codons. A single G inserted in nucleotide position 3824 has created a frameshift and new start and stop codons for OFR3B and 3C. In between nucleotide positions 2600 and 2639, four and six nucleotide insertions were detected in Gap5RC OFR 3A and 3B junction (Fig. 10b). These 4 nt and (AAGG) 6 nt insertion (AATTAG) appear to be duplications of the neighboring sequences, which creates a frameshift and stop codon in the Gap5RC genome. Similarly in Gap6RC 4 nt insertion can be seen in between nucleotide positions 2600 and 2639, which is also a result of a duplication event. The sequences around the 3C1and 3C2 junctions of GAP5RC are 100% conserved with CaYMV, but this

frameshift seems to have created two additional start and stop codons. We also observed a major nucleotide deletion in Gap6RC compared to CaYMV genome, where 1511 nucleotides were deleted in between nt positions 4247 and 5758 (Fig. 10c). One possible scenario for the occurrence of these defective genomes could be illegitimate RNA recombination during transcription (Geering et al., 2001; Pahalawatta et al., 2008). We could not successfully confirm the presence of these two genomes through PCR and cloning so they might not occur as adjoining sequence elements by infecting canna.

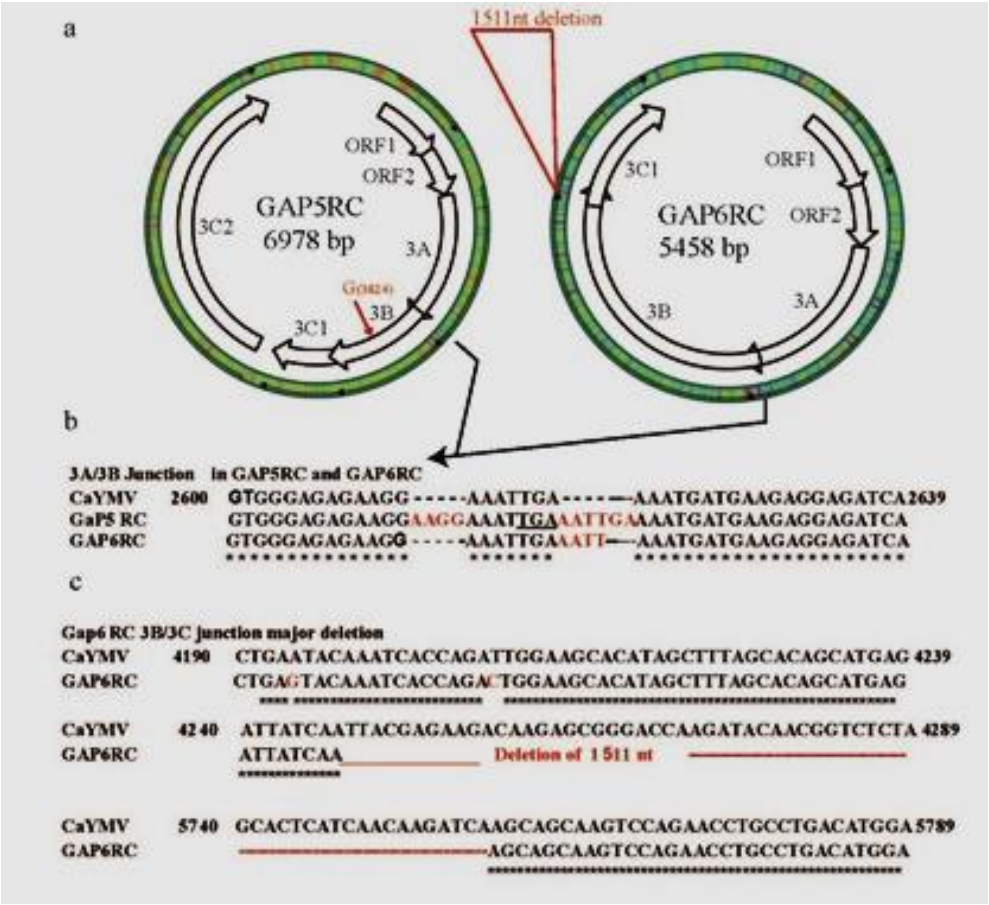
a
1511nt deletion
ORF1
ORF2
GAP5RC
6978 bp
3C2
3A
3B
3C1
3C1
ORF1
GAP6RC
5458 bp
ORF2
3B
3A
b
3A/3B Junction in GAP5RC and GAP6RC
CaYMV 2600 GTGGGAGAGAAGG-----AAATTGA-----AAATGATGAAGAGGAGATCA 2639
GaP5 RC GTGGGAGAGAAGGAAGGAAATTGAAATTGAAAATGATGAAGAGGAGATCA
GAP6RC GTGGGAGAGAAGG-----AAATTGAAATT---AAATGATGAAGAGGAGATCA
c
Gap6 RC 3B/3C junction major deletion
CaYMV 4190 CTGAATACAAATCACCAGATTGGAAGCACATAGCTTTAGCACAGCATGAG 4239
GAP6RC CTGAGTACAAATCACCAGACTGGAAGCACATAGCTTTAGCACAGCATGAG
CaYMV 4240 ATTATCAATTACGAGAAGACAAGAGCGGGACCAAGATACAACGGTCTCTA 4289
GAP6RC ATTATCAA Deletion of 1511 nt
CaYMV 5740 GCACTCATCAACAAGATCAAGCAGCAAGTCCAGAACCTGCCTGACATGGA 5789
GAP6RC AGCAGCAAGTCCAGAACCTGCCTGACATGGA

Fig. 10. (a). Diagrammatic representation of two CaYMV defective genomes, Single G inserted in nucleotide position 3824 in GAP5RC that causes a frameshift in the sequence, followed by a newly created 3B stop codon and 3C1 start codon.

(b). Alignment of CaYMV, GAP5RC, GAP6RC sequences located between CaYMV nucleotide positions 2600 and 2639. The sequence changes in GAP5RC and GAP6RC are highlighted in red.

(c). Alignment of CaYMV and GAP6RC sequences located between the CaYMV nucleotide positions 4190 and 5789. GAP6RC has 1511 nucleotides deleted in the region between nucleotide positions 4247 and 5758. Two base changes are also highlighted in red.

Phylogenetic analysis of CaYMV, EVS-0036 and EVS-00206

A maximum likelihood (ML) tree was constructed using full length genome sequences (Fig. 11a) and revealed that CaYMV can be classified among 30 other badnavirus species (Table 2). This tree is rooted with *Rice tungro bacilliform virus* (RTBV) as an outgroup and contains three major clades. Clade I contains 15 species including CaYMV, *Banana streak GF virus* (BSGFV), *Banana streak IM virus* (BSIMV), *Banana streak OL virus* (BSOLV), *Banana streak VN virus* (BSVNV), and *Banana streak MY virus* (BSMYV). Nested within Clade I are three sister subclades. In particular, Clade 1B contains two branches and four related viruses: BSGFV, BSIMV, CaYMV, and FBV-1 (Fig. 11a). Three of these viruses, aside from CaYMV, are described to be globally distributed (Iskra-Caruana et al., 2014; Laney et al., 2012).

Among Badnavirus species, ORF1 and ORF2 are the most divergent whereas the ORF3 polyprotein contains the essential viral replicase components, sharing similar linear arrangement of the aspartic acid proteinase, RT and RNAse H domains. Sequence comparisons using badnavirus ORF3s tend to reveal strong phylogenetic relationships. In Fig. 11b, Clade I contains CaYMV clustered with FBV-1, BSOLV, BSIMV and BSVNV. This ML tree confirms the relatedness of CaYMV ORF3 polyprotein to BSV species as well as FBV-1. However, Clade 1 broadly encompasses other monocot and dicot infecting badnavirus species. While these combined trees do not speak to badnavirus ontogeny (Fig. 11a, b), it is reasonable to hypothesize that the ORF3 polyprotein is a conserved module that shows less divergence among species because it contains most of the essential features for virus replication. Phylogenetic analysis was carried out comparing nucleotide sequences encoding the RT/RNAse H domains of EVS-0036 and

EVS-00206 with 22 badnavirus species including CaYMV (Fig. 12). As expected, EVS-0036 is closely related to the RT/RNAse H domain of CaYMV. The EVS-00206 is closely related to EVS-0036, suggesting that it may have arisen through faulty replication of that element, although we cannot rule out the possibility that it is also derived from CaYMV.

Canna yellow mottle virus was reported in Florida in 2004 ((Momol et al., 2004). On May 10, 2015 a search of the nucleotide database of Genbank produced 65 sequences all of which are 565 bp, that were amplified by researchers at various institutions. A ML tree was constructed using these sequences (Fig. 13). This tree was rooted with CaYMV genome reported here. Interestingly EVS-0036 is embedded within a group of sequences reported from cannas in India and Thailand. EVS-00206 seems more closely related to CaYMV. Given the relationship of EVS-0036 to these other reported sequences, and the lack of evidence of complete genomes for CaYMV, it is highly likely that the reported sequences in Genbank are endogenous virus sequences that have been dispersed through vegetatively propagated germplasm and not the CaYMV genome itself. Fig. 13 also reports the geographic locations of these reported sequences, many of which are in Asia, although some are in Africa and Europe. If these 565 bp sequences reported in Genbank are in fact endogenous viral elements, then their geographic dispersal would likely be due to vegetative propagation of canna rhizomes and dispersal through trade. Considering the likely scenario that these sequences related to EVS-0036 are endogenous sequences, then horizontal gene transfer would be a force in their divergent evolution.

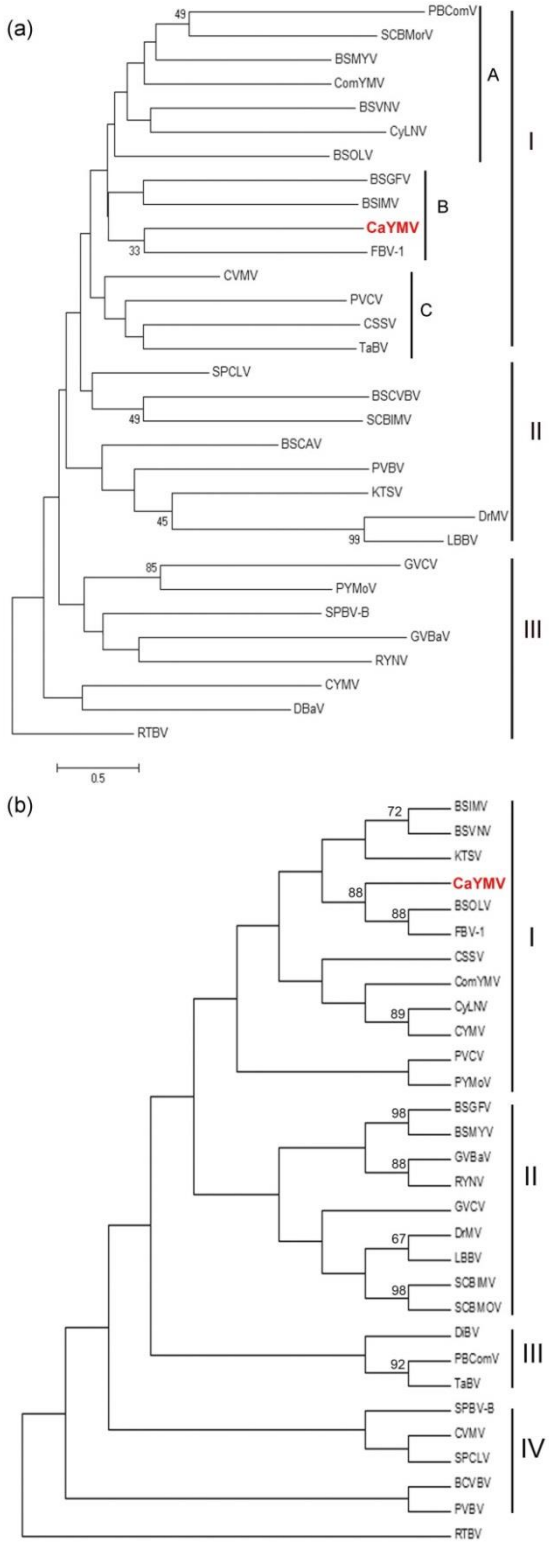
(a)
PBComV
SCBMorV
BSMYV
ComYMV
BSVNV
CyLNV
BSOLV
BSGFV
BSIMV
CaYMV
FBV-1
CVMV
PVCV
CSSV
TaBV
SPCLV
BSCVBV
SCBIMV
BSCAV
PVBV
KTSV
DrMV
LBBV
GVCV
PYMoV
SPBV-B
GVBaV
RYNV
CYMV
DBaV
RTBV
A
B
C
I
II
III
49
33
49
45
99
85
0.5
(b)
BSIMV
BSVNV
KTSV
CaYMV
BSOLV
FBV-1
CSSV
ComYMV
CyLNV
CYMV
PVCV
PYMoV
BSGFV
BSMYV
GVBaV
RYNV
GVCV
DrMV
LBBV
SCBIMV
SCBMOV
DBV
PBComV
TaBV
SPBV-B
CVMV
SPCLV
BCVBV
PVBV
RTBV
I
II
III
IV
72
88
88
89
98
88
67
98
92

Fig. 11. ML trees indicating relationship of CaYMV to 30 badnavirus species.

(a) ML tree comparing complete nucleotide sequences of 30 badnavirus species, including CaYMV, listed in Table 2. The ML tree was generated using MEGA6 software, selecting the Kimura model with 1000 bootstraps.

(b) ML tree comparing the ORF3 polyproteins of the same badnaviruses listed in Table 2. The ML tree was developed using MEGA6 software, selecting the Poisson model with 1000 bootstraps.

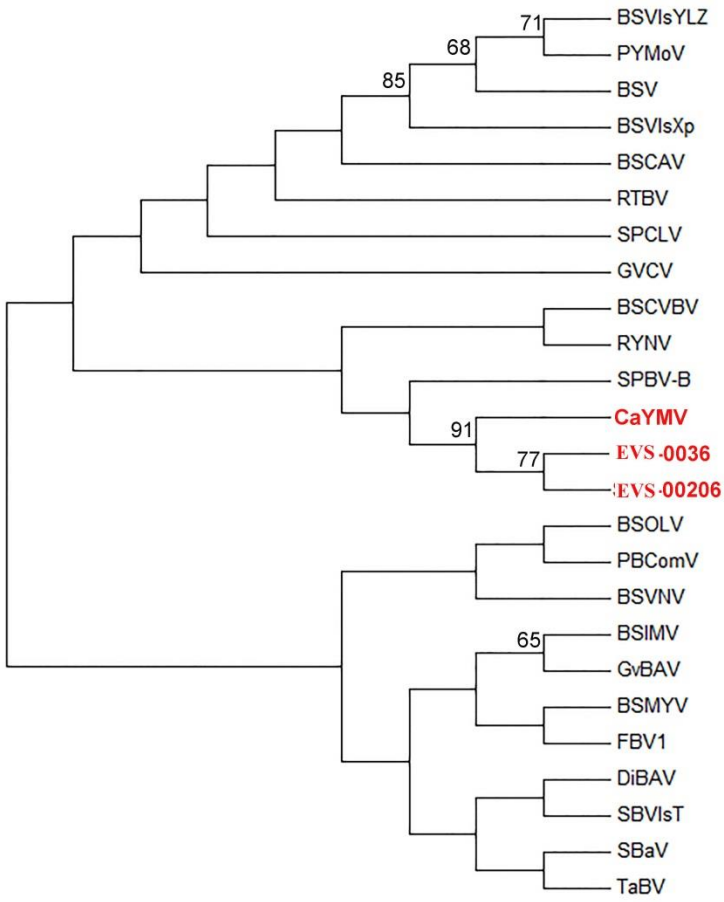

Fig. 12. ML tree comparing the nucleotide sequence of EVS-0036 and EVS -00206 with comparable RT/RNAse H encoding domains of badnaviruses listed in Table 2. MEGA6 software, selecting the Kimura model with 1000 bootstraps was used. The tree demonstrates that these EVS sequences are most closely related to CaYMV than other badnavirus sequences.

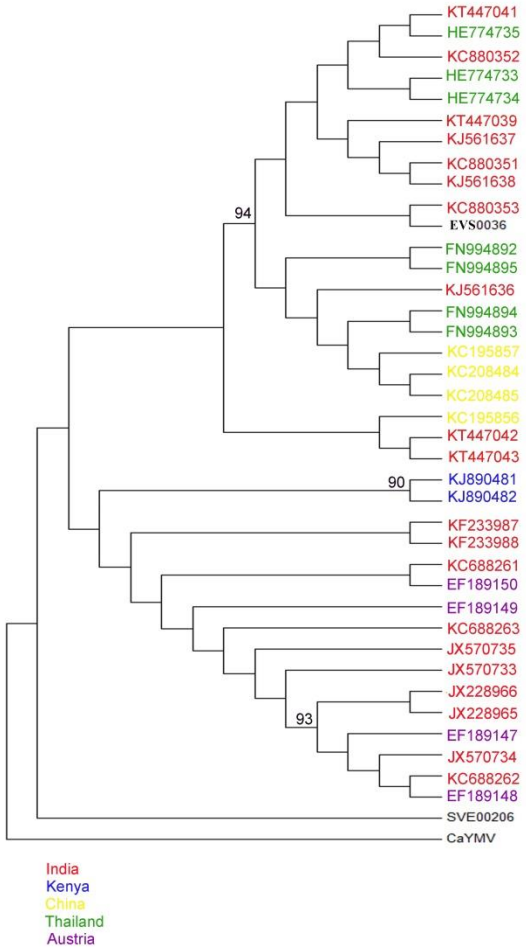

Fig. 13. ML tree resolving relationships among NCBI reported 565 bp containing partial CaYMV sequences with EVS-0036, EVS-00206 and tree was rooted to RT/RNAse H encoding domain of CaYMV. MEGA6 software, selecting the Kimura model with 1000 bootstraps was used. Geographical distributions of these sequences are indicated as Red- India, Blue- Kenya, Yellow- China, Green- Thailand and purple- Australia.

Southern and northern analysis

Southern blot hybridization was conducted using naturally infected canna plants (cultivars 'Striped Beauty' and 'Pretoria') and biotinylated probes that overlap CaYMV ORF1 or EVS-0036 (Fig. 14a). The 0036 probe unavoidably shares 70% sequence homology with the CaYMV genome and 86% homology with EVS-00206. Using ORF1 or 0036 probes, two bands were detected in undigested samples (Fig. 14b lanes 2& 5) which represent two conformations of the CaYMV genome. Since we report two particles, one that is twice the modal length of the smaller particle observed by electron microscopy in Fig. 9a, it is likely that the CaYMV genome forms concatamers. Others have reported that badnavirus genomes form concatamers during replication, encapsidation, or nuclear genome integration (Geijskes et al., 2004; Iskra-Caruana et al., 2014). The B*am*HI enzyme recognizes a single restriction site in the CaYMV genome and succeeded to only partially digest the episomal DNA producing a third lower migrating band (Fig. 14b lanes 1& 3). Badnaviruses have one or two discontinuities in one of the genome strands, which often affects the digestion pattern and migration through the gel. Given the lack of complexity in the banding pattern, CaYMV does not seem to resemble a class of endogenous pararetroviral sequences that has undergone extensive recombination and multiple insertions into the plant genome.

In the Southern treated with the 0036 probe, there is a higher molecular weight (HMW) band of undigested genomic DNA, in addition to the episomal CaYMV genomes. The HMW DNA band does not occur in Southern blots treated with the ORF1 probe and suggests that the 0036 probe detects subviral DNA elements that are incorporated into the plant genome (Fig. 14b). In repeated experiments we failed to detect

episomal 0036. Unfortunately we also found canna genomic DNA to be resistant to digestion with several restriction enzymes including H*ind*III, E*co*RI and B*am*HI.

Northern analysis was carried out using total RNA isolated from a diseased plant that tested positive by PCR using CaYMV diagnostic primers (Fig. 14c). Using biotinylated ORF1 or 0036 probes, and each blot presented a single band. CaYMV or EVS-0036 transcripts were not detected in healthy canna plants.

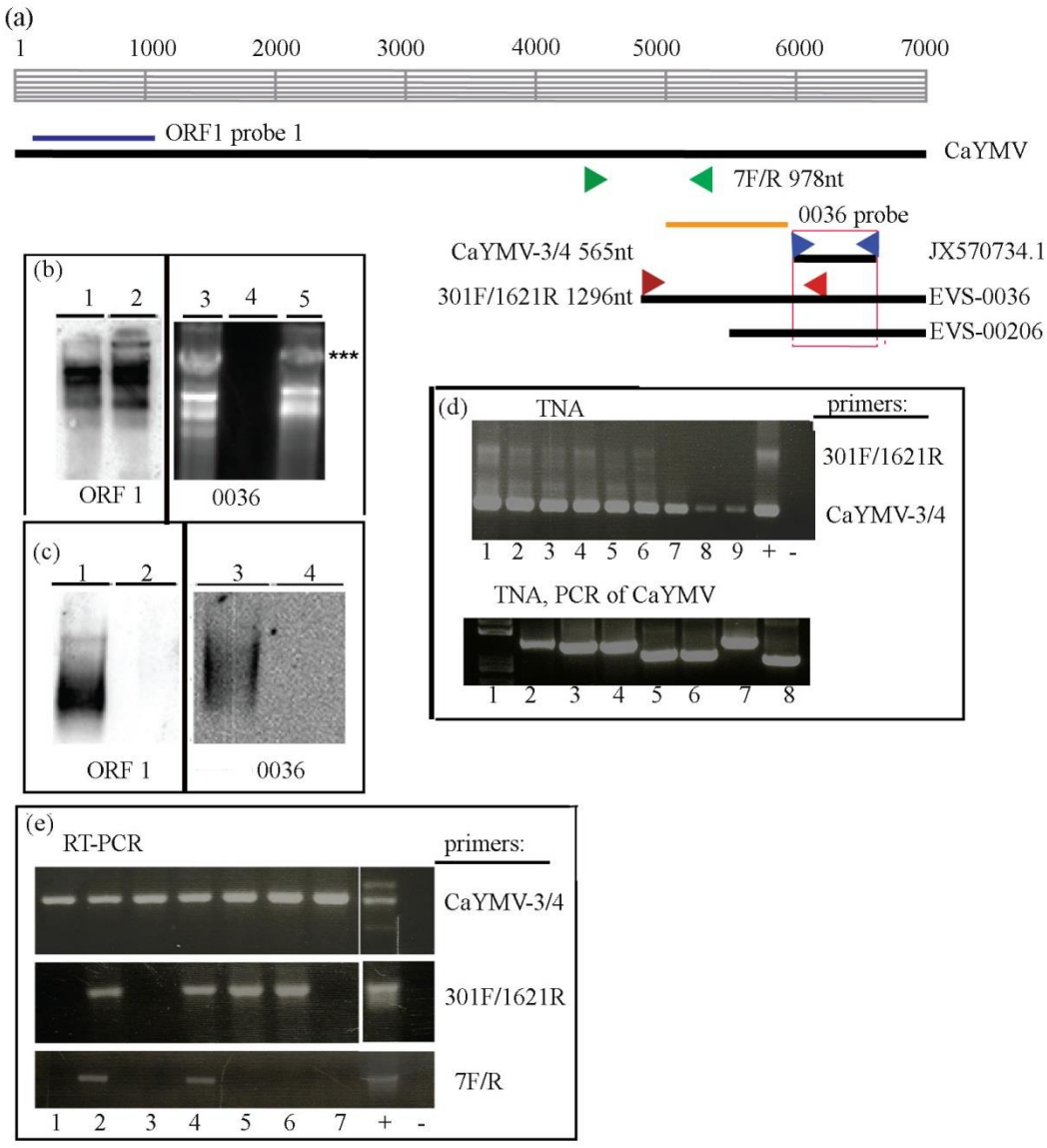
(a)
1
1000
2000
3000
4000
5000
6000
7000
ORF1 probe 1
CaYMV
7F/R 978nt
0036 probe
CaYMV-3/4 565nt
JX570734.1
301F/1621R 1296nt
EVS-0036
EVS-00206
(b)
1 2
3 4 5
***
ORF 1
0036
(c)
1 2
3 4
ORF 1
0036
(d)
TNA
primers:
301F/1621R
CaYMV-3/4
1 2 3 4 5 6 7 8 9 + -
TNA, PCR of CaYMV
1 2 3 4 5 6 7 8
(e)
RT-PCR
primers:
CaYMV-3/4
301F/1621R
7F/R
1 2 3 4 5 6 7 + -

Fig. 14. Molecular characterization of CaYMV using southern and northern analysis combined with PCR/RT-PCR

(a). A scaling ruler is presented at the top of the diagram identifying each 1000 bp position along the CaYMV genome. The black lines represent the CaYMV genome or EVS sequences and their identity is provided on the right on the figure. Probes that were used in southern and northern hybridization are labeled as ORF1 (highlighted in blue) and 0036 (highlighted in red). PCR primers are identified by colored arrows. All the primers and primer sequences are given in the Table 1.The red box highlights the region where EVS-0036 and EVS-00206 are aligned and primers detect both DNAs.

(b). Two representative Southern blots were probed with ORF1 probe or 0036 probe (listed at bottom of each gel). Genomic DNA was extracted from naturally infected canna leaves, cultivar 'Striped Beauty' (lanes 1 and 2) or 'Pretoria' (lanes 3 and 5). Lane 4 is blank. Lane numbers are listed above each blot. Genomic DNA was digested with B*am*HI in lanes 1 and 3. *** identify undigested genomic DNA.

(c). Two representative northern blots hybridized with ORF1 or 0036 probes. RNA was extracted from naturally infected canna leaves and the canna cultivars are identified: Lanes1 and 3, 'Striped Beauty' which tested positive for CaYSV, CaYMV and EVS-0036 by RT-PCR. Lanes 2 and 4 'Australia' plants that tested negative for CaYMV and EVS-0036 by RT-PCR.

(d). Two representative 1% ethidium bromide stained agarose gels. Top gel contains a 1 kb ladder (L) and PCR products developed using TNA.

(e). Representative 2% agarose gel electrophoresis presenting RT-PCR products of naturally infected canna plants grown from rhizomes in the greenhouse. TNAs were extracted approximately 8 weeks after planting. Primers used in each test are identified on the right of each gel. Lane numbers identify leaf samples taken from different plants, but tested in each gel.

PCR diagnostic detection of CaYMV

Having characterized the complete CaYMV genome, we then re-examined previously reported CaYMV diagnostic primers, CaYMV-3 and -4 (Chauhan, RP et al., 2015) (Table 1) and discovered that these primers overlap with EVS-0036 and EVS-00206 and fail to align with the CaYMV genome. The CaYMV-3/4 primers were developed by Momol, et al., (2004) for diagnostic detection of CaYMV and have been widely used for detecting new outbreaks of disease in different regions (Momol, et al., 2004; Byrne, 2006; Marino et al., 2008; Pappu et al., 2008; Chauhan, R et al., 2015). Phylogenetic analysis was conducted using NCBI reported CaYMV sequences all of which are fragments that are less than 1500 bp. A ML tree was generated, which revealed that these sequences are more closely related to EVS-0036 than CaYMV (Fig. 13). To verify that EVS-0036 can be differentiated from CaYMV we developed PCR primers that specifically detect EVS-0036 and do not overlap with EVS-00206 or the CaYMV genome (Fig 14). Then we carried out duplex PCR using total nucleic acids isolated from canna plants and the diagnostic CaYMV-3/4 and 301F/1621R primers. Fig. 14d shows a representative data set (top gel). Nine canna plants were sampled, and six produced PCR products using the 301F/1621R and CaYMV-3/4 primers, but only three produced PCR products representing the CaYMV-3/4 primers. Samples that tested positive using both primer sets contained EVS-00306. Samples that did not test positive using the 301F/1621R primer pair contained EVS-00206. These PCR products were sequenced to verify their origins as EVS-0036 or EVS-00206. Direct PCR was also carried out to clone the entire CaYMV genome and each of the EVS elements PCR product was sequenced (Fig. 14d, bottom gel). Since badnaviruses replicate through an RNA

intermediate, RT-PCR was conducted to detect the badnavirus replication intermediate as well as any transcript that may be templates for translation (Periasamy et al., 2006). We selected a set of canna samples that did not test positive for the potyviruses BYMV or CaYSV. RT-PCR was conducted using CaYMV-7F/R, CaYMV-3/4, and 301F/1621R primers in separate reactions. Representative 1% ethidium bromide stained gels demonstrated that PCR products were amplified in all samples using the CaYMV-3/4 primer but only four of those represented EVS-0036. Only two of the seven plants tested positive for CaYMV and EVS-0036 using the CaYMV-3/4, 301F/1621R and 7F/R primers (Fig. 14e). Two samples tested positive for EVS-0036 using the 301F/1621R and CaYMV-3/4 primers. Three samples tested positive using the CaYMV-3/4 primers and none of the other primers suggesting that it detects EVS-00206 (Fig. 14e). These data indicate that CaYMV occurs in samples that also contain endogenous elements, and that both CaYMV and the endogenous elements produce transcripts.

CaYSV but not CaYMV or EVS-0036 can be transmitted to *P. vulgaris*

While most badnaviruses have a restricted host range, only BSV species and SCBV have been shown to be transmissible to plants belonging to other families (Bouhida et al., 1993). Therefore, experiments were conducted to learn if CaYMV is mechanically transmissible to other plants beyond *Cannaceae*, and to learn if EVS-0036 is a component of a disease complex involving CaYMV or CaYSV. In the first experiment, leaf sap was taken from 8-week old canna plants that tested positive for CaYMV+CaYSV+EVS-0036 and inoculated to *P. vulgaris, L.*, *P. coccineus. L.*, *N. benthamiana*, and *C. indica L.* ('Verde') (Table 6). Nucleic acids were extracted from upper leaves at 21 days post inoculation (dpi) and RT-PCR was conducted. CaYMV

transcripts were never detected in an alternative host, indicating that CaYMV failed to replicate through an RNA intermediate in these test plants. Similarly when sap was taken from plants that tested positive for CaYMV and EVS-0036 was used to inoculate plants, none of these became infected. When we used CaYSV containing sap inoculum, 26% (5/19) of *P. vulgaris L* plants and 14% (3/22) *P. coccineus L*. plants became systemically infected. When we used sap from plants that tested positive for CaYSV and EVS-0036 transcripts, only CaYSV was transmitted to 26% of *P. vulgaris* plants. Plants showed mosaic symptoms (Fig.15).

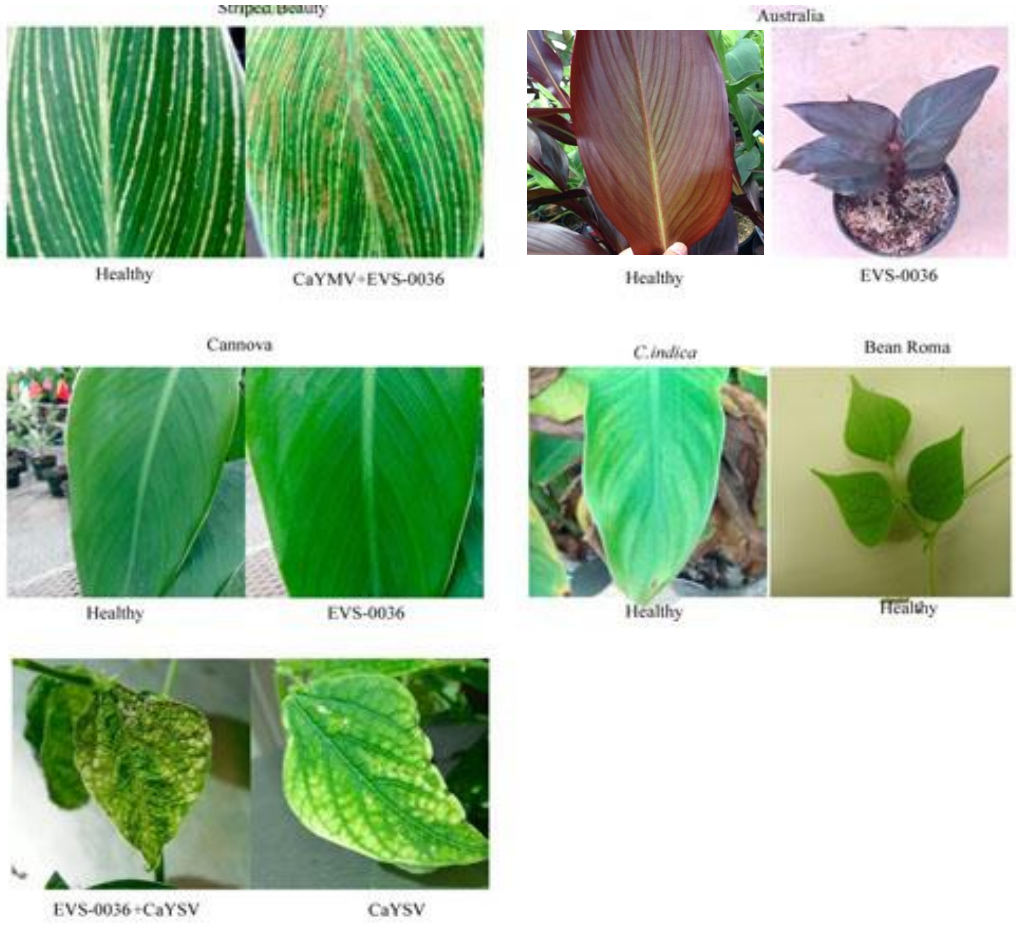

Fig. 15. CaYMV, SVE-0036 and CaYSV infected plants

The symptom development in different canna cultivars and mechanically inoculated bean plants with EVS-0036 with CaYSV.

Table 6. Canna sap used to inoculate alternative plants to test mechanical transmission of CaYSV, CaYMV, and EVS-0036

Inoculum Sap	***P. vulgaris L.***	***C. Indica* (verde)**	***N. benthamiana***	***P. coccineus L.***
CaYSV, CaYMV, EVS-0036	0/52	0/5	0/20	0/4
CaYMV, EVS-0036	0/22*	NT	NT	0/10
CaYSV alone	5/19	NT	NT	3/22
CaYSV sap& EVS-0036 sap[b]	5/19 CaYSV 0/19 EVS	NT	NT	0/22 CaYSV 0/22 EVS
EVS-0036 sap[b]	0/10	NT	NT	0/4

[a]These plants compiled from three to five experiments. Plants that tested positive by RT-PCR showed yellow symptoms in upper leaves. However these yellowed and necrotic upper leaves tested negative by RT-PCR which led us to further investigations.

[b] Sap from a plant that tested positive by RT-PCR for CaYSV and sap from another plant for EVS-0036 were mixed and used to inoculate plants. EVS-0036 sap was from 'Cannova' series plant

CHAPTER V

CONCLUSION AND DISCUSSION

During recent years there has been an increase in reported badnavirus species among vegetative propagated staple crops. Alongside these new species are the simultaneous reports of integrated pararetrovirus sequences into the genomes of these same crops. In the case of banana, such sequences can be activated by high temperatures or other stresses and causes disease (Dahal et al., 1998). In other examples, such as the tobacco endogenous pararetroviruses (TEPRVs), which lack an episomal form of the virus, they exist to enhance host immunity to virus infection (Mette et al., 2002). The pervasiveness of such endogenous sequences in flowering plants has implications for virus-indexing, breeding, and international movement of germplasm. This is especially true for ornamental hybrid canna lily whose rhizomes are traded globally. Furthermore, there is a canna agricultural group known generically as achira (*C. edulis*) that is identified among the Andean Roots and Tuber Crops (ATRCs) as having significant market potential and its development as a biofuels or edible starch crop is encouraged for subsistence farms in regions outside of the Andes (Cisneros et al., 2009; Dewi, 2009; Flores et al., 2003; Hermann and Heller, 1997; Leonel et al., 2002). Although this study failed to detect CaYMV or EVS-0036 in *C. indica* plants that were grown from seed, we

have not yet examined whether CaYMV or endogenous badnavirus sequences occur in *C. edulis*. Such studies to determine if CaYMV or endogenous badnavirus sequences occur in *C. edulis* will be necessary alongside any breeding effort and prior to movement of germplasm for the goal of diversifying cropping systems.

Badnaviruses are double stranded DNA genomes that include a reverse transcription-based replication step (Bousalem et al., 2008). The genome has a highly conserved $tRNA^{met}$ binding site and a conserved polyprotein gene containing reverse transcriptase and RNAse H domains, which resembles animal retroviruses. One drawback to studying badnaviruses is that they are rarely mechanically transmissible. *Sugarcane bacilliform vi*rus (SCBV) is the only badnavirus reported to infect plants of two different families. SCBV has been found to infect banana that belongs to *Musaceae* and rice, sorghum, and brachiaria that belong to *Poaceae* (Borah et al., 2013). In this study we failed to identify an alternative host for CaYMV. In naturally infected canna, CaYMV never occurred alone but was always present alongside CaYSV and/or EVS-0036. These data suggested that CaYMV by itself is poorly infectious and may require the presence of a helper virus or subviral agent in cannas. During the last two years of our studies, the majority of attempts to mechanically inoculate CaYMV to cannas or a wider range of alternative hosts than presented here, failed to produce infection (data not shown). A subset example is provided in Table 6 using leaf sap from canna plants infected with CaYSV, CaYMV and EVS-0036 to inoculate test plants. Only CaYSV was mechanically transmissible. Thus, if there is a synergistic interaction among these species, it likely occurs in canna, and does not involve favoring expanded transmission to alternative hosts.

This study reports the first genome sequence of CaYMV, which is closely related to BSV species and FBV-1. The relatedness of CaYMV to BSV species is particularly interesting since canna (sole member of the family *Cannaceae*) and banana (family *Musaceae*) are in families of the order *Zingiberales*. Most badnaviruses are reported to have evolved closely with their host species and rarely cause natural disease outside their primary hosts. The relatedness of CaYMV with BSV species is most intriguing and points to the possibility that CaYMV arose through evolution after the divergence of their hosts *Musaceae* (banana) and *Cannaceae* (canna), both of which are in the order *Zingiberales* and are classified as commelinids (Bousalem et al., 2008; Iskra-Caruana et al., 2014). With the new discovery of CaYMV sequence these data provide a unique opportunity to study how divergence of Musa and Canna may have played a role in co-evolution of badnavirus species. On the other hand, the relatedness of CaYMV and FBV-1 is puzzling since their hosts are not phylogenetically related. Since it is not yet possible to relate divergence of these viruses to divergence of their host species, perhaps as more Badnavirus species are discovered or the vectors for these viruses are identified, the evolutionary pressures that contributed to their divergence will be revealed.

We developed successful diagnostic PCR primers that detect CaYMV and EVS-0036. It was interesting to observe that certain cultivars have EVS-0036 in the absence of CaYMV or CaYSV and do not show disease. The occurrence of CaYMV together with EVS-0036 is observed. One explanation is that EVS-0036 acts within the plant to create an environment that helps CaYMV. EVS-0036 encodes a reverse transcriptase and RNAse H. While it is not known if the EVS-0036 encoded polypeptide is translated and enzymatically active, it is worth considering the possibility that EVS-0036 is contributing

to a silencing suppression mechanism that could favor CaYMV infection. Perhaps in the absence of this element, the plant immune system may prevent active replication of CaYMV.

Iskra-Caruana et al. (2014) described in depth the co-evolution of banana streak virus, its endogenous counterparts, and its host Musa species. There are two classes of integrants. One class is capable of producing active infectious episomal virus. Another class of integrants in banana genomes are defective, fragmented, and can be dispersed among various loci and chromosomes. The latter class is disarmed and over millions of years, has become part of the plant genetic material (Iskra-Caruana et al., 2014). It is reasonable to consider ancestral Musaceae, Cannaceae, and badnaviruses co-existed in tropical environments and that a model for co-evolution for Musa and badnaviruses proposed by Iskra-Caruna et al (2014) could also extend to include *Cannaceae* and badnaviruses. Thus an ancestral CaYMV infected an ancestral Canna species. In this ancestor, there was a process of CaYMV genome decay and an integration event, leading to a pseudogenisation (Iskra-Caruana et al., 2014). Detection of EVS-0036 in hybrid cannas likely originated from a donor canna species and was incorporated into the hybrid genome through breeding programs. Thus, it is reasonable to expect that in modern domesticated cannas, which are interspecific hybrids, to observe outbreaks of CaYMV but also the occurrence of detectable badnavirus fragments in the canna genome. Further research is needed to understand the diversity of integrant canna virus sequences and their distribution among chromosomes to gain a better appreciation of the co-evolution of badnaviruses and cannas.

REFERENCES

Andrew, MQK, Michael, JA, B.C, E, and J.L, E (2012): *"Virus Taxonomy" Ninth report of the International Committee on Taxonomy of Viruses*, EC 44 Academic Press, Leuven.

Ausubel, FM, Brent, R, Kingston, RE, Moore, DD, Seidman, JG, and Struhl, K (1988): *Current Protocols in Molecular Biology*. John Wiley & Sons.

Bhat, AI, Siljo, A, and Deeshma, KP (2013): Rapid detection of Piper yellow mottle virus and Cucumber mosaic virus infecting black pepper (Piper nigrum) by loop-mediated isothermal amplification (LAMP). *J Virol Methods* **193**, 190-6.

Borah, BK, Sharma, S, Kant, R, Johnson, AM, Saigopal, DV, and Dasgupta, I (2013): Bacilliform DNA-containing plant viruses in the tropics: commonalities within a genetically diverse group. *Mol Plant Pathol* **14**, 759-71.

Bouhida, M, Lockhart, BEL, and Olszewski, NE (1993): An analysis of the complete sequence of a sugercane bacilliform virus genome infectious to banana and rice. *Journal of General virology* **74**, 15-22.

Bousalem, M, Douzery, EJ, and Seal, SE (2008): Taxonomy, molecular phylogeny and evolution of plant reverse transcribing viruses (family Caulimoviridae) inferred from full-length genome and reverse transcriptase sequences. *Arch Virol* **153**, 1085-102.

Bousalem, M, Durand, O, Scarcelli, N, Lebas, BS, Kenyon, L, Marchand, JL, Lefort, F, and Seal, SE (2009): Dilemmas caused by endogenous pararetroviruses regarding the taxonomy and diagnosis of yam (Dioscorea spp.) badnaviruses: analyses to support safe germplasm movement. *Arch Virol* **154**, 297-314.

Chauhan, RP, Hamon, HF, Rajakaruna, P, Webb, MA, and Verchot, J (2015a): Reliable Detection for Bean yellow mosaic virus, Canna yellow streak virus, and Canna yellow mottle virus in Canna Varieties with Red Foliage. *Plant Disease* **99**, 188-194.

Chauhan, RP, Rajakaruna, P, and Verchot, J (2015b): Complete genome sequence of nine isolates of canna yellow streak virus reveals its relationship to the sugarcane mosaic virus (SCMV) subgroup of potyviruses. *Arch Virol* **160**, 837-44.

Chauhan, RP, Wijayasekara, D, Webb, MA, and Verchot, J (2015c): A reliable and rapid multiplex RT PCR assay for detection of two potyviruses and a pararetrovirus infecting canna plants. *Plant Disease* **99**, 1-9.

Cisneros, FH, Zevillanos, R, and Cisneros-Zevallos, L (2009): Characterization of Starch from two Ecotypes of Andean Achira Roots (Canna edulis). *Journal of Agricultural and Food Chemistry* **57**, 7363-7368.

Covey, SN, Noad, RJ, al-Kaff, NS, and Turner, DS (1998): Caulimovirus isolation and DNA extraction. *Methods Mol Biol* **81**, 53-63.

Dahal, G, Hughes, JD, Thottappilly, G, and Lockhart, BEL (1998): Effect of temperature on symptom expression and reliability of banana streak badnavirus detection in naturally infected plantain and banana (Musa spp.). *Plant Disease* **82**, 16-21.

Dahal, G, Ortizc, R, Tenkouano, A, Hughes, J, Thottappilly, G, Vuylsteke, D, and Lockhart, BEL (2000): Relationship between natural occurrence of banana streak badnavirus and symptom expression, relative concentration of viral antigen, and yield characteristics of some micropropagated Musa spp. *Plant pathology* **49**, 68-79.

Deeshma, KP, and Bhat, AI (2015): Complete genome sequencing of Piper yellow mottle virus infecting black pepper, betelvine, and Indian long pepper. *Virus Genes* **50**, 172-5.

Dewi, K (2009): Prospect of Canna (Canna edulis Ker.) rhizome as a feedstock for bioethanol production. *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology* **153A**, S215-S215.

Flores, HE, Walker, TS, Guimaraes, RL, Bais, HP, and Vivanco, JM (2003): Andean root and tuber crops: Underground rainbows. *Hortscience* **38**, 161-167.

Franck, A, Guilley, H, Jonard, G, Richards, K, and Hirth, L (1980): Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* **21**, 285-94.

Gade, DW (1966): Achira, the Edible Canna, Its cultivation and use in the Peruvian Andes. *Economic Botany* **20**, 407-415.

Gayral, P, and Iskra-Caruana, ML (2009): Phylogeny of Banana Streak Virus reveals recent and repetitive endogenization in the genome of its banana host (Musa sp.). *J Mol Evol* **69**, 65-80.

Gayral, P, Noa-Carrazana, JC, Lescot, M, Lheureux, F, Lockhart, BE, Matsumoto, T, Piffanelli, P, and Iskra-Caruana, ML (2008): A single Banana streak virus integration event in the banana genome as the origin of infectious endogenous pararetrovirus. *J Virol* **82**, 6697-710.

Geering, AD, Olszewski, NE, Dahal, G, Thomas, JE, and Lockhart, BE (2001): Analysis of the distribution and structure of integrated Banana streak virus DNA in a range of Musa cultivars. *Mol Plant Pathol* **2**, 207-13.

Geering, AD, Parry, JN, and Thomas, JE (2011): Complete genome sequence of a novel badnavirus, banana streak IM virus. *Arch Virol* **156**, 733-7.

Geering, AD, Scharaschkin, T, and Teycheney, PY (2010): The classification and nomenclature of endogenous viruses of the family Caulimoviridae. *Arch Virol* **155**, 123-31.

Geering, ADW, McMichael, LA, Dietzegn, RG, and Thomas, JE (2000): Genetic Diversity Among *Banana streak virus* Isolates from Australia. *Virology* **90**, 921-927.

Geijskes, RJ, Braithwaite, KS, Smith, GR, Dale, JL, and Harding, RM (2004): Sugarcane bacilliform virus encapsidates genome concatamers and does not appear to integrate into the Saccharum officinarum genome. *Arch Virol* **149**, 791-8.

Givord, L, Dixon, L, Rauseo-Koenig, I, and Hohn, T (1988): Cauliflower mosaic virus ORF VII is not required for aphid transmissibility. *Ann Inst Pasteur Virol* **139**, 227-31.

Grabowski, M (2010): Canna Yellow Streak Virus. In U. o. M. Extension (Ed.): *Yard and Garden News*.

Haas, M, Bureau, M, Geldreich, A, Yot, P, and Keller, M (2002): Cauliflower mosaic virus: still in the news. *Molecular Plant Pathology* **3**, 419-429.

He, X, Fütterer, J, and Hohn, T (2002): Contribution of downstream promoter elements to transcriptional regulation of the rice tungro bacilliform virus promoter. *Nucleic Acids Res* **30**, 497-506.

Hermann, M, and Heller, J (1997): Andean roots and tubers: Ahipa, arracacha, maca, and yacon. International Plant Genetic Resources Institute,, Rome, Italy.

Hohn, T, Fiitterer, J, Basel, FMI, and Zurich, ETH (1997): The proteins and functions of plant Pararetroviruses: Knowns and Unknowns. *Critical Reviews in Plant Sciences* **16**, 133-161.

Hohn, T, and Rothnie, H (2013): Plant pararetroviruses: replication and expression. *Curr Opin Virol* **3**, 621-8.

Hsu, YC, Wang, CS, Lin, YR, and Wu, YP (2016): Structural Diversity of a Novel LTR Retrotransposon, RTPOSON, in the Genus Oryza. *Evol Bioinform Online* **12**, 29-40.

Hull, R (2007): Caulimoviridae (Plant Pararetroviruses). *Encyclopedia of life sciences*.

Iskra-Caruana, ML, Chabannes, M, Duroy, PO, and Muller, E (2014): A possible scenario for the evolution of Banana streak virus in banana. *Virus Res* **186**, 155-62.

Jakowitsch, J, Mette, MF, van Der Winden, J, Matzke, MA, and Matzke, AJ (1999): Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. *Proc Natl Acad Sci U S A* **96**, 13241-6.

Kazmi, SA, Yang, Z, Hong, N, Wang, G, and Wang, Y (2015): Characterization by Small RNA Sequencing of Taro Bacilliform CH Virus (TaBCHV), a Novel Badnavirus. *PLoS One* **10**, e0134147.

Kress, WH (1990): The Phylogeny and Classification of The Zingiberales. *Annals of the Missouri Botanical Garden* **77**, 698-721.

Kubiriba, J, Legg, JP, Tushemereirwe, W, and Adipala, E (2001): Vector transmission of Banana streak virus in the screenhouse in Uganda *Annals of Applied Botany* **139**, 37-43.

Laney, AG, Hassan, M, and Tzanetakis, IE (2012): An integrated badnavirus is prevalent in fig germplasm. *Phytopathology* **102**, 1182-9.

Leonel, M, Sarmento, SBS, Cereda, MP, and Guerreiro, LMR (2002): Extração e caracterizão do Amido de Biri (Canna edulis). *Brazilian Journal of Food Technology* **5**, 27-32.

Lockhart, BEL, and Autrey, LJC (1988): Occurrence in sugarcane of a bacilliform virus related serologically to banana streak virus *Plant Disease* **72**, 230-233.

Marino, MT, Ragozzino, E, Lockhart, BEL, Miglino, R, and Alioto, D (2008): First report of Canna yellow mottle virus (CaYMV) in Italy and in Netherlands. *Plant Pathology* **57**, 394.

Mass-Van De Kamer, H, and Mass, PJM (2008): The Cannaceae of the world. *Blumea, National herbarium Nederland* **53**, 247-318.

Mette, M, Kanno, T, Aufsatz, W, Jakowitsch, J, van der Winden, J, Matzke, M, and Matzke, A (2002): Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. *EMBO J* **21**, 461-9.

Momol, MT, Lockhart, BEL, Dankers, H, and Adkins, S (2004): Canna yellow mottle virus detected in Canna in Florida: *Plant Health Progress*. Online.

Monger, WA, Adams, IP, Glover, RH, and Barrett, B (2010): The complete genome sequence of Canna yellow streak virus. *Arch Virol* **155**, 1515-8.

Monger, WA, Harju, V, Skelton, A, Seal, SE, and Mumford, RA (2007): Canna yellow streak virus: a new potyvirus associated with severe streaking symptoms in canna. *Arch Virol* **152**, 1527-30.

Muller, E, Dupuy, V, Blondin, L, Bauffe, F, Daugrois, JH, Nathalie, L, and Iskra-Caruana, ML (2011): High molecular variability of sugarcane bacilliform viruses in Guadeloupe implying the existence of at least three new species. *Virus Res* **160**, 414-9.

Ndowora, T, Dahal, G, LaFleur, D, Harper, G, Hull, R, Olszewski, NE, and Lockhart, B (1999): Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. *Virology* **255**, 214-20.

Oana, D, Ziegler, A, Torrance, L, Gasemi, S, and Danci, M (2009): Potyviridae Family Short review. *Hortyculture, Forestry and Biochemistry* **13**, 410-420.

Pahalawatta, V, Druffel, K, and Pappu, H (2008): A new and distinct species in the genus Caulimovirus exists as an endogenous plant pararetroviral sequence in its host, Dahlia variabilis. *Virology* **376**, 253-7.

Pauli, S, Rothnie, HM, Chen, G, He, X, and Hohn, T (2004): The cauliflower mosaic virus 35S promoter extends into the transcribed region. *J Virol* **78**, 12120-8.

Pfeiffer, P, and Hohn, T (1983): Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* **33**, 781-9.

Ploetz, RC, Kema, GH, and Ma, LJ (2015): Impact of diseases on export and smallholder production of banana. *Annu Rev Phytopathol* **53**, 269-88.

Prince, LM (2010): Phylogenetic Relationships and Species Delimitation in Canna (Cannaceae) *Cann Molecular Phylogeny*, 307-330.

Rajakaruna, P, Shafiekhani, M, Kim, T, Payton, M, Chauhan, R, and Verchot, J (2013): Production of discernible disease phenotypes in Canna by five plant viruses belonging to the genera Potyvirus, Cucumovirus and Badnavirus. *Plant Pathology* **63**, 821-830.

Remans, T, Grof, CP, Ebert, PR, and Schenk, PM (2005): Identification of functional sequences in the pregenomic RNA promoter of the Banana streak virus Cavendish strain (BSV-Cav). *Virus Res* **108**, 177-86.

Specht, CD, Kress, WJ, Stevenson, DW, and DeSalle, R (2001): A molecular phylogeny of Costaceae (Zingiberales). *Mol Phylogenet Evol* **21**, 333-45.

Spetz, C, Taboada, AM, Darwich, S, Ramsell, J, Salazar, LF, and Valkonen, JP (2003): Molecular resolution of a complex of potyviruses infecting solanaceous crops at the centre of origin in Peru. *J Gen Virol* **84**, 2565-78.

Staginnus, C, and Richert-Pöggeler, KR (2006): Endogenous pararetroviruses: two-faced travelers in the plant genome. *Trends Plant Sci* **11**, 485-91.

Viswanathan, R, Alexander, KC, and Garg, ID (1996): Detection of sugarcane bacilliform virus in sugarcane germplasm. *Acta Virol* **40**, 5-8.

Xu, D, Mock, R, Kinard, G, and Li, R (2011): Molecular analysis of the complete genomic sequences of four isolates of Gooseberry vein banding associated virus. *Virus Genes* **43**, 130-7.

Zhang, Y, Angel, CA, Valdes, S, Qiu, W, and Schoelz, JE (2015): Characterization of the promoter of Grapevine vein clearing virus. *J Gen Virol* **96**, 165-9.

Zhuang, J, Wang, JH, Zhang, X, and Liu, ZX (2011): Molecular characterization of Banana streak virus isolate from Musa Acuminata in China. *Virol Sin* **26**, 393-402.

APPENDICES

Full genome sequence of CaYMV

>CaYMV

TGGTATCAGAGCGAGGTTGAAGGTATGTCTTTCATGGGGTGAAACCCTTAGATAAGGGCCTAA
GGGCTCTGTTTATATTTTCTAAGGAAAGCTGTTGAGCAAGTAGTTTTGGTTGATAAGTCGAAT
GTCATAATAACGTTTCTACTAGAAGAACTATGATGCATGCTTAGGAAAGATACCTGTTGGGAG
TAGCATCGTTTAAGGAGGAAACAATGTTATCCATGGTTCCATTCAAGGGACTAGCTAAGCTGT
GAAAAGAACAAAAGGCCGCAAAGATACCCGCAGGCAGGAGGCCGTTAGGGTGAAATAGTCG
GATGAGTTCTAGTTGGCTTGCGAAATCCTGGAACTAGATAGACAATTTACCTCCCAAGGTGTG
AAACCCGTAGGCGATTTCCAATTAAGTATGTTGAATTAGTTCTTTGAAAAAGAAACTTAAAAG
ACAGTAAGACGTAAGGCCATAGACTACCACTGGATCTGACAGTAATCCTTGAAGTATAAGCA
TCCCGAGGGGTACCCTTATCTTCTTTATCATAGAGAGACACATCTACCTTAAATCTCGTTTTAT
GTCCACAAAACAATTTTGGGATAGTAAGTTTTCTGAATTTAAAGAACATTCTGCTCTAATTTCT
GCAAATTTAGAGTATTTAGATTTAGCTTCACAGGCGAAGGTCACCGTAAAGGACCTATCGCAC
AATTTTCAAATTTTCTGTTATAGACAGGACCTTGGTCATAAGGTTTTGCTCGCGGCAACCCACA
AGAGTTGCAAAGAAGTTACAGAAGCACTTCAAGAAGACAACGAGAAGCTTCAGAAGGATCTA
AAGGCGAACAAAGCTGCTGTCTTGCTACAGCAACAGACGATCCAGAAGCTCGTGGAATCTTG
TGATAACCTTCACAAGGAGATCACTAAGCTGAAAAAGATTCCAAAACCTTTATCAAAAGAGG
ACGTTGAAGAGCTCGTGGTCAAGATATCTGAACAGCCAAAGTTTATCGAAAGACAAACTGAG
GCCCTAATTGAAGAGCTTTCCTCAAAGGTCAACAAAATCGAAGCCTTGATCCACAGGCTTGAG
AAGGTACTACTCACTTAATGAGCCTTGCCAACTCAAGAGCAAGTGCAGCCTATCAAGAAGCGT
TAGCAGCCACAACTCAAGATTGGGACACAGCGTCTGGTTTCACTGCAAAGTCAGACCTCCCTA
ATATAGCAACTATCTCAAGGCAGTTAAATTCAGCCTTATTCCTATTAGTCCGTCTTGACACCAA
GTTAGCTGCTCTAGAAGACAAGCTGCTCAGACTTGAGGCAAGAGTTCGAAATATTGAAGCGG
CAAAAGCTCAAATCGGAACAGGAGCTTCTGATTGGAAGGAAGAGCTTGAGAGACTTACTGAT
AAACTCAGTAATTTAAAGATCGGTGACTCTCAAAAGCTAAGGGAGATCGGAGGAAATCTCAA
AGTTATCAAGTCCCCTTACGAAATCCTCCAAAGCATCCAGTAAGAAGATGGCGACAAGAGGA
AGAGTGCTCGGAACAGGAGGTTCCTCAACAAATACTGAACCTGGTCAACCCCTAATAGAAGA
CCAAATTAGGGATTATAGAAGAGCACAAAGGGCGCGGTATGAAGCCGCAAGAATTGCCAGAA
ACATCGGCAATGTCGCAAGAACAATCGTAGGAAGACAACCAAGAGAGCACACTTTGTCTCTC
TTAATGGATCCTGAAGTGGAACTTCAGAGGTCTATGTCCGAAAGGGCACGGACTGTACCAGCT
GAGGTACTGTATATGACTCGGCGGGATGATGTACACCACCGTGTATACCATCATCGGAGTGAG
GAAAGAATGTTGGCTCTCACCAATGATCAGCAAGATAGGACGTTCATTCAACCGGAGTCCTAT
GAAGCGCTCGCAAGAGCAGGATTCGAGTTTATTCATCTGGGAGTCATGCAAGTTCGATTCCAG
ATATTGCACAGAAGATATGCCGGAACACTGGCCTTTTTAGTGTTTCGAGACACAAGATGGGGC
GAAGACGACAGATCTATAATCGCAGCAATGGAAATAGATCTCGCTGAAGGTAACCAACTGGT
TTACATTATACCTGATATTATGATGACAATTAAGGATTTCTGGAGGCATATCCAAATCTCTATC

AGAACCAGAGGTTACGAGACATGGGAAGGAGGAGAAGCTAATATCCTCATAACCAGAAGTAT
CACCTCAAGACTCAGCAACACCCCCAATGTTGGGTTTGCATACAAGATTGAGCAGGTAGCTGA
ATACCTGAGATCAAAGGGAGTCAAGGCCATCAACGCAACAAAACATTCTGGAAAACAGTTAC
ATGGCCAAGAATGGATGCTTAGACCATCCAAGGTGATAGCCCCCATGCAGCCATCATCAATG
AACACGGCTACTAGATACGATGGCCAGATCTCCATCAAGTTCGGAGATTATGAAGCAGCCTCA
ACAAGTAAGGCACCTACCTACAATCGTCACGACGACGAGATCGACGAGGAAGAGATCATCGC
GTTCCTGAGCGATGAAGAAGAAGAAGACCCGTGGGCTGAACTCACGGCAGAATTTGAGA
AATTTTGTCCTCAAAATTCCTCCCGGAAGGTGGTGGGAGAGAAGGAAATTGAAAATGATGAA
GAGGAGATCATTGACACTTTCTTAAAGGCTTGTGCAAGTAATGAGCCTGAAAGCTCAGACGAC
GAAAGATTAAGTTTGAAAAGCTACAACATTGCAGTTATGGAATACCCAGAGCTGAAGCCAGA
GATCAACGACATTCTGGTAAATTCCTCTGTCACTTCTGATTACAGACCACCAGAAGTTGACAT
GACTGGACCTACTGGTTATGCACCAGCTACGTCAAGAATTGGGTATCACACTGGAGAAATAA
GTGAGAGAATCGGAGGAGGGTTCAGATGGAAAAACCCTTCAGAGAATTTCCAACTACCATCA
GCGCAACAACAGTCTGGCGCTATGTTTGTCATGCCATCCAACTTTGATCCCAAAGTCTTTGAA
AGATGGGAATCAGTGGTACTAAACCACTTAGCAGATAAGAGCTTTGTGACTGCGGAAGATAA
GCTCATCTACATGGAGAATCTGCTAGGAGAGATGGAGAAGATCACGTTCCAAACATGGAGAA
TGCAATACACTGCAGAATATGACACCATGAAGGGACAAGCATTGGGAAACAACGGGACACAA
AATGTCCTAAACCAAATCAGGAGAATCTTTTATTTGGAAGACCCAAAGAGTGGAACTACTGTC
ACTCAAGATGCAGCATACAAGGCTATCAAGAGTTTGGTATGCAATGAAATGAGTGGAGAAGC
CATCAGAAGATATATGGTATCCTATCTTGACCTTGCAGCAAGAACAGGTCGAATGTGGGTATC
ATCTGAACTATCGGATGAGTTCTTCACCAAATTGCCAAGCGGCCTTGGAGACAAAGTAGGAA
AAGCTTTCAAGGAGAAGTATCCAGGCAACACTGTAGGTGTACCGGTAAGAATCACCTTCACA
CAAAACTACCTGGAGGAAATATGCAGAGAAGCTGCCTACCAAAGGAGTCTGAAGAATCTTAA
CTTCTGCAAAGAATTTCCAATACCAGGCTACTACAAGAAGCCAAATAAGAAGTATGGCAGAA
GAAGATCTACAACCTACAAAGGTAAACCTCATAAGACCCACATACGGATAGATAGAACGAAG
AATCTGAGAAAGAGAAAATGCAAATGCTACGCATGTGGCGAGGAGGGTCATTTTGCTAAGGA
TTGTACAAATCCCAGGAAGATCGTGGATAGAGTACATATCCTAGAAGATCTAGAACTGAAAG
ATGGAGTAGATGTTCTCTCAGTAGGAGAAGACGAAGACGAGCTGTCTGATATCTACTCAATCA
CATCAGGAGAAGGAGAGCCTGATGAAGAGTTAAATTCTTTAGTTTTCGCACTAGAAGCCGAA
CCAACTGAACAGCTTTTAGTAGGAGATGCAGTGAAGTCACCTTGGAGAACTCAGATGAGGGT
GACAAGAAGAGAATACTACTGCATACATAAGTGGGAATTTGAAACCACAGAACCAAGGACTT
GCAGAGCATGCAAGTTAGAAGCAAGAAAAGGCGAGAGGATGGAATGCTCTATATGTGACATG
GTAGTCTGCTGCCTTTGCAGTAACTATTGTTATGGCATACACATCCCGAGAAACAGAGTCCAT
GCTGAATACAAATCACCAGATTGGAAGCACATAGCTTTAGCACAGCATGAGATTATCAATTAC
GAGAAGACAAGAGCGGGACCAAGATACAACGGTCTCTACAACATCAAAGTAAAGCTAGAAG
TCGACGGAAGAGAAATCTCTCTTAATGCAATACTGGACACAGGAGCAACAATCTGTGTCGTG
AGAATGGAAAAACTGACGGAAGACTGGTTAGAAGATGCAGCCATGGATTACACCATCAGAGG
TGTAAATTCAGTAACAAAGGCTAACAAGGTCCTCAAGAAAGGAAAATTATGGATCGGGGAGC
AGTTCTTCAGAATCCCGAGGACAATGGCTGTAGACATGACCTTGTCAGCAGGCATCGACATGA
TCTTAGGTTGCAACTTTATAAGATCAATGGAAGGAGGTCTTAGAATAGAAGGAGAGAACATC
ACCTTCTACAAACTGACTACACACATCGAGGCATCAAAAACTGCACACGAGGTAGCCTCAATT
GAAGAGTTGGACCTCAATGAAGATGAATACTATGACATAGCCCTCACAGAAGACATCAAAGG
GTACGTCAACAGAGAAATAGTTGATACACAACTATTCAAAGAGCTAAAAGAAAATGGATACA
TGGGAGAAGAGCCATTAAAGCACTGGAAGAAAAACCAGATCAAATGTAAGCTTGAGATCAAG
AATCCGGACCTTATCATAGAAGACAGGCCCCTCAAGCACGTTACTCCCAAGATGAAAGAAGA
ATGGCAAAACATGTGAATCAGTTGTTAAAGCTTGGAGTTATCAGACCATCAAACAGCAAGCA
CAGGACAACTGCCATGCTTGTTGAATCTGGAACAGAAGTTGATCCGAAAACTGGAGAAGAGA
AGAGAGGAAAACAAAGGCTAGTCTTTAACTACAAAAGACTCAATGACAACACAGAGAAAGA
CCAGTACTCTCTGCCAGGAATCAATACAATTATCCAAAGGATCGGGCGATCAAGGGTATACTC

AAAATTTGATTTAAAATCAGGTTTTCATCAGGTAGCCATGGAAGAAGAATCCATCCCCTGGAC
GGCCTTTTGGGCTATTGATGGATTATATGAGTGGCTTGTTATGCCCTTTGGATTGAAGAATGCC
CCGGCATGTTTTCAGAGAAAGATGGATAACTGCTTCAGAGGTAAAGAACACTTCATAGCAGTT
TACATTGATGACATTCTAATATTCTCAGAAAACAAGGAGCAGCATGTGCAGCATCTGAAAGA
ATTTCTGCGCATCGTGAAGAAGGAAGGGTTGGTTCTAAGCCCAACCAAAATGAAAATTGGAG
TACCAAAAGTGGATTTCCTAGGAGCTACAATTGGCGAATCAAGAATCAAGCTACAACCTCATA
TCATCAAGAAGGTTGTGAATTTTAAAAACGAAGATCTGAAAGAAACAAAGGGATTGAGAAGT
TTCTTAGGCATTCTGAACTACGCAAGAAACTATATCCCTAATCTTGGTAAAACTCTAGGGCCA
CTTTATTCTAAGACTTCGCCGAATGGTGAGAGGCGAATGAATGCACAAGACTGGGCACTCATC
AACAAGATCAAGCAGCAAGTCCAGAACCTGCCTGACATGGAATTACCTCCAGACAACGCAGT
CATCACATTGGAAACTGACGGCTGTATGGACGGATGGGGTGGTGTTTGTAAGTGGAAAATGC
CAGGAGAACCCAAGTCAGCAGAGAAAATTTGTGCATACGCTAGTGGGAAGTTTCCTGCAATC
AAGAGCACTATAGATGCCGAGATCCAGGCTGTGATCAATAGCCTAGACAAGTTTAAAATCTAT
TATCTCGACAAAAAGGAAATTCTCATAAGAACTGATTGCCAGGCAATTGTTGCCTTCTATGCC
AAGACATCCCAGAACAAGCCTTCGCGGGTAAGATGGCTAACGTTTTCTGATTATATTACAGGG
CTGGGAATGATCGTCAAGTTCGAGCATATTGACGGCAAAGACAACACAATTGCTGATACATTG
TCAAGAATGGTGGTCATGTTCCTGAAAGAAGACTCATTCAAGAAGGAGATCCCCAAAGTTCTA
GCATGTCTTAAATATGAAGAAGACGTGAACGTTCTGACAAGGAGGCCAGTCCTCAAATGCTC
ATGCAACAAACCCGTGAAACACTGGACATCAAGGACCAGTAGAAACCCTGGACGAGTTTTCG
TGGCATGTGCTGAGCAAAAATGTCATGCATGGTGGTGGGACGATCTCATCGAAAACTATCTTG
AAGAAATTTGGAGCATGGAAGACGAAGAAGTTGGGAAGACTCCGGTCACTATGTCGCCTGGA
AGGATAGAAGACATATTTGACCTTGCAGACGTAAGCAATGACGATTGAGGAAAAGAAATCGT
AAGCAGTGACGTCAAGGCGGAAGTGGTGGACCCATACCTTTATTCCGCACTATCTAAAGTATA
GTAATTATTGTACTGTATGTGTGATTCCACTAAGTTGTGTTTGTACCACTTTGCGTTTGCTTTAC
GCAAAGGTAGATGCCCGAGCATGTGATGCGCTACCAAATGCATTATTGGTAAGCCTTCAGGCT
CTATATAAGGGGAAGCTTCAGAGCAACCCAAACCATCCAACCATCCAAGAAACCACCCAAAA
ACAAAAACCCAGTAAGCTTGTAAGAGTTTAAGTGTAAGCTTTCAGAAGTTTGAGTAATAAACT
TTCCAGTTCTAAGTTTCATACCTTGTGTTCCTGTAAGACTAACCAAGCTAGGAACAAGATCCG
CAAAAGCCAATTCCGCACGCGATTCGTT

Endogenous viral sequence 0036

>EVS-0036

ATGAATTATACAATCAAAGGCGTCAACTCCATAACAAAAGCAAACCAAGTCTTGAAAGGAGG
AAAGATCTGGATCGGGGAGCAGTATTACAGAATCCCTAGGACGATGGCCTTGGAGGTGGCCT
TATCAGAAGGGATTGACATGATAATTGGATGTAATTTCATAAGAAGTCTCGAAGGTGGAGTAC
GAATTGAAGGAGATAAGATCACATTCTACAAACTGACCACCCACATAGATACATCCAAGGAG
GCATTCACTGTGGCCTCAATAGAAGAACTCGACCTGAATGAGGACGAGTACTATGACATAGC
CCTGATGGACGAGGAGAAGGGCTACATCAACAGGGAGATAGTGGACTCCCAGCTGTTCAGAA
GCCTGAAGGAGGCAGGCTACATAGGGGAGGAGCCCCTGAAACATTGGTCGCAAAACCAGATT
AAGTGCCGTCTGGAGATCAAGAACCCAGACATGGTCATTGAGGACAGACCATTGAAGCATGT
CACGCCAAAACTGAAAGAAGACATGCGGAAGCACATAGACCAATTGCTGAAGCTGAAGGTTA
TCAGGCCCTCTGCAAGTAAGCATAGGACAACTGCTATGCTAGTGGAGTCTGGGACGGAGGTG
GATCCCAAAACAGGCCTGGAAAAGAAAGGAAAACAGAGACTGGTATTCAACTACAAGAGGTT
AAATGACAACACTGAAAAGGACCAGTATTCTCTGCCTGGGATAAATACAATTATCCAGAGGA
TTGGAAGGTCGAAGATTTACTCCAAATTCGACCTCAAAAGTGGTTTCCACCAAGTTGCCATGG
AAGCTGAGTCCATACCATGGACAGCGTTTTGGGCAATTGATGGCTTATACGAATGGCTTGTGA
TGCCATTCGGACTGAAGAATGCACCTGCAGTGTTCCAAAGAAAGATGGACAGTTGCTTTAGGG

GGACTGAAGAGTTCATTGCCGTCTACATTGACGACATATTAGTATTTTCTGAAACTCCCCAGC
AGCACGTTCAACATTTAAGGAAGTTCCTGGAGATCGTGAAGAAGAACGGTCTGGTTCTGAGCC
CAACTAAAATGAAAATCGGGGTTTCTCAGATAGACTTCCTGGGTGCAACAATAGGGCAGTCA
AGGATTAAGCTCCAGCCTCATATCATCAAGAAGGTGACTGAATTTCAAGAGGAAAAGCTTAA
TGACACTAAGGGACTGCGTCAGTTTCTCGGGATCCTGAACTACGCCAGGAACTACATCCCGAA
CCTTGGAAAGACTTTGGGGCCTCTGTACTCAAAAATATCCCCCAAAGGTGAGAAAAGAATGA
ACAAGCAGGATTGGGCGCTTGTTCAGCAGATCAAGAAGCAGGTTGCGAACCTGCCTGAAATG
GAATTACCCCCCAGGGACGCCGTCGTGGTACTTGAAACAGACGGATGTATGGATGGGTGGGG
TGGCATCTGCAAGTGGAAGATGCCCGGAGCAACAAAGGCGAGCGAGAAGGTCTGCGCCTATG
CCAGTGGGAGGTTCCCCGCTGTGAAGAGCACCATAGATGCTGAGATCCAGGCCGTGATCAAT
AGCCTGGACAAATTTAAAATTTATTACTTGGACAAGAAGGCGCTCACAATCAGAACGGACTG
CCAGGCCATAGTCTCGTTCTACGCCAAAATTGCACAGAACAAGCCATCCCGGGTAAGGTGGCT
AACGTTTTCTGATTTCATTACAGGTTTAGGTGTCGTGGTGACATTTGAACACATTGACGGCAA
GGACAACGTTCTCGCCGATACACTGTCAAGGTTAGTAGTTATGTTCCTCCAAGAGGACAAGTA
TGATGACTTGATCCAGAAGGCCTTGAAACATATCAAGGTTGAAGAGGAGGCCTACATGCTAG
CCTGGCCATGCCCGAAGCCCAATTTACTGTGCAACTGCGGGAAGCCCGCAAAAACGTGGATA
TCCAGGACAAGCAGGAACCCTGGAAGAGCATTTGTGGCGTGTGCAGAGCAGAGATGCCACGC
CTGGTGGTGGGATGATCTACTGGAAGACTGGGTTGATAGAGCCATGGAGATGAAGGACAGAT
GGCTAAGACCCTACGATGGCTATGACCCTAGCTATGGCTACCCCTGCCACGATGCACAAGATT
CCCTGGATACAAAGGACATTGCCAACGATGATTGAGCGGAAGTGGTGGACCCACCTACGCAC
TTATTGGCACCAATAATTGTACCTTTGCGTGAAGGCTAAGCAAAGGTGTTTCAATAATTGTAC
GTCGCTAGTTGTAAAGATGCTATCCTGAGCTGTAAAGGATAGACTCTTAGTAGTGTAGTGCCA
GAGAGCATGTGATACGCCACCTAAAGTTATGGTGTCGATGGGGCCCAAAGAGCACCCGACCA
TACTTAGCGTGAGCCTTCAAAGCTCTATATAATGAGGCACCCCCTCTCGTTGCTAGGCATCGA
GTCATACACTCAGTTGCTTACTTCGAGCTTGTGAGTTTGTAAGAGTTCCAGTTCT

VITA

DULANJANI S WIJAYASEKARA

Candidate for the Degree of

Master of Science

Thesis: SEQUENCING AND SHARACTERIZATION OF *CANNA YELLOW MOTTLE VIRUS*

Major Field: Plant Pathology

Biographical:

Education:

Completed the requirements for the Master of Science in Plant Pathology at Oklahoma State University, Stillwater, Oklahoma in July, 2016.

Completed the requirements for the Bachelor of Science in Botany at University of Peradeniya, Sri Lanka in 2013.

Experience:

Graduate Research Assistant, Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK, January 2014 to July 2016

Teaching Assistant (2013 February- November) Department of Botany, University of Peradeniya, Sri Lanka

Professional Memberships:

Member, American Plant Pathological Society (APS) 2014-2016