AN ABSTRACT OF THE DISSERTATION OF

Alfredo Diaz Lara for the degree of Doctor of Philosophy in Botany and Plant Pathology presented on December 16, 2016.

Title: Identification of Endogenous and Exogenous Pararetroviruses in Red Raspberry (*Rubus idaeus* L.) and Blueberry (*Vaccinium corymbosum* L.).

Abstract approved: _____

Robert R. Martin

The Pacific Northwest (Oregon and Washington in the United States and British Columbia in Canada) is one of the major producers of red raspberry (*Rubus idaeus* L.) and blueberry (*Vaccinium corymbosum* L.) in the world. The expansion of growing area with these crops has resulted in the emergence of new virus diseases that cause serious economic losses. The majority of viruses affecting plants (including blueberry and red raspberry) contain RNA genomes. In contrast, plant viruses with DNA genomes are relatively rare and most of the time ignored in virus surveys. The family *Caulimoviridae* is a group of plant pararetroviruses (reverse-transcribing viruses) with the ability to integrate their DNA into the host genome, resulting in complex molecular interactions that lead to inconsistencies in terms of detection and disease symptoms. Albeit, few studies have been conducted to determine the nature of plant pararetroviruses and their relationships with the associated host. To investigate the presence of pararetroviruses in blueberry and red raspberry, and their possible integration events, different plant material suspected to be infected with

viruses was collected in nurseries, commercial fields and clonal germplasm repositories for a period of four years. For blueberry, using rolling circle amplification (RCA) a new virus was identified and named *Blueberry fruit drop-associated virus* (BFDaV) because of its association with fruit-drop disorder. Based on the genome organization and phylogenetics, BFDaV may represent a novel genus in the *Caulimoviridae*, and it has the largest genome known thus far in this virus family. Until recently, *Rubus yellow net virus* (RYNV) has been the only plant pararetrovirus reported to occur naturally in *Rubus* spp. Over the past few years, it was observed that several plants regarded as negative for RYNV, based on graft transmission onto indicator plants, produced positive results by PCR-based assays. Additionally, these plants tested positive for after virus clean-up using thermal therapy and meristem tip-culture, which has been shown to successfully eliminate RYNV when assays were based on graft-transmission. A series of methods including bioassays, reverse transcription-PCR, RCA and Southern blotting was used to demonstrate that RYNV not only exists as infectious particles, but also as an integrated element into the red raspberry genome. This was confirmed by a further characterization of the plant-virus junctions (virus genome inserts) by next-generation sequencing. Simultaneously, an atypical strain of RYNV was described, which differed in genome organization from the previously described isolate. In addition, it was shown that the new RYNV variant lacked aphid transmissibility. Altogether, this work highlights the importance of plant pararetroviruses and describes a methodology for their identification and discrimination between endogenous and exogenous viruses in red raspberry and other plant species.

Identification of Endogenous and Exogenous Pararetroviruses in Red Raspberry (*Rubus idaeus* L.) and Blueberry (*Vaccinium corymbosum* L.)

by
Alfredo Diaz Lara

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented December 16, 2016
Commencement June 2017

Doctor of Philosophy dissertation of Alfredo Diaz Lara presented on December 16, 2016

APPROVED:

Major Professor, representing Botany and Plant Pathology

Head of the Department of Botany and Plant Pathology

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Alfredo Diaz Lara, Author

ACKNOWLEDGEMENTS

From my personal point of view, the acknowledgement section is the most important part of this manuscript, without the support of the people mentioned below to accomplish this task alone would be impossible.

I would like to begin by expressing my most sincere acknowledgement to my advisor Dr. Robert R. Martin (Bob), who allowed me to achieve my dream to attend graduate school; since the first day in his lab, he demonstrated unconditional support and immense patience. The Martin lab represents more than a workplace to me, it represents a family. There is no better way to explain it than my foster mothers: Nola, Ruth and Karen, thanks for taking care of me during these five years. Also, I need to mention to my friend, co-worker and future collaborator Patrick (Pat), and the rest of the crew (including technicians and undergrads), thanks for your help guys.

I have to acknowledge the assistance received from the researchers, office and maintenance staff at the Horticultural Crops Research Unit (HCRU) and the National Clonal Germplasm Repository of the USDA-ARS (Corvallis, OR).

My friends Javier (Pablito), Carlos (Roommate), Mario (Panchito) and Diego (Colega), thanks for all the adventures and good moments. Special mention to Diego, who without knowing me completely, he trusted and supported me even before the start of my studies. Additionally, I want to thank to the members of the Association of Latin America Students (ALAS) at OSU for making this journal more comfortable.

Gratitude to my committee members: Drs. Niklaus J. Grunwald, Valerian V. Dolja, Christopher C. Mundt and Glenn T. Howe for their guidance. In the same way, I really

appreciate the aid obtained from the folks at the Botany and Plant Pathology Department (OSU).

Finally, the most significant acknowledge is to my parents and brother, whose love and support encouraged me to reach my goals. I just hope that you are proud of me as I am of you.

“Gracias”

CONTRIBUTION OF AUTHORS

Dr. Robert R. Martin provided advice and assistance during the entire research project, also, he helped to edit the manuscript. Nola J. Mosier designed the primers used to assess the presence of RYNV in red raspberry cv. 'Baumforth's Seedling A' (Chapter 3). As well, Nola J. Mosier conducted the preliminary tests for Chapter 4, which included the PCR-based test for RYNV, the thermal therapy combined with meristem-tip culture and the graft transmission assay. For both Chapters 3 and 4, Karen E. Keller assisted in the genome-assembly process and sequence analysis. Final versions of chapters were approved by all co-authors.

TABLE OF CONTENTS

Page

Chapter 1: Introduction……..1

Chapter 2: *Blueberry fruit drop-associated virus*: A new member of the family *Caulimoviridae* isolated from blueberry exhibiting fruit-drop symptoms.…….....6

Abstract……..7

Introduction……7

Materials and methods……8

Plant material……..8

DNA extraction, rolling circle amplification (RCA), enzymatic digestion, cloning and resequencing.……..9

Genomic analysis of BFDaV……..9

Phylogenetic analysis……10

Virus detection……...10

Results……11

Virus characterization…….……...11

Taxonomic placement……13

Virus detection and diversity……..13

Discussion…..……15

References……..16

Chapter 3: A variant of *Rubus yellow net virus* with altered genomic organization……23

Abstract……..24

Introduction……24

TABLE OF CONTENTS (Continued)

Page

Materials and methods……...27

Plant material…….27

Rolling circle amplification, enzymatic digestion, cloning, primer walking and resequencing……..27

Genomic analysis of RYNV-BS……28

Phylogenetic analysis……29

Aphid transmission experiment……..30

Results……30

Virus characterization……..……...30

Phylogenetic analysis.……32

Acquisition and transmission of viral particles by aphids…….....32

Discussion…..…….33

References……..35

Chapter 4: Evidence of *Rubus yellow net virus* integration into the red raspberry genome……...42

Abstract……..43

Introduction…….……...44

Materials and methods…….……...45

Plant material and preliminary tests…….……..45

Graft transmission……..……46

Nucleic acids extraction……..……...46

Molecular techniques: RT-PCR, RCA and Southern blotting.……...47

TABLE OF CONTENTS (Continued)

Page

Next-generation sequencing……..…….48

Results……49

Graft transmission…….…….49

RT-PCR…….……49

RCA…….…….....49

Southern blotting…….…….....49

Next-generation sequencing……..……50

Discussion…..…….50

References……..53

Chapter 5: Conclusions……..61

Bibliography……..66

Appendix……73

LIST OF FIGURES

Figure Page

2.1. Symptoms of fruit abortion caused by *Blueberry fruit drop-associated virus*……..19

2.2. Phylogenetic inference of *Blueberry fruit drop-associated virus* (BFDaV) in relation to members of the family *Caulimoviridae* ……20

2.3. Restriction profile of *Blueberry fruit drop-associated virus* (BFDaV) and amplicons of BFDaV from blueberry 'Bluecrop', using detection primers……...21

2.4. Genomic organization of *Blueberry fruit drop-associated virus* (BFDaV), *Cauliflower mosaic virus* (CaMV, genus *Caulimovirus*) and *Petunia vein clearing virus* (PVCV, genus *Petuvirus*)……22

3.1. Representation of *Rubus yellow net virus* isolate "Canada" (RYNV-Ca) and *Rubus yellow net virus* isolate "Baumforth's Seedling A" (RYNV-BS) genomes.....38

3.2. Phylogenetic analysis of the genus *Badnavirus*……..39

3.3. Alignment of two different ORFs present in RYNV-BS and RYNV-Ca genomes……..40

4.1. RT-PCR assay for RYNV in red raspberry samples……..…….56

4.2. Agarose gel electrophoresis of RCA products previously digested with *Bam*HI restriction enzyme.…….57

4.3. Southern blot analysis for RYNV using red raspberry DNA treated with *Bam*HI….58

4.4. Diagram of RYNV-like contigs mapped against the RYNV genome……59

LIST OF TABLES

Table Page

3.1. Comparison of coding regions present in *Rubus yellow net virus* isolates "Canada" and "Baumforth's Seedling A", determining positions and lengths of ORFs in both genomes……41

4.1. *Rubus* cultivars involved in preliminary tests for RYNV presence……60

LIST OF APPENDICES

Appendix Page

Supplemental Table S 2.1. Primers used for BFDaV sequencing……74

Supplemental Table S 2.2. PCR detection of BFDaV from 'Bluecrop' blueberries in a row of 66 blueberry plants with approximately one half of plants exhibiting symptoms from a field in northern Washington state…….....76

Supplemental Table S 3.1. Primers used for RYNV-BS sequencing……78

Supplemental Table S 3.2. Viruses and sequences used in the phylogenetic analysis, including GenBank accession numbers.……79

Supplemental Table S 4.1. Contigs searched against the RYNV genome……80

DEDICATION

I want to dedicate this work to my lovely wife Norma and my daughter Maia (my baby).
Every achievement and goal that I accomplish is because of you.

Chapter 1:

Introduction

Alfredo Diaz-Lara

Berries (also known as small fruits) are a group of flowering plants, which produce edible nutritious fruits. Two of the most popular berries are red raspberry (*Rubus idaeus* L., family *Rosaceae*) and blueberry (*Vaccinium corymbosum* L., family *Ericaceae*), both of which are cultivated widely in the Pacific Northwest (PNW) (Oregon and Washington in the United States and British Columbia in Canada). According to data from the United Nations Food and Agriculture Organization (http://faostat3.fao.org), the PNW is the first and third largest producer of blueberry and red raspberry in the world, respectively.

Berry production is relatively more expensive than other commercial crops, and profits depend greatly on yield and market price (Julian et al., 2011). Current production practices for blueberry and red raspberry include management of virus diseases, because of their negative impact on production. In red raspberry, viruses have been reported to reduce fruit yields by as much as 70 percent (Ellis and Nita, 2008). Red raspberry and blueberry are perennial crops, thus, they are exposed to virus threats during the life of a planting. Additionally, these plants are propagated vegetatively, which increases the risk of spreading infection with planting stock.

A basic classification of viruses is based on nucleic acids type (RNA or DNA) of their genomes. In the case of plant viruses the majority of species contain an RNA genome. Of the total number of viruses reported in *Vaccinium* spp., less than 20% have DNA as genetic material (Martin et al., 2012b); a similar pattern is observed in *Rubus* spp., where, *Blackberry virus A* and *Rubus yellow net virus* (RYNV) are the only viruses with DNA genomes (Jones et al., 2002; Martin et al., 2012a; Shahid et al., 2016). Because of the preponderance of RNA viruses affecting plants, in the event of an emerging disease, a DNA viral element is rarely considered.

One example of DNA plant viruses is the group *Caulimoviridae* or plant pararetroviruses, whose distinctive feature is their replication via reverse transcription. Briefly, such replication begins with the transcription of the viral double-stranded DNA (dsDNA) in the nucleus of the host cell and then the resulting RNA is retro-transcribed into new dsDNA in the cytoplasm (Harper et al., 2002). Caulimoviruses are responsible for several severe diseases in commercial crops, including banana streak, rice tungro and cacao swollen shoot (Geering, 2007).

In recent years, a new syndrome was observed in the blueberry cultivar 'Bluecrop' in the PNW. The disease symptoms involved premature fruit-drop and reddening of young leaves and flowers (Martin et al., 2008). Numerous studies failed to identify the causal agent of the disorder, comprising fungi, bacteria or viruses (Martin et al., 2006); in the last case specifically, attempts of virus purification, mechanical transmission to indicator plants and dsRNA extraction were unsuccessful (Martin et al., 2006). As a result failure to identify a causal agent for fruit-drop with other methods, a novel detection method (rolling circle amplification) focused on DNA viruses was tried.

RYNV infects raspberry and blackberry species and belongs to the genus *Badnavirus,* one of the eight genera in the family *Caulimoviridae* (other genera are: *Caulimovirus*, *Cavemovirus*, *Petuvirus*, *Solendovirus*, *Soymovirus*, *Tungrovirus,* and *Rosadnavirus*) (Anonymous, 2016; Geering and Hull, 2012; Jones et al., 2002; Regenmortel et al., 2000). RYNV is part of a complex of aphid-borne viruses of *Rubus* spp. that cause raspberry mosaic disease, which severely effects plant vigor as well as fruit yield and quality (Jones et al., 2002). After this work was initiated, there was one complete genomic sequence of RYNV deposited in the GenBank (Kalischuk et al., 2013). This work resulted in

characterization of an isolate of RYNV believed to have originated in Scotland that is quite diverse from the previously reported sequence. The strain of RYNV reported here was isolated from 'Baumforth's Seedling A' red raspberry obtained from the National Clonal Germplasm Resource in Corvallis, Oregon that was originally obtained from Scotland. This European strain of the virus (RYNV-BS) was characterized at the molecular level, and in terms of transmission properties.

Another peculiarity about caulimoviruses is their ability to integrate into host genomes, though this is not required for replication. Pararetroviruses lack long terminal repeats and do not encode for an integrase, which are basic elements necessary for integration of retroviruses (Staginnus and Richert-Pöggeler, 2006). At present, there are reports of endogenous pararetroviruses (EPRVs) in species of dicotyledonous and monocotyledonous plants (Harper et al., 2002; Staginnus and Richert-Pöggeler, 2006). Most of EPRVs are benign components for the host, because the viral genome is segmented and essential open reading frames have been lost (Kunii et al., 2004). On the other hand, if the whole pararetroviral genome is nearly intact, the EPRV can 'pop out' from the host DNA and initiate an infection as has been shown for incorporated sequences of *Petunia vein clearing-virus* (Richert-Poggeler et al., 2003) and *Banana streak virus* (Ndowora et al., 1999), which were reactivated after abiotic stress or tissue culture of the host plant.

Over the past decade, it was observed that some red raspberry plants thought to be free of RYNV based on aphid or graft transmission onto biological indicators (*R. occidentalis*), produced positive results when indexed by PCR-based assays. Moreover, the same plants tested positive for RYNV after treatment by heat therapy and meristem-tip culture for virus elimination, which is the standard procedure to eradicate RYNV and other plant viruses.

By comparison, raspberry plants with graft-transmissible RYNV that were treated by thermal therapy and meristem-tip culture were freed of graft-transmissible RYNV. RYNV is on the quarantine pest/pathogen list of several importing countries. Once an RYNV specific PCR assay was developed, certified raspberry plants were rejected on multiple occasions by quarantine personnel without further confirmation. In two instances, rejected plants were obtained for further testing and were negative in bioassays and the bioassay test plants were negative by PCR, suggesting a lack of transmission rather than a symptomless isolate of the virus (Martin, unpublished). Overall, the PCR-positive results contradicting the lack of induced symptoms in bioassays suggest the presence of an EPRV. With the purpose of exploring the possibility that RYNV is integrated into the red raspberry genome, different techniques were used to discriminate between exogenous and endogenous forms of the virus. Such methods included bio-indexing, reverse transcription-PCR, rolling circle amplification and Southern blotting, which have been employed previously in other works with EPRVs (Harper et al., 1999; Laney et al., 2012; Shahid et al., 2016). Finally, next-generation sequencing was used to identify the virus-like sequences in the red raspberry genome, as recently done in 'Carrizo' citrange (Roy et al., 2014).

In summary, this dissertation describes the application of rolling circle amplification for the characterization of DNA viruses is blueberry and raspberry. Also a diverse strain of RYNV is described and the integration (RYNV) into the red raspberry genome is documented.

Chapter 2:

Blueberry fruit drop-associated virus: A new member of the family *Caulimoviridae* isolated from blueberry exhibiting fruit-drop symptoms

Alfredo Diaz-Lara and Robert R. Martin

Plant Disease
3340 Pilot Knob Road
Saint Paul, MN, 55121
United States of America
100: 2211-2214

Abstract

This study describes the nucleotide sequence and genome organization of a new DNA virus isolated from 'Bluecrop' blueberry plants exhibiting fruit-drop symptoms and named *Blueberry fruit drop-associated virus* (BFDaV). Blueberry fruit-drop disease was first detected in blueberry plants in British Columbia, Canada in the late 1990's, and in a single field in northern Washington state in the U.S. in 2012. Infected bushes abort nearly 100% of their fruit about three weeks prior to harvest, when the berries are about 3-5 mm in diameter. At harvest the affected plants appear taller than healthy ones as there is no fruit weighing down the branches. The virus was amplified from diseased material using rolling circle amplification, followed by enzyme digestion, cloning and sequencing. The full genome of BFDaV is 9,850 bp in length, and contains a single open reading frame (ORF), encoding for a polyprotein and a large noncoding region. Based on the genome size and organization, and phylogenetics, BFDaV is proposed as a new and the largest member of the family *Caulimoviridae*. Finally, in mapping part of a field with fruit-drop symptoms there was a nearly perfect correlation between the presence of the virus and fruit-drop symptoms.

Introduction

The Pacific Northwest (Oregon and Washington in the U.S. and British Columbia in Canada) is the world's largest blueberry (*Vaccinium corymbosum;* family *Ericaceae*) production area with greater than 175,000 million pounds produced in 2014 (http://www.bcblueberry.com/; http://www.nass.usda.gov/Data_and_Statistics/). Recently, a fruit-drop symptom has been observed in several fields of 'Bluecrop', in the

Fraser River Valley in northwest Washington state and southwest British Columbia, Canada. Also, it was observed that young leaves exhibited a transient red coloration of the veins during the bloom period and the corolla of the flowers exhibited some red striping. After bloom the plants appear normal until about three weeks prior to harvest, when the fruit drops (Figure 2.1). Prior to harvest, affected bushes can be identified easily since they stand upright as there is no fruit weighing down the branches. Several studies failed to identify any fungi, bacteria or virus related to the new disorder (Martin et al., 2006). Later, a cryptic virus (*Blueberry latent virus*) was characterized but was not associated with this or any disease in blueberry, and it was widespread in all production blueberry areas, in multiple cultivars of highbush blueberry, and in multiple species of *Vaccinium* (Martin et al., 2011).

In this manuscript, we describe the genome of *Blueberry fruit drop-associated virus* (BFDaV), a new member of the family *Caulimoviridae*, whose properties differ from other genera in the family and provide evidence that it is the causal agent of the disease.

Materials and methods

Plant material. Leaves were collected from blueberry 'Bluecrop' plants that exhibited fruit-drop symptoms and from asymptomatic plants during 2014 and 2015 growing seasons from commercial fields in the state of Washington in the United States (one field) and British Columbia, Canada (two fields). The plant material, used for nucleic acid extractions and cloning, was tested for all known viruses of blueberry and found to be infected only with *Blueberry latent virus* (BBLV).

DNA extraction, rolling circle amplification (RCA), enzymatic digestion, cloning and resequencing. Total DNA was extracted from leaf tissue as described by Lockhart (1990), with the addition of a proteinase K digestion prior to the nucleic acid extractions with phenol-chloroform-isoamyl alcohol (25:24:1) and chloroform-isoamyl alcohol (24:1). Purified DNA was used as template for RCA, using the Illustra TempliPhi 500 Amplification Kit (GE Healthcare, Buckinghamshire, UK) and following the manufacturer's instructions. RCA exponentially amplifies single- or double-stranded circular DNA via random hexamer primers and Phi29 DNA polymerase (Dean et al., 2001; Lizardi et al., 1998). The RCA product then was digested with *Eco*RI (New England Biolabs, Ipswich, MA, USA) and separated using agarose gel electrophoresis. The *Eco*RI-digestion resulted in several restriction fragments, which were cloned into pBluescript II KS(+) vector using Quick T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA) and transformed into One Shot Top 10 chemically competent cells of *Escherichia coli* (Invitrogen Corp., Carlsbad, CA, USA). Plasmids were sequenced using M13 primers, followed by primer walking, and finally primers were designed and used to sequence across the *Eco*RI restriction sites (Supplemental Table S 2.1). Sequences were assembled using Geneious v4.8 (Kearse et al., 2012) to produce a preliminary sequence of BFDaV. In order to confirm the BFDaV sequence, additional primers were designed that overlapped and the entire genome of BFDaV was resequenced threefold developing a consensus sequence.

Genomic analysis of BFDaV. Nucleotide (nt) and translated amino acid (aa) sequence of BFDaV were compared to existing sequences in GenBank. ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) was used to identify ORFs, which were also

analyzed independently using BLASTp (Gish and States, 1993) and the Conserved Domain search tool (Marchler-Bauer and Bryant, 2004). Also, aa sequences were aligned and the pairwise scores calculated using ClustalW2 (Larkin et al., 2007) and PIR (protein information resource) (http://pir.georgetown.edu/pirwww/search/comp_mw.shtml) was employed to calculate the molecular mass of the polyprotein. Nuclear export signals and cleavage sites of protease were predicted using NetNES1.1 (La Cour et al., 2004) and PROSPER (Song et al., 2012), respectively. Patterns and profiles were predicted using ELM (Dinkel et al., 2013). Finally, PLACE (Higo et al., 1999) was used to search for plant *cis*-acting regulatory DNA elements.

Phylogenetic analysis. The phylogenetic analysis of BFDaV was performed in Mesquite (Maddison and Maddison, 2001) and RAxML (Stamatakis, 2014), using the JTT aa model of substitution and was confirmed with a bootstrap of 100 pseudoreplicates. A maximum likelihood tree was obtained, based on the aa sequence alignment of reverse transcriptase (RT) present in different members of the family *Caulimoviridae* (viruses and the accession numbers are listed in Figure 2.2). The phylogram was visualized in FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/) and shows all clades with a bootstrap value 50% or greater.

Virus detection. Specific primers (FruitdropFr1F1: 5'-GACAACAGCATCTACATCTCTGC-3' and FruitdropFr1R1: 5'-GGTCGTTCTACCACGTTTGTG-3') that flank conserved regions of the virus aspartate protease and amplify a 395-bp amplicon were designed to test for BFDaV in plant tissue

by polymerase chain reaction (PCR). DNA extracts previously obtained were used as templates for the PCR amplification, using *Taq* polymerase (GenScript, Piscataway, NJ, USA) and a program consisting of 3 min at 95 °C; 40 cycles of 40 s at 95 °C, 40 s at 53 °C, 40 s at 72 °C; and a final elongation step of 7 min at 72 °C.

Results

Plants in the three fields studied tested negative for *Blueberry leaf mottle*, *Blueberry scorch*, *Blueberry shoestring*, *Blueberry shock*, *Peach rosette mosaic*, *Tobacco ringspot* and *Tomato ringspot viruses* by enzyme-linked immunosorbent assay (ELISA) and for *Blueberry red ringspot virus* and *Blueberry virus A* by RT-PCR. The plants did test positive for BBLV, which has not been associated with any symptoms in *Vaccinium* species. These results suggested the possibility that an unknown agent was responsible for the observed fruit-drop symptoms.

Virus characterization. Three restriction fragments (bands) were observed in the agarose gel and later fully sequenced (Figure 2.3A). The two larger restriction fragments of the RCA products encompassed the entire viral genome; furthermore, only two *Eco*RI recognition sites (GAATTC) were identified in the BFDaV genome, confirming accuracy of the restriction profile from the RCA product. The third band from the *Eco*RI-digested RCA reaction (1,281 bp) aligned with chloroplast sequence from blueberry and was not present in the resequenced BFDaV.

The complete genomic sequence of BFDaV consisted of 9,850 bp (GenBank accession number KT148886), making it the largest known member in the family *Caulimoviridae*.

The intergenic region (IR) of BFDaV was 2,056 bp and contained different plant *cis*-acting regulatory DNA elements, such as five GATA-boxes and one TATA-box; these promoter elements were identified previously in other *Caulimoviridae* viruses (Lam and Chua, 1989; Medberry et al., 1992). A plant $tRNA^{Met}$ ($TGGTATCAGAGC_{1-12}$) and priming site for reverse transcription was predicted within the IR (Medberry et al., 1990). This noncoding RNA was chosen as the starting location for the circular genome, as has been done for other caulimoviruses (Bouhida et al., 1993; Laney et al., 2012).

The nt analysis identified similarity between BFDaV and a reduced number of members of the family *Cauilimoviridae*. The genome of BFDaV shared 66% nt sequence identity with *Strawberry vein banding virus* (SVBV) and *Cauliflower mosaic virus* (CaMV), but only over a small portion of the genome (4% and 5% respectively), the rest of the BFDaV genome does not match with any identified viral sequence.

Unlike SVBV and CaMV, the new species contains a single ORF encoding for a large polyprotein (Figure 2.4). Homology was observed between aa sequences of the BFDaV polyprotein and SVBV ORF 5 (50% aa identities), *Rudbeckia flower distortion virus* (RuFDV) putative enzymatic polyprotein (48% aa identities) and *Horseradish latent virus* polyprotein (45% aa identities), only over 24% of the polyproteins. Except for RuFDV (unclassified genus), the viruses mentioned that exhibit some homology to BFDaV belong to the genus *Caulimovirus*, whose genomes contain six or seven ORFs and contrast with the single polyprotein identified in BFDaV. This polyprotein of BFDaV has a length of 2,597 aa, molecular mass of 301.6 kDa and contains the following signature domains: a movement protein (Ile_{327}-Ser_{446}), previously identified and characterized from CaMV (Kasteel et al., 1996); a zinc knuckle (Cys_{997}-Pro_{1011}), which is a zinc binding motif

$CX_2CX_4HX_4C$ (where X can be any amino acid) from retroviral gag proteins (nucleocapsid) (Zhou et al., 2007); pepsin-like aspartate protease (Leu_{1380}-Phe_{1461}); reverse transcriptase (Lys_{1752}-Gly_{1902}) responsible for the replication of retroelements including caulimoviruses (Xiong and Eickbush, 1990) and a ribonuclease H (RNase H) (Val_{2000}-Arg_{2122}). Additionally, a coiled-coil region was found upstream of the zinc finger domain as well as one nuclear export signal. Based on previous information and identification of several protease cleavage sites, it is suggested that the BFDaV polyprotein is post-translationally cleaved to yield a movement protein, a coat protein and a replicase, which are key components of most plant pararetroviruses.

Taxonomic placement. Phylogenetic analyses placed BFDaV within the family *Caulimoviridae* but distinct from the currently recognized genera (Figure 2.2). In addition, the new virus is evolutionary distant from *Petunia vein clearing virus* (PVCV), another species in the family *Caulimoviridae*, genus *Petuvirus*, with a single ORF (Richert-Pöggeler and Shepherd, 1997). Although, the genomic arrangement of BFDaV is similar to the *Petuvirus*; no sequence similarities were identified between BFDaV and the *Petuvirus* using BLASTn and BLASTx, and a ClustalW2 alignment between polyproteins of PVCV and BFDaV produced a low pairwise score (14.82%). Additionally, the genome of BFDaV is 2,644 bp larger than PVCV.

Virus detection and diversity. During the summer of 2014 and spring of 2015, a total of 344 plants were tested via PCR, resulting in 71 positive results for BFDaV (data not shown). The infected samples were collected in three different blueberry fields in the Fraser

River Valley that were separated by 50-60 km from each other. Subsequently, the RNase H and coat protein of an isolate of BFDaV from each of the three fields were sequenced and compared (data not shown) to estimate the virus diversity, the three sources were very similar, having nt sequence identities ranging from 96 to 100% in the case of the RNase H and 92 to 93% for the coat protein. This is a limited number of isolates, but at this time these are the only three fields identified with fruit-drop symptoms.

To examine the relationship of BFDaV with symptoms observed in the field, a section of the field with approximately 50% incidence of fruit-drop symptoms was used. Sixty-six plants in a single row of blueberries in the field in Washington state were evaluated for virus symptoms and tested for BFDaV by PCR in the spring and summer of 2015 (Supplemental Table S 2.2). In spring (mid-May) there were 34 plants that exhibited vein reddening in the young leaves, 31 plants without any vein reddening, and one plant that was not rated, since it had symptoms of *Blueberry shock virus* (BlShV), in which most leaves were necrotic and subsequently tested positive for BlShV by ELISA. The 34 symptomatic plants as well as the plant with BlShV were positive for BFDaV by PCR and the 31 asymptomatic plants were negative. The 34 plants with leaf vein reddening symptoms in the spring exhibited fruit-drop symptoms prior to harvest in July and tested positive for BFDaV, when tested using the detection primers in PCR assays in July. The 31 asymptomatic plants in spring did not show any fruit-drop symptoms in the summer and they all tested negative for BFDaV in PCR assays in July. The one plant that had symptoms of BlShV was not rated for fruit-drop symptoms, since BlShV causes a severe flower necrosis and there was very little fruit produced on the plant, but it was positive for BFDaV in July.

Fields of ‘Duke’, ‘Liberty’, and ‘Bluejay’ growing adjacent to infected blocks of ‘Bluecrop’ did not exhibit any fruit drop symptoms and tested negative for BFDaV in PCR assays of 96 plants.

Discussion

There are eight genera (*Caulimovirus*, *Badnavirus*, *Cavemovirus*, *Petuvirus*, *Solendovirus*, *Soymovirus, Tungrovirus,* and *Rosadnavirus*) in the family *Caulimoviridae,* but BFDaV did not have a level of sequence similarity that placed it in any of the eight genera (Hull and Geering, 2012). Even considering the low sequence homology to known members of the genus *Caulimovirus*, the genome arrangement of the new virus was similar to that of genus *Petuvirus* (Figure 2.4). Consequently, we propose that BFDaV represents a novel genus in this family.

Using RCA and later detection primers (PCR), we confirmed the presence of the virus from symptomatic plants but not from healthy plants (Figure 2.3), supporting the hypothesis that BFDaV is the cause of the fruit-drop symptom observed in ‘Bluecrop’ blueberry. Also, there was an excellent correlation between the presence of BFDaV and the fruit-drop symptoms, which supports the assertion that BFDaV is the causal agent of the syndrome. At the present time this disease is restricted to a few commercial fields in the Fraser River Valley in southwestern British Columbia and northwest Washington state. Measures are to be taken to limit the distribution of this virus and prevent it entering nurseries and germplasm repositories. PCR testing for BFDaV has been added to G1 (top tier) plants in the National Clean Plant Network. Additionally, efforts are underway to determine if this virus can be detected in native vegetation in the Fraser River Valley adjacent to fields

where the virus has been detected and to work with growers to eradicate this virus from production fields. BFDaV has been detected only in blueberry ‘Bluecrop’, and not in several other blueberry fields, ranging from 5 to 35 years old, adjacent to infected blocks with infected ‘Bluecrop’ plants. Graft transmission studies to other blueberry cultivars is underway.

Finally, mechanical transmissions were attempted from flower tissue and young leaves of symptomatic blueberry plants to several species of herbaceous host indicator plants (*Cucurbita pepo, Glycine max, Nicotiana occidentalis, N. tabacum, N. sylvestris, N. benthamiana, Cucumis sativus, Phaseolus vulgaris, Chenopodium amaranticolor, Chenopodium quinoa, Beta vulgaris, Solanum lycopersicum* and *Brassica rapa*). All attempts at mechanical transmission from blueberry plants exhibiting the fruit-drop symptom have been unsuccessful.

References

Bouhida, M., Lockhart, B.E., Olszewski, N.E., 1993. An analysis of the complete sequence of a *Sugarcane bacilliform virus* genome infectious to banana and rice. J. Gen. Virol. 74, 15–22.

Dean, F.B., Nelson, J.R., Giesler, T.L., Lasken, R.S., 2001. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. 11, 1095–1099.

Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Gibson, T.J., 2013. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. D259-D266.

Gish, W., States, D.J., 1993. Identification of protein coding regions by database similarity search. Nature Genetics 3, 266-272.

Higo, K., Ugawa, Y., Iwamoto, M., Korenaga, T., 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 27, 297-300.

Hull, R., Geering, A., 2012. Virus taxonomy. Family *Caulimoviridae.* IXth report of the ICTV. In: King, A.M.Q., Lefkowitz, E., Adams, M.J. Carstens, E.B. (eds.). Elsevier Academic Press. 429-443.

Kasteel, D.T.J., Perbal, M.C., Boyer, J.C., Wellink, J., Goldbach, R.W., Maule, A.J., Van Lent, J.W.M., 1996. The movement proteins of *Cowpea mosaic virus* and *Cauliflower mosaic virus* induce tubular structures in plant and insect cells. J. Gen. Virol. 77, 2857–2864.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

La Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K., Brunak, S., 2004. Analysis and prediction of leucine-rich nuclear export signals. Protein Engineering Design and Selection 17, 527-536.

Lam, E., Chua, N.H., 1989. ASF-2: a factor that binds to the *Cauliflower mosaic virus* 35S promoter and a conserved GATA motif in Cab promoters. Plant Cell Online 1, 1147–1156.

Laney, A.G., Hassan, M., Tzanetakis, I.E., 2012. An integrated badnavirus is prevalent in fig germplasm. Phytopathology 102, 1182–1189.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., Ward, D.C., 1998. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat. Genet. 19, 225–232.

Lockhart, B.E.L., 1990. Evidence for a double-stranded circular DNA genome in a second group of plant viruses. Phytopathology 80, 127-131.

Maddison, W.P., Maddison, D.R., 2001. Mesquite: a modular system for evolutionary analysis, version 0.98. http://mesquiteproject.org

Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32, W327-W331.

Martin, R.R., Tzanetakis, I.E., Sweeney, M., Wegener, L., 2006. A virus associated with blueberry fruit drop disease. Acta Hortic. 715, 497–501.

Martin, R.R., Zhou, J., Tzanetakis, I.E., 2011. *Blueberry latent virus*: An amalgam of the *Partitiviridae* and *Totiviridae*. Virus Res. 155, 175–180.

Medberry, S.L., Lockhart, B.E.L., Olszewski, N.E., 1990. Properties of *Commelina yellow mottle virus*'s complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. Nucleic Acids Res. 18, 5505–5513.

Medberry, S.L., Lockhart, B.E., Olszewski, N.E., 1992. The *Commelina yellow mottle virus* promoter is a strong promoter in vascular and reproductive tissues. Plant Cell Online 4, 185–192.

Richert-Pöggeler, K.R., Shepherd, R.J., 1997. *Petunia vein-clearing virus*: A plant pararetrovirus with the core sequences for an integrase function. Virology 236, 137–146.

Song, J., Tan, H., Perry, A.J., Akutsu, T., Webb, G.I., Whisstock, J.C., Pike, R.N., 2012. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. PloS one 7, e50300.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353–3362.

Zhou, J., Bean, R.L., Vogt, V.M., Summers, M., 2007. Solution structure of the *Rous sarcoma virus* nucleocapsid protein: μΨ RNA packaging signal complex. J. Mol. Biol. 365, 453–467.

Figure 2.1. Symptoms of fruit abortion caused by *Blueberry fruit drop-associated virus*. Healthy plant with normal fruit load in the foreground, fruit drop plants in the background without any fruit.

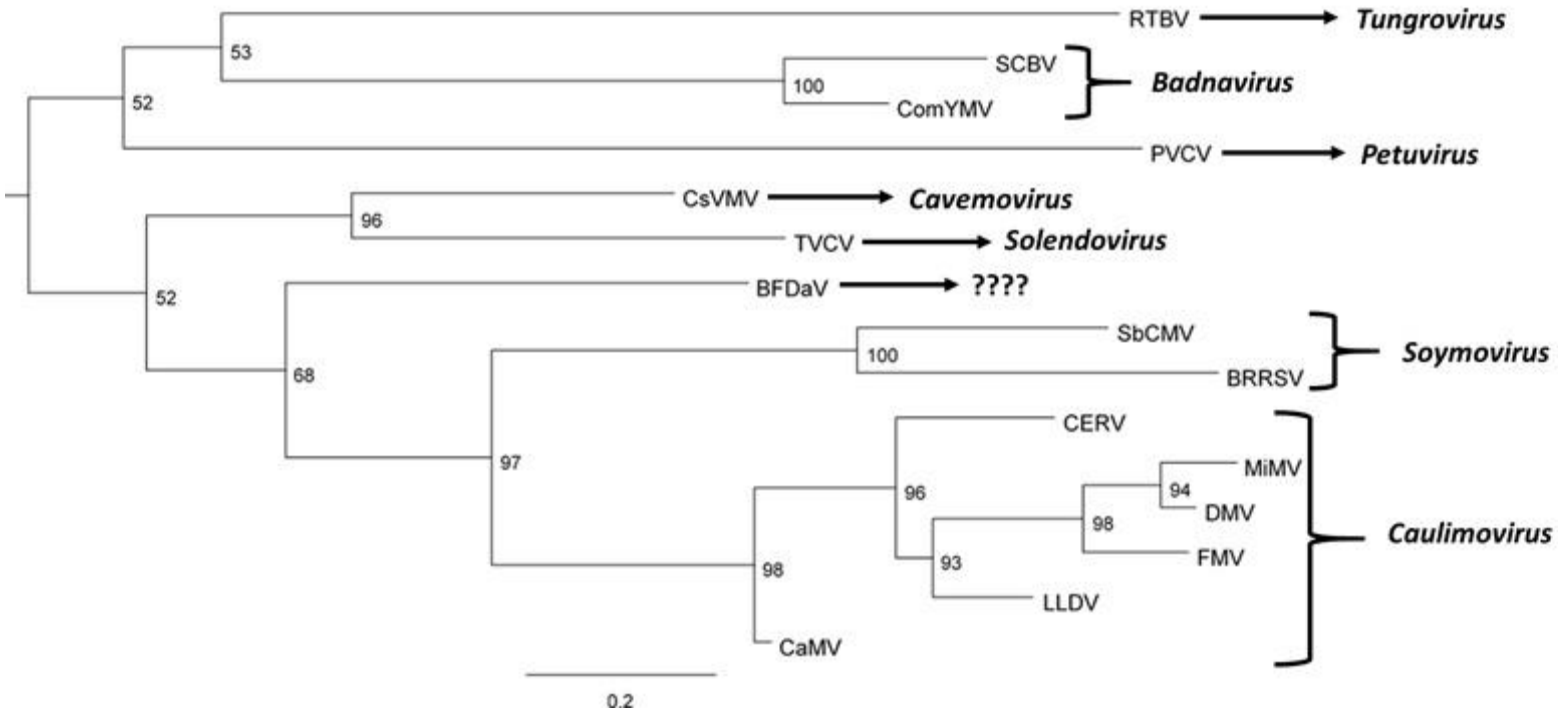

Figure 2.2. Phylogenetic inference of *Blueberry fruit drop-associated virus* (BFDaV) in relation to members of the family *Caulimoviridae*. *Cauliflower mosaic virus* (CaMV, M90543.1); *Carnation etched ring virus* (CERV, NC_003498.1); *Lamium leaf distortion associated virus* (LLDV, NC_010737.1); *Figwort mosaic virus* (FMV, NC_003554.1); *Dahlia mosaic virus* (DMV, NC_018616.1); *Mirabilis mosaic virus* (MiMV, NC_004036.1); *Blueberry red ringspot virus* (BRRSV, NC_003138.2); *Soybean chlorotic mottle virus* (SbCMV, NC_001739.2); *Blueberry fruit drop-associated virus* (BFDaV, KT148886); *Cassava vein mosaic virus* (CsVMV, NC_001648.1); *Tobacco vein clearing virus* (TVCV, NC_003378.1); *Petunia vein clearing virus* (PVCV, NC_001839.2); *Rice tungro bacilliform virus* (RTBV, D10774.1); *Sugarcane bacilliform virus* (SCBV, NC_013455.1); *Commelina yellow mottle virus* (ComYMV, NC_001343.1). Maximum likelihood tree based on the amino acid sequences of the conserved reverse transcriptase (RT) using the JTT model of substitution. Horizontal branch length is proportional to genetic distance; the scale bars represent changes per site. Bootstrap values less than 50% are not shown.

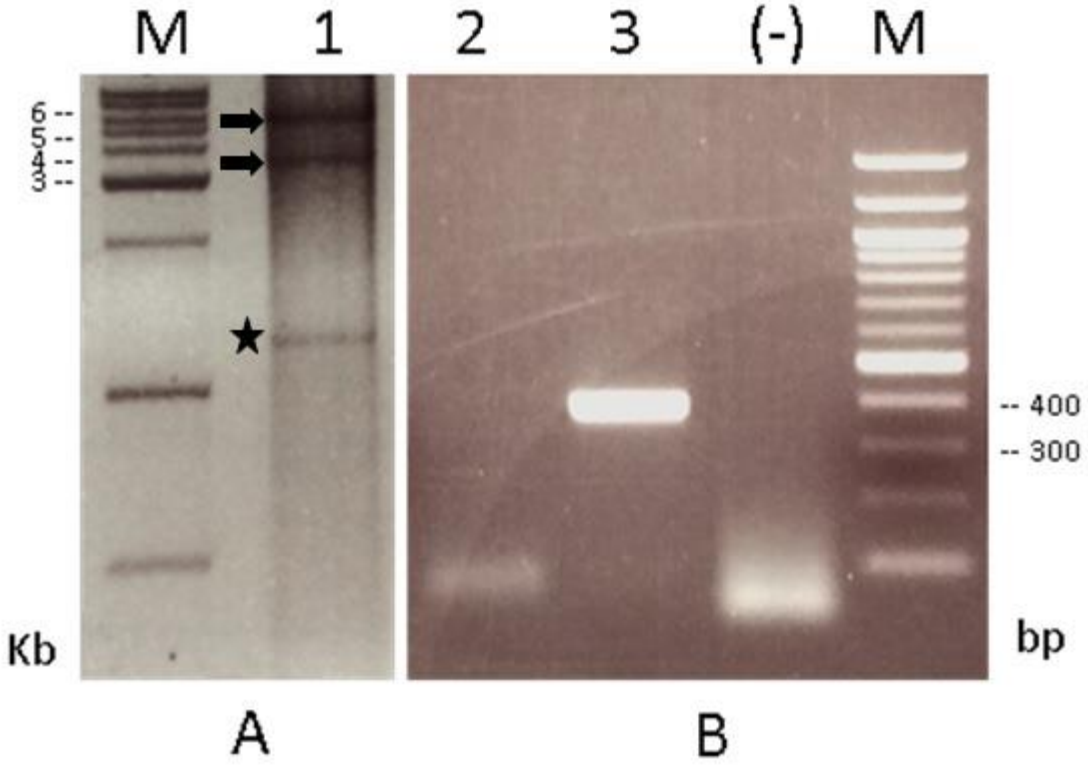

Figure 2.3. A, Restriction profile of *Blueberry fruit drop-associated virus* (BFDaV). Lane 1, The entire genome of BFDaV, amplified with rolling circle amplification and digested with *Eco*RI enzyme, arrows indicate restriction fragments (6.2 kb and 3.6 kb) belonging to BFDaV; the star indicates blueberry chloroplast sequence; lane M, 1-kb ladder. **B,** Amplicons of BFDaV from blueberry 'Bluecrop', using detection primers. Lane 2, Asymptomatic plant; lane 3, Symptomatic plant; lane (-), water control. Expected amplicon size 395 bp; lane M, 100-bp ladder.

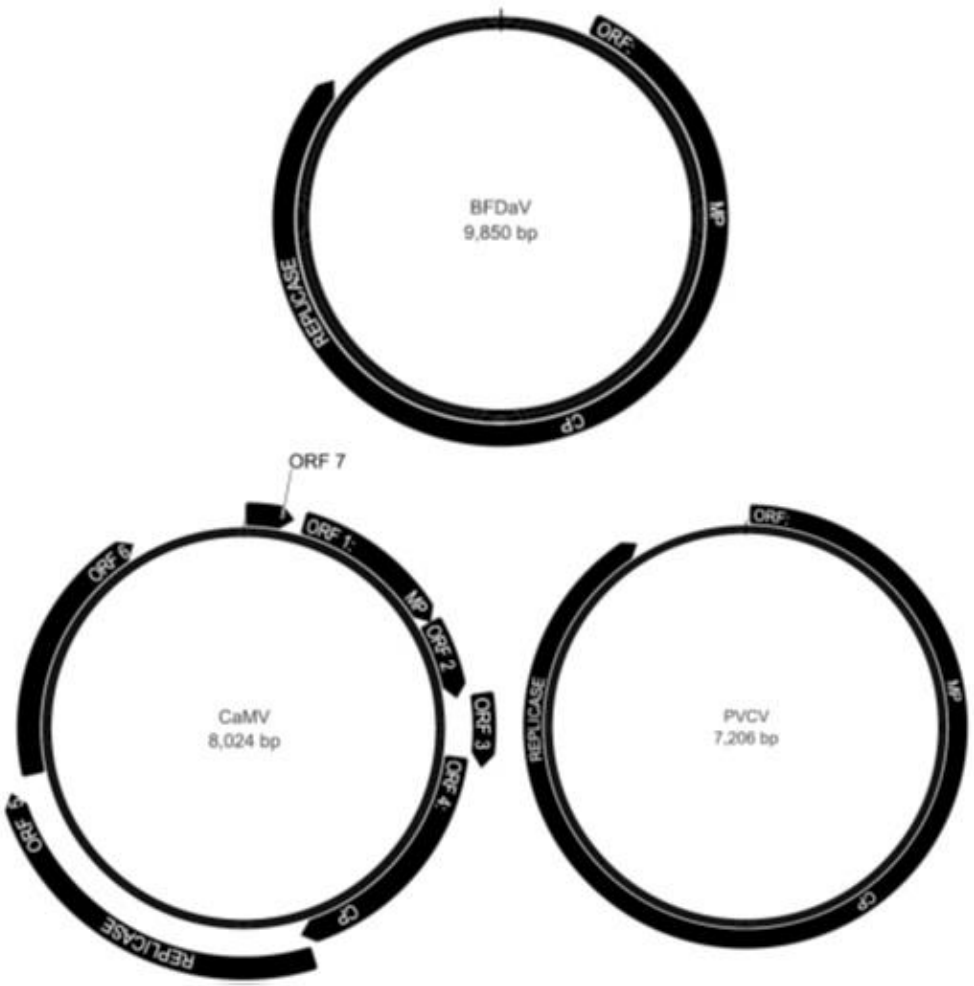

Figure 2.4. Genomic organization of *Blueberry fruit drop-associated virus* (BFDaV), *Cauliflower mosaic virus* (CaMV, genus *Caulimovirus*) and *Petunia vein clearing virus* (PVCV, genus *Petuvirus*). ORF, open reading frame; MP, movement protein; and CP, coat protein.

Chapter 3:

A variant of *Rubus yellow net virus* with altered genomic organization

Alfredo Diaz-Lara, Nola J. Mosier, Karen E. Keller, and Robert R. Martin

Virus Genes
233 Spring St.
New York, NY 10013
United States of America
50: 104-110

Abstract

Rubus yellow net virus (RYNV) is a member of the genus *Badnavirus* (family *Caulimoviridae*). RYNV infects *Rubus* species causing chlorosis of the tissue along the leaf veins, giving an unevenly distributed netted symptom in some cultivars of red and black raspberry. Recently, a strain of RYNV was sequenced from a *Rubus idaeus* plant in Alberta, Canada, exhibiting such symptoms. The viral genome contained seven open reading frames (ORFs) with five of them in the sense-strand, including a large polyprotein. Here we describe a graft-transmissible strain of RYNV from Europe infecting cultivar 'Baumforth's Seedling A' (named RYNV-BS), which was sequenced using rolling circle amplification, enzymatic digestion, cloning and primer walking, and it was resequenced at a 5X coverage. This sequence was then compared with the RYNV-Ca genome and significant differences were observed. Genomic analysis identified differences in the arrangement of coding regions, promoter elements and presence of motifs. The genomic organization of RYNV-BS consisted of five ORFs (four ORFs in the sense-strand and one ORF in the antisense-strand). ORFs 1, 2 and 3 showed a high degree of homology to RYNV-Ca, while ORFs 4 and 6 of RYNV-BS were quite distinct. Also, the predicted ORFs 5 and 7 in the RYNV-Ca were absent in the RYNV-BS sequence. These differences may account for the lack of aphid transmissibility of RYNV-BS.

Introduction

Rubus yellow net virus (RYNV) infects *Rubus* species and cultivars in North America and Europe. It infected all red raspberry cultivars and most blackberry and hybrid berry cultivars (Stace-Smith and Jones, 1987). Leaves of infected plants on some genotypes

develop a netlike chlorosis of the tissue along the veins, giving the plant a pale green appearance with some of the leaves being slightly cupped downward, but there is no distortion or stunting, while many genotypes of *Rubus* remain symptomless (Stace-Smith, 1955). In North America, RYNV in mixed infections with *Black raspberry necrosis virus* (BRNV) causes raspberry mosaic disease (RMD), a virus disease complex that produces serious decline in plant vigor and productivity (Stace-Smith and Jones, 1987). The impact on yield in red raspberry is minimal in single infections but can be 30-75% in mixed infections with BRNV in the first cropping year with losses 0-15% in subsequent years (Stace-Smith and Jones, 1987). In Europe, RYNV together with BRNV and *Raspberry leaf mottle virus* have been reported to cause RMD (McGavin and MacFarlane, 2010).

RYNV is a member of the genus *Badnavirus* (family *Caulimoviridae*) (Jones et al., 2002; Kalischuk et al., 2013), which are pararetroviruses. Pararetroviruses are circular double-stranded DNA viruses that encode for a reverse transcriptase required for their replication. Badnaviruses primarily infect host plants in tropical and sub-tropical regions with a few examples from regions with temperate climates. Most badnaviruses are transmitted by mealybugs (Hull and Covey, 1995), however there are exceptions; RYNV is spread by the large raspberry aphid, *Amphorophora idaei* Börner in Europe and *A. agathonica* Hottes in North America, probably in a semi-persistent manner in both locations (Stace-Smith and Jones, 1987). These vectors are common in commercial *Rubus* fields and are able to transmit other viruses involved in the RMD complex (Quito-Avila et al., 2012). Jones et al. (1974) suggested that small bacilliform (80-150 nm x 25-30 nm) virus-like particles observed in plants with RMD may represent those of RYNV, which resemble the particle morphology of badnaviruses. In 2002, a highly conserved portion of the virus genome was

sequenced using degenerate badnavirus-specific primers (Jones et al., 2002), confirming RYNV as a badnavirus.

Recently, RYNV was cloned and sequenced from one red raspberry (*Rubus idaeus* L.) plant in Lethbridge, Alberta, Canada that exhibited symptoms characteristic of this virus (Kalischuk et al., 2013). The viral genome consisted of 7,932 base pairs (bp) and contained seven putative open reading frames (ORFs), with five of them located on the sense-strand, including a large polyprotein (ORF 3) (Kalischuk et al., 2013), which is the main feature of badnaviruses. The polyprotein is cleaved post-translationally by the viral-encoded aspartic protease to produce a movement protein, coat protein, and replicase comprised of a reverse transcriptase and ribonuclease H (Laney et al., 2012; Sether et al., 2012). All badnaviruses, including RYNV, share three ORFs, which have approximately the same size and location within the genome (Kalischuk et al., 2013; Lockhart, 1990; Medberry et al., 1990). An uncommon characteristic in the genome of the RYNV from Canada (RYNV-Ca) when compared to the genomes of other badnaviruses was the presence of four additional ORFs; two small ORFs in the sense-strand and two ORFs in the antisense-strand (Kalischuk et al., 2013). RYNV-Ca is the only complete RYNV sequence in GenBank, along with two partial sequences, including an isolate from United Kingdom.

In this work, we describe a graft-transmissible strain of RYNV that likely originated in Europe (RYNV-BS), because it was isolated from a plant of ‘Baumforth’s Seedling A’ that came from the U.K. in 1997.

Materials and methods

Plant material. RYNV was isolated from a clone of red raspberry cv. 'Baumforth's Seedling A' (PI number RUB 9020), which is in the National Clonal Germplasm Repository of the USDA (Corvallis, Ore.), the plant originally came from England in 1997. This *Rubus* plant expresses symptoms characteristic of RMD.

Rolling circle amplification, enzymatic digestion, cloning, primer walking, and resequencing. Total DNA was extracted from leaf tissue as described by Lockart (1990), with minor modifications. These modifications included a proteinase K treatment, followed nucleic acid extractions, first with phenol-chloroform-isoamyl alcohol (25:24:1) and then with chloroform-isoamyl alcohol (24:1).

Oligonucleotides designed from previously determined partial sequence of RYNV (GenBank Accession No. AF468454) (Jones et al., 2002) were employed to assess the DNA extraction by polymerase chain reaction (PCR), confirming the presence of RYNV nucleic acids. The RYNV-BS genome was amplified using the Illustra TempliPhi 500 Amplification Kit (GE Healthcare, Buckinghamshire, UK), following the manufacturer's instructions (Dean et al., 2001; Lizardi et al., 1998). Then, the amplified product was digested with *Eco*RI (New England Biolabs, Ipswich, MA, USA) based on previous observation during experimental work for production of polymorphic restriction profiles using several restriction endonucleases (data not shown). The digestion resulted in three restriction fragments, which were ligated into *Eco*RI digested pBluescript II KS(+) using Quick T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions, and later transformed into One Shot Top 10 chemically

competent cells of *Escherichia coli* (Invitrogen Corp., Carlsbad, CA, USA). Finally, plasmids were extracted using the Zyppi Plasmid Miniprep Kit (Zymo Research, Irvine, CA, USA).

Plasmids were sequenced on an ABI 3730XL DNA analyzer by Macrogen (Seoul, Korea) using M13 universal primers, followed by internal primers (Supplemental Table S 3.1) designed to sequence by primer walking. Additionally, the junctions of the three fragments were confirmed using specific sequencing oligonucleotides (Supplemental Table S 3.1) positioned near the ends of each fragment to amplify fragments that included the *Eco*RI restriction sites and the resulting amplicons were sequenced. All sequencing reads were assembled by Geneious v4.8 (Kearse et al., 2012) to produce a preliminary sequence of RYNV-BS, this sequence was used to design primers used to resequence the virus genome. In order to confirm the sequence and correct any residual errors, the virus genome was fully resequenced five times using specific primers (forward and reverse, Supplemental Table S 3.1) and a consensus sequence was obtained.

Genomic analysis of RYNV-BS. The nucleotide (nt) and translated amino acids (aa) sequences of RYNV-BS were compared with other sequences in GenBank using BLASTn and BLASTx (Altschul et al., 1990). ORFs were identified using ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/), later these coding regions were analyzed independently using BLASTp and the Conserved domain-search tool at the NCBI site (Gish et al., 1993; Marchler-Bauer and Bryant, 2004).

Molecular mass was calculated using the protein information resource (http://pir.georgetown.edu/pirwww/search/comp_mw.shtml) and plant *cis*-acting

regulatory DNA elements were searched through PLACE (Higo et al., 1999). Furthermore, nuclear export signals (NES) were predicted using NetNES1.1 (http://www.cbs.dtu.dk/services/NetNES/), patterns and profiles were projected using ELM (http://elm.eu.org/search/). The RYNV-BS sequence was compared with the RYNV-Ca genome using BLAST programs. In the cases of ORFs 4 and 6, the sequences were aligned and the pairwise scores were calculated utilizing ClustalW2 (Larkin et al., 2007).

Phylogenetic analysis. Different approaches were employed to infer the phylogeny of RYNV-BS. Initially, sequence similarity within the RT/RNase H-coding region of badnaviruses was used for species demarcation as this is the most conserved coding region in the genus (James et al., 2010). An evolutionary tree was obtained based on the aa sequence alignment of RNase H. Three phylograms were developed using the entire aa sequence of ORFs 1, 2 and 3. These three coding regions were chosen since they are present in all badnaviruses sequenced to date (Xu et al., 2011).

Isolates from different badnaviruses (details of viruses and sequences used in the analysis are in Supplemental Table S 3.2) were selected to create each phylogenetic analyses, which included *Rice tungro bacilliform virus* (RTBV, *Tungrovirus*) as the outgroup taxon. The matrices were generated in Mesquite software (Maddison and Maddison, 2001) and aligned using MAFFT (Katoh and Standley, 2013) using the L-INS-I method. These aligned sequences were analyzed in MRBAYES 3.0 (Huelsenbeck et al., 2001) for a Bayesian inference using two runs and four chains of 1 million MCMCMC (Markova Chain Montecarlo Metropolis Coupling) generations with 25% burn-in and a sampling frequency of 1,000 generations. The JTT (Jones–Thornton–Taylor) amino acid model of substitution

was used. The MCMCMC results were summarized in a majority consensus tree, showing all clades with estimated probabilities of 50% or greater. Trees were visualized on FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Aphid transmission experiment. Transmission studies were performed as described by Jones et al. (2002) with minor modifications. Individuals of *A. agathonica* were obtained from young raspberry plants in several commercial fields in Benton County, Ore. in July 2013. Upon collection, aphids from different plants and fields were consolidated to form a new colony. Aphids were reared at 23 ˚C and 14 h of light on healthy raspberry cv. 'Meeker' obtained from virus tested *Rubus* collection maintained at the USDA-ARS HCRU in Corvallis, Ore. (Quito-Avila et al., 2012). Later, adult aphids were transferred to detached leaves of virus-infected red raspberry cv. 'Baumforth's Seedling A', maintained in a Petri dish with a moist paper towel. Also, groups of aphids were transferred to young potted 'Baumforth's Seedling A' plants in an aphid proof cage in a greenhouse. Insects were removed from virus sources and frozen (-80 ˚C) or transferred to healthy 'Meeker' plants immediately. Different acquisition access feeding periods (AAFP) and transmission access feeding periods (TAFP) were assigned to groups of five aphids for each assay. After completing TAFPs, aphids were frozen as well. All frozen samples were analyzed by PCR to determine if insects acquired virus.

Results

Virus characterization. The viral genome of RYNV-BS consists of 7,836 nts (GenBank accession no. KM078034). This isolate shares 82% nt sequence identity with RYNV-Ca in

a query covering of 96% its genome. Surprisingly, the RYNV-BS isolate contained only five ORFs rather than seven reported for RYNV-Ca (Figure 3.1). The first protein with unknown function has 210 aa and molecular mass of 23.9 kDa and shares 90% aa sequence identity to ORF 1 of RYNV-Ca, a coiled-coil region was observed inside this frame. ORF 2 encodes for a polypeptide of 139 aa with a predicted molecular mass of 15.2 kDa, and shares 81% aa sequence identity with the corresponding ortholog of RYNV-Ca. ORF 3 shares 90% aa sequence identity to the corresponding ORF of RYNV-Ca. This polyprotein has a length of 1,976 aa and a molecular mass of 224.7 kDa, comprising the main conserved domains of badnaviruses including: ribonuclease H (Ile_{1706}-Leu_{1833}) and reverse transcriptase (Gly_{1424}-Gly_{1607}), which have long terminal repeats (Xiong and Eickbush, 1990); a pepsin-like aspartate protease (Lys_{1210}-Arg_{1295}); a zinc knuckle (Cys_{868}-Lys_{885}), which is a zinc binding motif from retroviral gag proteins (nucleocapsid) (Zhou et al., 2007) and a ribosomal L25/TL5/CTC N-terminal 5S rRNA binding domain (Thr_{60}-His_{102}) that is related to the movement of the virus inside the plant. In addition, a coiled-coil region was found upstream of the zinc finger domain. ORF 4 encodes for a putative 15.4 kDa protein of 138 aa, which shares 67% aa sequence identity with the putative protein of ORF 4 in RYNV-Ca, this frame contains five NES. RYNV-BS genome contained only one ORF in the antisense strand; this ORF encodes for a putative protein of 69 aa with a molecular mass of 7.3 kDa and six NES. This polypeptide lacks conserved domains, but interestingly has the same degree of similarity (Identities:38/67=57% aa) with putative proteins encoded by ORFs 5 and 6 of RYNV-Ca.

The RYNV-BS contains the negative strand of 3'ACCAUAGUCUCGGUCCAA5'$_{1\text{-}18}$, which is related to a plant $tRNA^{Met}$ and priming site for reverse transcription (Medberry et

al., 1990), similar to that reported for the RYNV-Ca (Kalischuk et al., 2013). This non-coding RNA was selected as the starting location for the circular genome. The intergenic region (849 nts) includes nine GATA-boxes and two TATA-box motifs, promoter elements related to badnavirus replication (Lam and Chua, 1989; Medberry et al., 1992).

Phylogenetic analysis. Phylogenetic analyses using the three ORFs present in all badnaviruses and a conserved domain (RNase H) placed RYNV-BS and RYNV-Ca into the same monophyletic group sharing the most recent common ancestor (Figure 3.2), this inference is supported by 100% of posterior probabilistic in all cases, additionally, *Gooseberry veinbanding associated virus* (GVBAV) was the sister taxon of the RYNV clade in the four phylograms as reported previously (Jones et al., 2002; Kalischuk et al., 2013). Together these results not only confirm the relationship between RYNV-Ca and RYNV-BS but also the existence of divergent isolates of RYNV in the genus *Badnavirus*.

Acquisition and transmission of viral particles by aphids. RYNV was not detectable in aphids following the acquisition access feeding times reported in previous studies (Jones et al., 2002) using either detached leaves or young infected plants. However, when acquisition access feeding periods were extended to 2 weeks to ensure feeding from plant tissue and the aphids immediately frozen when removed from the source plants, PCR testing yielded an amplicon in all cases. Testing of cohorts of aphids that were transferred to healthy plants for 2, 4, 8, 24 or 48 hours and then tested for RYNV produced erratic results with only some of the aphids testing positive for RYNV, suggesting that the virus was not replicating in the aphids. In no case did the aphids transmit RYNV-BS to healthy

raspberry plants in these assays. The test plants did not develop any symptoms and were consistently negative when tested for RYNV by PCR. These data suggests that RYNV-BS is a non-aphid transmissible strain of RYNV, or that the population of *A. agathonica* selected for this study was not capable of transmitting the virus.

Discussion

The molecular characterization of a virus involved in RMD resulted in the identification of a new strain of *Rubus yellow net virus*, which differed significantly from the previously reported RYNV sequence (Table 3.1). RYNV-Ca has seven ORFs while RYNV-BS has five ORFs. Furthermore, when the putative proteins encoded by ORFs 4 and 6 were compared to orthologs encoded by the RYNV-Ca (Figure 3.3), significant differences in molecular mass were observed as well as a low degree of homology. The divergence observed in ORFs 4 and 6 contrasts with the high degree of aa sequence identity in ORFs 1-3 (average 90%).

Interestingly, the predicted coding regions 5 and 7 in the RYNV-Ca were absent in the RYNV-BS genome. A short region with limited homology to ORF 5 was detected in the RYNV-BS sequence, however it lacked a start and a stop codon and thus cannot be considered a real ORF. Similarly, there were short nt sequences with limited homology to ORF 7 from RYNV-Ca, but in the sense-strand rather than the antisense-strand as reported for RYNV-Ca, which suggests possible genetic rearrangement, but again these regions lacked start and stop codons.

A highly conserved portion of the virus genome, previously sequenced in the U.K. (Jones et al., 2002), was aligned with the same conserved region present in RYNV-BS using

BLASTn (Altschul et al., 1990) and a near perfect homology was obtained (Identities:1695/1709=99%, Gaps:3/1709=<1%). In contrast, this partial sequence showed much lower homology with the corresponding sequence in the RYNV-Ca genome (Identities:1158/1460=79%, Gaps:21/1460=1%). This result suggests that RYNV-BS is very close to the isolate that Jones et al. published in 2002 (Jones et al., 2002) and supports the suggestion that the RYNV in the 'Baumforth Seedling A' used in this study was infected prior to it being obtained by the NCGR in Corvallis, Ore. and represents a European isolate of RYNV. Further testing of RYNV in native and commercial *Rubus* spp. in North America and Europe should provide information on the diversity of RYNV and possibility that these two strains are the result of geographic isolation. It is possible that RYNV-Ca isolated from native *Rubus* in Alberta, Canada represents a North American lineage of this virus and the RYNV-BS reported here represents a European lineage.

RYNV-BS was graft-transmissible (data not shown) but all aphid transmission attempts failed (five repetitions). This could be due to several changes in the genomic arrangement of RYNV-BS including deletions that resulted in the loss of aphid transmissibility. In the case of *Cauliflower mosaic virus* (genus *Cauilimovirus*) factors interacting with both virus particles and vector mouthparts, thereby mediating virus aphid transmission; were identified (Schmidt et al., 1994). On the other hand, if the origin of RYNV-BS is indeed in Europe, the lack of transmission of RYNV-BS particles by the large raspberry aphid from North America (*A. agathonica*) may simply be a matter of this aphid being a non-vector. Thus, further studies are necessary to understand the dynamics of the two strains of RYNV in North America.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Dean, F.B., Nelson, J.R., Giesler, T.L., Lasken, R.S., 2001. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. 11, 1095–1099.

Gish, W., States, D.J., 1993. Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

Higo, K., Ugawa, Y., Iwamoto, M., Korenaga, T., 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 27, 297–300.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Hull, R., Covey, S.N., 1995. Retroelements: propagation and adaptation. Virus Genes 11, 105–118.

James, A.P., Geijskes, R.J., Dale, J.L., Harding, R.M., 2010. Development of a novel rolling-circle amplification technique to detect *Banana streak virus* that also discriminates between integrated and episomal virus sequences. Plant Dis. 95, 57–62.

Jones, A.T., McGavin, W.J., Geering, A.D.W., Lockhart, B.E.L., 2002. Identification of *Rubus yellow net virus* as a distinct badnavirus and its detection by PCR in *Rubus* species and in aphids. Ann. Appl. Biol. 141, 1–10.

Jones, A.T., Roberts, I.M., Murant, A.F., 1974. Association of different kinds of bacilliform particle with vein chlorosis and mosaic diseases of raspberry (*Rubus idaeus*). Ann. Appl. Biol. 77, 283–288.

Kalischuk, M.L., Fusaro, A.F., Waterhouse, P.M., Pappu, H.R., Kawchuk, L.M., 2013. Complete genomic sequence of a *Rubus yellow net virus* isolate and detection of genome-wide pararetrovirus-derived small RNAs. Virus Res. 178, 306–313.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Lam, E., Chua, N.H., 1989. ASF-2: a factor that binds to the *Cauliflower mosaic virus* 35S promoter and a conserved GATA motif in Cab promoters. Plant Cell 1, 1147–1156.

Laney, A.G., Hassan, M., Tzanetakis, I.E., 2012. An integrated badnavirus is prevalent in fig germplasm. Phytopathology 102, 1182–1189.

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

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., Ward, D.C., 1998. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat. Genet. 19, 225–232.

Lockhart, B., 1990. Evidence for a double-stranded circular DNA genome in a second group of plant viruses. Phytopathology 80, 127–131.

Maddison, W., Maddison, D., 2001. Mesquite: a modular system for evolutionary analysis. http://mesquiteproject.org

Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32, W327–W331.

McGavin, W. J., MacFarlane, S.A., 2010. Sequence similarities between *Raspberry leaf mottle virus*, *Raspberry leaf spot virus* and the closterovirus *Raspberry mottle virus*. Ann. Appl. Biol. 156, 439–448.

Medberry, S.L., Lockhart, B.E., Olszewski, N.E., 1992. The *Commelina yellow mottle virus* promoter is a strong promoter in vascular and reproductive tissues. Plant Cell Online 4, 185–192.

Medberry, S.L., Lockhart, B.E.L., Olszewski, N.E., 1990. Properties of *Commelina yellow mottle virus*'s complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. Nucleic Acids Res. 18, 5505–5513.

Quito-Avila, D.F., Lightle, D., Lee, J., Martin, R.R., 2012. Transmission biology of *Raspberry latent virus*, the first aphid-borne reovirus. Phytopathology 102, 547–553.

Schmidt, I., Blanc, S., Esperandieu, P., Kuhl, G., Devauchelle, G., Louis, C., Cerutti, M., 1994. Interaction between the aphid transmission factor and virus particles is a part of the molecular mechanism of *Cauliflower mosaic virus* aphid transmission. Proc. Natl. Acad. Sci. 91, 8885–8889.

Sether, D.M., Melzer, M.J., Borth, W.B., Hu, J.S., 2012. *Pineapple bacilliform CO virus*: diversity, detection, distribution, and transmission. Plant Dis. 96, 1798–1804.

Stace-Smith, R., 1955. Studies on rubus virus diseases in British Columbia: I. Rubus yellow-net. Can. J. Bot. 33, 269–274.

Stace-Smith, R., Jones, A.T., 1987. Rubus yellow net. Agric. Handb. - U. S. Dep. Agric. Comb. For. Pest Res. Dev. Program USA.

Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353–3362.

Xu, D., Mock, R., Kinard, G., Li, R., 2011. Molecular analysis of the complete genomic sequences of four isolates of *Gooseberry vein banding associated virus*. Virus Genes 43, 130–137.

Zhou, J., Bean, R.L., Vogt, V.M., Summers, M., 2007. Solution structure of the *Rous sarcoma virus* nucleocapsid protein: μΨ RNA packaging signal complex. J. Mol. Biol. 365, 453–467.

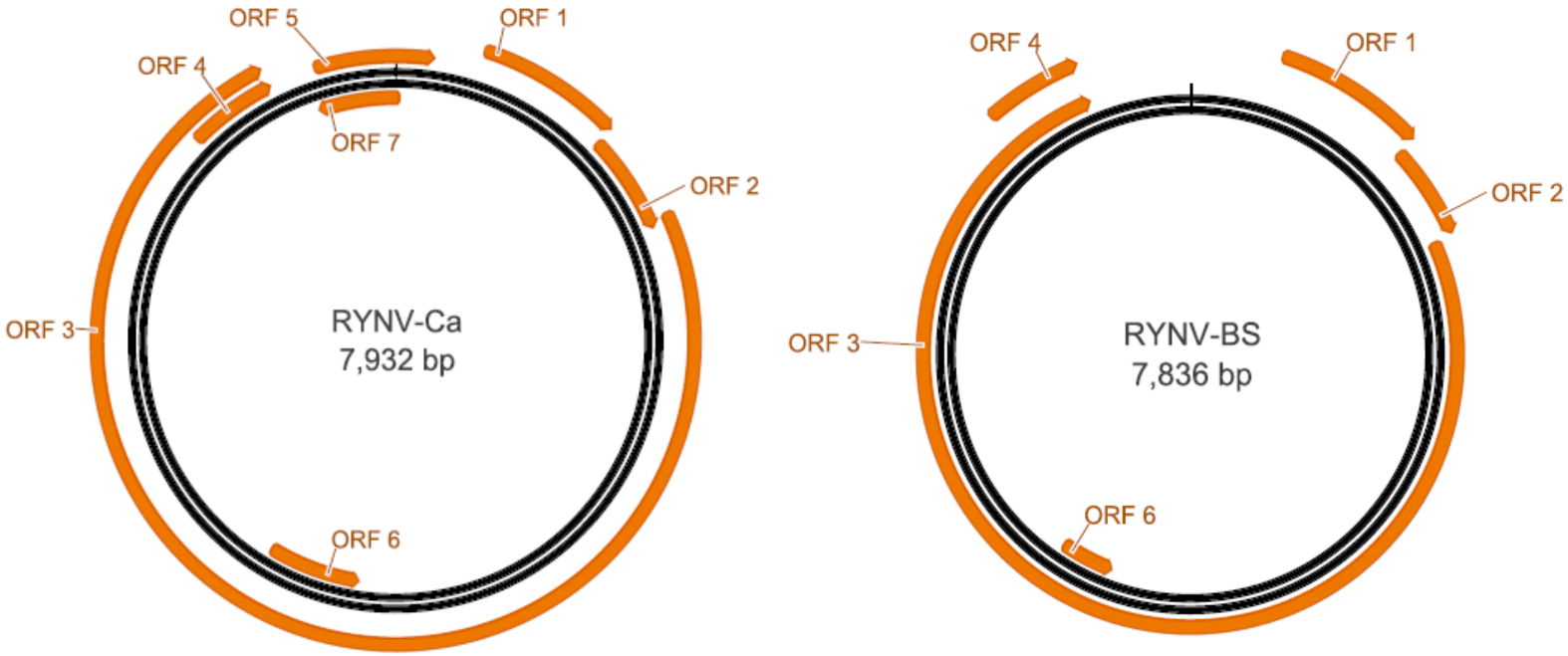

Figure 3.1. Representation of *Rubus yellow net virus* isolate "Canada" (RYNV-Ca) and *Rubus yellow net virus* isolate "Baumforth's Seedling A" (RYNV-BS) genomes. Orange arrows denote the coding regions; ORFs 1-5 located on the sense-strand and ORFs 6 and 7 located in the antisense-strand.

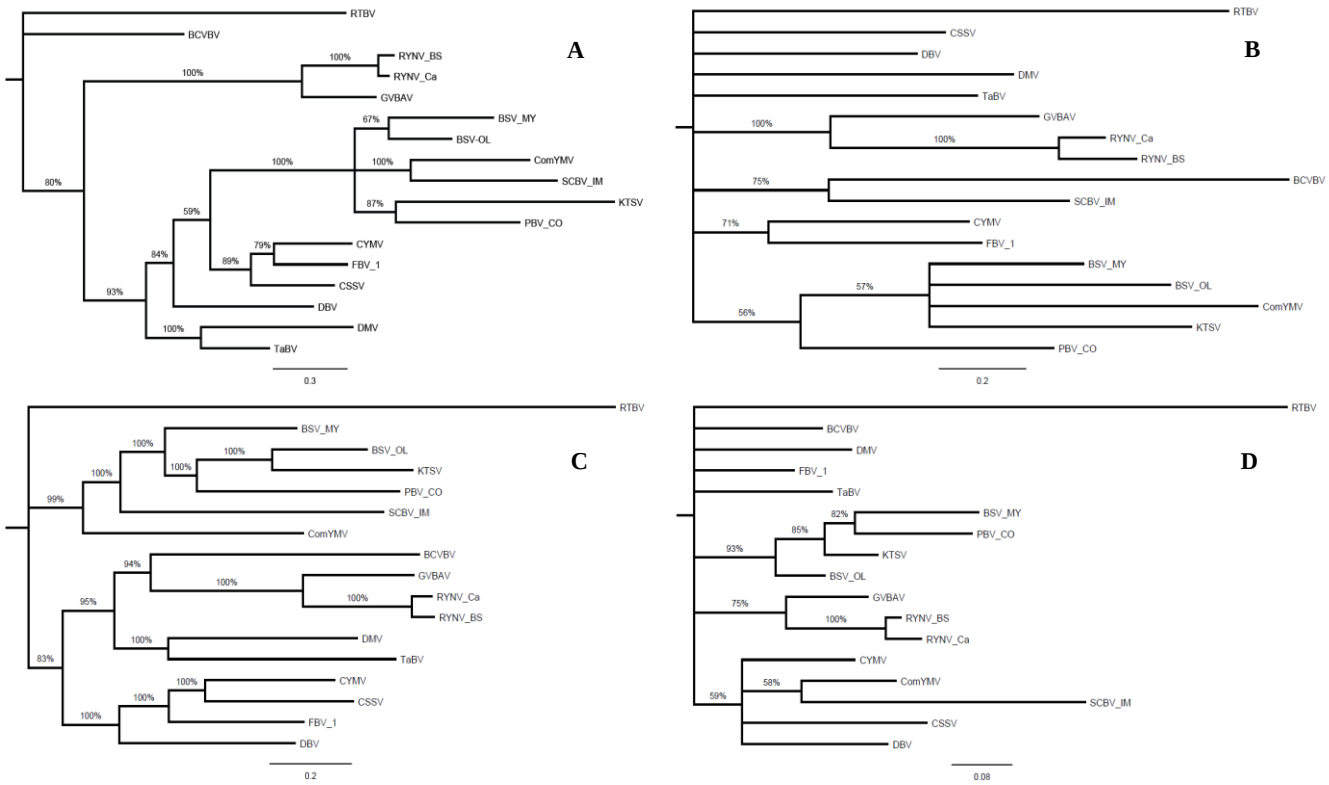

Figure 3.2. Phylogenetic analysis of the genus *Badnavirus*. Bayesian inference using two runs and four chains of 1 million MCMCMC generations with 25% burn-in and a sampling frequency of 1,000 generations. **A**, Relationships determined using the aa sequence of ORF 1 present in badnaviruses. **B**, Relationships determined using the aa sequence of ORF 2. **C**, Relationships determined using the aa sequence of ORF 3. **D**, Evolutionary inference using the aa sequence of the RNase H conserved domain. The MCMCMC results were summarized in a 50% majority consensus tree. Numbers above branches are posterior probabilities recovered by the Bayesian analysis. Details of the accession used in the analysis are in the Supplemental Table S 3.2. *Rice tungro bacilliform virus* (RTVB) was used as outgroup. Horizontal branch length is proportional to genetic distance; the scale bars represent changes per site. Values less than 50 are not shown.

Figure 3.3. Alignment of two different ORFs present in RYNV-BS and RYNV-Ca genomes. **A**, ORF 4. **B**, ORF 6. An asterisk denotes an exact match, a double dot indicates conservation between groups of strongly similar properties and a single dots denotes conservation between groups of weakly similar properties. Sequences were aligned using ClustalW2.

Table 3.1. Comparison of coding regions present in *Rubus yellow net virus* isolates "Canada" and "Baumforth's Seedling A", determining positions and lengths of ORFs in both genomes.

ORF	Starting nucleotide		Ending nucleotide		Length nts		Length aa		BLASTp results	
	Ca	BS	Ca	BS	Ca	BS	Ca	BS	Identity	Similarity
1 (+)	386	386	1018 (TGA)	1018 (TGA)	633	633	210	210	189/210(90%)	198/210(94%)
2 (+)	1015	1015	1476 (TGA)	1434 (TGA)	462	420	153	139	103/127(81%)	115/127(90%)
3 (+)	1433	1431	7348 (TGA)	7361 (TGA)	5916	5931	1971	1976	1777/1978(90%)	1863/1978(94%)
4 (+)	6940	6953	7356 (TAG)	7369 (TAG)	417	417	138	138	93/138(67%)	108/138(78%)
5 (+)	7654	*	180 (TGA)	*	459	*	152	*		
6 (-)	3330	4365	3767 (TGA)	4574 (TGA)	438	210	145	69	38/67(57%)	45/67(67%)
7 (-)	7906	*	405 (TAA)	*	432	*	143	*		

Sequences of both RYNV isolates were aligned and compared using BLASTp at the amino acids level. Data from RYNV-Canada was obtained from Kalischuk et al., 2013.

ORF, open reading frame; Ca, RYNV-Canada; BS, RYNV-Baumforth's Seedling A; nts, nucleotides; aa, amino acids; BLASTp, protein basic local alignment search tool; +, sense-strand; -, antisense-strand; *, missing.

Chapter 4:

Evidence of *Rubus yellow net virus* integration into the red raspberry genome

Alfredo Diaz-Lara, Nola J. Mosier, Karen E. Keller, and Robert R. Martin

Abstract

Rubus yellow net virus (RYNV) infects *Rubus* spp. in Europe and North America, causing a severe decline when present in mixed infections and a vein chlorosis in indicator plants. RYNV belongs to the family *Caulimoviridae* (plant pararetroviruses), which can exist as infectious particles (exogenous) or incorporated elements (endogenous). Integrated pararetroviruses do not cause infection, however they can lead to false positives in PCR assays. Graft transmission onto indicator plants (*R. occidentalis*) has been the standard test method for RYNV. Recently, it was noticed that some plants that were PCR positive for the virus did not induce symptoms in bioassays, suggesting an incorporated element or a symptomless strain of the virus. In this study, different molecular techniques were employed to differentiate between endogenous and exogenous RYNV sequences. RT-PCR using RYNV specific oligonucleotides after DNase treatment generated positive results for the virus in graft transmissible isolates only. To confirm these results, rolling circle amplification (RCA) on DNA preparations from the same samples resulted in amplicons identified as RYNV. In Southern blot hybridization, using *Rubus* genomic DNA digested with *Bam*HI (one recognition site in the RYNV genome) and a unique probe for RYNV, multiple signals, differing in size from digested RYNV genomic DNA, were detected. Next-generation sequencing was used to identify the RYNV-like sequences present in the plant's DNA. These results demonstrate the incorporation of RYNV into the red raspberry genome and highlight the necessity to recognize this phenomenon in future *Rubus* quarantine and certification programs.

Introduction

Raspberries are classified in the genus *Rubus* (family *Rosaceae*), which includes hundreds of species and hybrid berry cultivars (Martin et al., 2012). Approximately 19,750 acres of red raspberry (*R. idaeus*) and black raspberry (*R. occidentalis*) were harvested in the USA in 2014 (https://www.nass.usda.gov/Statistics_by_Subject/), making this a horticulturally important crop. Raspberry is susceptible to numerous diseases, including those triggered by viruses; there are reports of more than 30 viruses or virus-like diseases affecting *Rubus* spp. (for a review see Martin et al., 2012).

In North America and Europe, *Rubus* spp. are hosts of *Rubus yellow net virus* (RYNV), whose symptoms on plants range from netlike chlorosis of the tissue along the veins to asymptomatic appearance (Stace-Smith, 1955; Stace-Smith and Jones, 1987). RYNV in mixed infections with other aphid-borne viruses (*Raspberry leaf mottle virus* and *Black raspberry necrosis virus*) causes raspberry mosaic disease (RMD), a disease complex that severely affects plant vigor and yield (Jones et al., 2002). RYNV produces bacilliform particles (virions) containing a circular double-stranded DNA (dsDNA) genome that is 7.9 kb in size and encodes a large polyprotein (Jones et al., 2002; Kalischuk et al., 2013; Diaz-Lara et al., 2015), features typical of the genus *Badnavirus*, family *Caulimoviridae*.

Caulimoviruses or plant pararetroviruses replicate via cellular RNA intermediate (reverse transcription step) without integrating into the host genome, as in the case of animal retroviruses (Lyttle et al., 2011). Despite their non-integrative replication cycle, there is enough evidence that pararetroviruses can exist as exogenous and endogenous forms (Harper et al., 2002). Episomal (exogenous) viruses produce free virions capable of causing infection in the host, on the other hand, endogenous pararetroviruses (EPRVs) are

incorporated in the host genome and are common in the plant kingdom (Staginnus and Richert-Pöggeler, 2006).

Over the past few years, several *Rubus* plants that indexed negative for RYNV in aphid or graft transmission assays onto the biological indicator produced positive results in PCR-based tests, leading to rejection of plants by quarantine personnel in importing countries. In this work, we investigated the possibility that RYNV is integrated into the red raspberry genome using grafting, PCR, reverse transcription-PCR (RT-PCR), rolling circle amplification (RCA) and Southern blotting. Once the presence of endogenous RYNV was confirmed, the next stage was to characterize the RYNV-like sequences, in order to determine diversity of insertions via next-generation sequencing (NGS).

Materials and Methods

Plant material and preliminary tests. Diverse red raspberry cultivars were obtained from the National Clonal Germplasm Repository and the Horticultural Crops Research Unit (HCRU) of the USDA-ARS in Corvallis, Oregon. These plants were PCR-positive for RYNV using the detection primers RYNV6F (5'-CGTGATAACGGCTTGGTTTT-3') and RYNV6R (5'-CGTAAGCGCAGATTTCTTCC-3') that target the viral reverse transcriptase and ribonuclease H domains respectively and amplify a 463-bp amplicon (red raspberry cultivars are listed in Table 4.1). PCR amplification was performed using *Taq* polymerase (GenScript, Piscataway, NJ) and a program consisting of 3 min at 95 °C; 40 cycles of 40 s at 95 °C, 40 s at 56 °C, 40 s at 72 °C; and a final elongation step of 10 min at 72 °C. Additionally, all the amplicons were sequenced to confirm RYNV origin.

Two of the cultivars that were positive for RYNV in the above PCR test (‘Cascade Bounty’ and ‘Cascade Harvest’) were exposed to heat therapy and meristem-tip culture protocols for the elimination of the virus (protocols available upon request). Eight months later, the produced plantlets were tested by PCR and continued generating positive results.

Graft transmission. All PCR assessed *Rubus* cultivars were employed as donor sources for graft transmission onto indicator plants (*R. occidentalis* cv. ‘Munger’), which were obtained from a commercial nursery in Oregon with a disease-free certification program. In addition, all the ‘Munger’ plants were tested with the RYNV-primers to confirm the virus-free status before the start of the experiment. Three repetitions (‘Munger’ plants) per donor plant were used, and each indicator plant was grafted twice (two grafts). Lastly, the grafted plants were maintained in a greenhouse at the USDA-ARS HCRU at 25 °C with 16 h daylight; and the plants were evaluated for RYNV symptoms three, six and eight weeks post-grafting as well as tested for RYNV (PCR) at the end of the observation period.

Nucleic acids extraction. Different extraction protocols were used to assure the correct performance of the upcoming molecular tests. In the case of the RT-PCR, the protocol described previously (Quito-Avila and Martin, 2012) was employed to obtain total RNA from symptomatic and asymptomatic red raspberry leaves. For RCA and Southern blot analysis, total DNA was extracted from plant material as previously described by Diaz-Lara et al. (2015). Finally, the FastDNA SPIN Kit and the FastPrep Instrument (MP Biomedicals, Santa Ana, CA) were utilized, according to manufacturer’s recommendations to isolate high-quality genomic DNA for downstream NGS.

Molecular techniques: RT-PCR, RCA and Southern blotting. From the preliminary tests and graft transmission experiments, seven cultivars (‘Caroline’, ‘Comox’, ‘Glen Clova’, ‘Meeker’, ‘Willamette’, ‘Cascade Bounty’ and ‘Baumforth’s Seedling A’) were selected and analyzed by RT-PCR, RCA and Southern blotting to determine whether RYNV is actively replicating or is incorporated in the genome of the host plant. RNA extracts from the samples were digested with 1 µl of RQ1 RNase-Free DNase (Promega, Madison, WI) following the manufacturer’s recommendations prior to RT-PCR. 2.5 µl of DNase-digested product was used as template for the reverse transcription (RT) using Maxima Reverse Transcriptase (Thermo Scientific, Hudson, NH) and the RT reaction was incubated 60 min at 50 °C, then 5 min at 85 °C, followed by PCR as described previously.

1 µg of sample DNA was treated with 10 units of Exonuclease V (New England Biolabs, Ipswich, MA) at 37 °C for 30 min before the RCA, in order to eliminate the linear DNA (potential interference with the RCA reaction). The whole circular genome of RYNV was amplified via RCA using the Illustra TempliPhi 500 Amplification Kit (GE Healthcare, Buckinghamshire, UK) following the commercial protocol. The RCA product was digested with *Bam*HI (New England Biolabs, Ipswich, MA) and then size analyzed by electrophoresis (1.5% agarose gel). The RYNV genome contains only one *Bam*HI recognition site.

The last method used to detect endogenous RYNV sequences was Southern blot hybridization. 10 µg of total DNA from red raspberry samples was treated overnight with *Bam*HI restriction enzyme and later electrophoresed (1% agarose gel) for 2.5 h at 50 V. Immediately, the DNA was transferred onto positively charged Nylon Membrane (Roche Diagnostics, Indianapolis, IN) and fixed by cross linking with UV-light. The DIG-High

Prime DNA Labeling and Detection Starter Kit II (Roche Diagnostics, Indianapolis, IN) was utilized for the DIG-DNA labeling, hybridization and immunological detection according to the manufacturer's instructions. The DNA probe was the 463-bp amplicon obtained with the RYNV6F/RYNV6R oligonucleotide primers.

Next-generation sequencing. Aiming to characterize the RYNV-like sequences present in the red raspberry genome, genomic DNA from the cultivar 'Meeker' was sequenced using the Illumina HiSeq 3000 Platform (Illumina, San Diego, CA), using a paired-end (2x150 bp) configuration. Briefly, an entire flow cell lane was loaded with the single sample, producing around 300 million paired reads. To obtain a draft version of the 'Meeker' genome, a *de novo* assembly of a sub-set of 30% of the total number of reads was generated, using VelvetOptimiser (Zerbino, 2002). VelvetOptimiser optimized the use of the Velvet assembler (Zerbino and Birney, 2008) using different *k*-mer values (21, 25, 27, 29 and 31). After assembly, contigs were expanded; to perform the expansion, the resultant contigs from Velvet (*k*=31, best *k*-mer value) and the total number of sequencing reads were combined.

Extended contigs were compared with two available RYNV genomes in the GenBank (Accession no. NC_026238 and KF241951), using BLASTn (Altschul et al., 1990). BLASTn search was executed using the default parameters, except for the word size=20. Finally, Geneious v7.1.9 (Kearse et al., 2012) was utilized to map contigs against the RYNV genome (KF241951).

Results

Graft transmission. Indicator plants grafted with ‘Baumforth’s Seedling A’, ‘Glen Esk’ and ‘Glen Moy’ started to show RYNV-symptoms (vein chlorosis) 6 weeks post-grafting, later, they were confirmed positive by PCR. On the other hand, the rest of raspberry cultivars failed to induce symptoms after graft transmission (*Rubus* cultivars and assay results are listed in Table 4.1).

RT-PCR. Except for the cultivar ‘Baumforth’s Seedling A’, the rest of samples (‘Caroline’, ‘Comox’, ‘Glen Clova’, ‘Meeker’, ‘Willamette’ and ‘Cascade Bounty’) were negative for RYNV after the RT-PCR test with the additional DNase-treatment (Figure 4.1). In contrast, all the samples were positive for the virus via RT-PCR when samples were not digested with DNase prior to RT-PCR, which suggests that the positive amplification originated from RYNV DNA, not RNA. These results suggested that a version of RYNV with replication capacity, such that RNA is produced *in vivo*, is present in the ‘Baumforth’s Seedling A’ sample and not in the other cultivars tested.

RCA. Supporting the RT-PCR results, RCA yielded an amplicon from the sample ‘Baumforth’s Seedling A’ but not from the other *Rubus* cultivars in the study. A DNA band of nearly 8 kb in size was observed after the RCA product was treated with *Bam*HI (Figure 4.2). The amplicon matched the predicted size and when sequenced was most closely related to RYNV.

Southern blotting. With the purpose of confirming the integration by RYNV, four red raspberry cultivars ('Comox', 'Glen Clova', 'Meeker' and 'Cascade Bounty') apparently free of viral particles were analyzed through Southern blotting. At least four different bands (around 1.3, 2.8, 7 and 8 kb in size) were observed (Figure 4.3). Also, a signal from undigested DNA occurred at high-molecular-weight (great than 10 kb). The pattern of hybridization signals was similar among cultivars suggesting the inserted RYNV sequences are closely linked.

Next-generation sequencing. The extension process resulted in 1,307,705 contigs (ranging from 61 to 2,250 nucleotides), all with above 100x coverage. Later, 93 contigs were found to share 82-100% nucleotide sequence identity with the complete genomic sequences of RYNV (Supplemental Table S 4.1); on the other hand, some contigs did not have full query coverage, which presumably originated from virus-plant junctions. 4 out of 93 contigs were more closely related to a European strain of the virus (NC_026238), while the rest matched more closely to a North American isolate (KF241951). Based on the previous result, RYNV-like contigs were mapped against the KF241951 sequence, observing several gaps along the viral genome (Figure 4.4); similar result was obtained when all the expanded contigs (1,307,705) were included (data not shown).

Discussion

The work presented clearly demonstrates that RYNV has integrated into the genome of red raspberry on at least one occasion. Using different molecular approaches (RT-PCR, RCA and Southern blotting) and grafting, we could differentiate between exogenous and

endogenous versions of RYNV in several red raspberry cultivars. Following the nomenclature proposed by Geering et al. (2010), the EPRV identified in red raspberry should be named endogenous-*Rubus yellow net virus* (eRYNV).

There are several examples of incorporated pararetrovirus sequences with nearly perfect identity to infectious viruses reported in flowering plants previously (Harper et al., 2002; Hansen et al., 2005). For example, EPRVs have been reported from banana (Harper et al., 1999), tobacco (Gregor et al., 2004; Jakowitsch et al., 1999), petunia (Richert-Poggeler et al., 2003; Richert-Pöggeler and Shepherd, 1997), rice (Kunii et al., 2004), fig (Laney et al., 2012) and recently from yam (Umber et al., 2014). The majority of integrated virus-like sequences were rearranged in the host DNA, suggesting illegitimate recombination (Harper et al., 2002). Such recombination may have occurred when the episomal virus was replicating in the host (Kunii et al., 2004).

Graft indexing is a popular technique for virus detection, especially in the case of unknown viruses; however, in this study we used grafting for a different reason, to determine whether red raspberry cultivars hosted a form of RYNV with infectious capacity or not. The aim of including a DNase digestion before the RT-PCR was to remove any RYNV-like sequence present in the genomic DNA of red raspberry, thus targeting only RNA transcripts of exogenous viruses. Similar methodology was employed in an initial study reporting a new badnavirus in blackberry with integration capacity (Shahid et al., 2016). RYNV belongs to the group *Caulimoviridae*, and the RCA amplifies the circular RYNV genome and not the endogenous virus since it is not a small circular DNA. Southern blot analysis of total DNA extracts from red raspberry cultivars using a probe specific for RYNV produced unexpected hybridization signals including at high-molecular-weight, signs of an

integration occurrence (Harper et al., 1999; Laney et al., 2012). The restriction endonuclease *Bam*HI has one recognition site in the RYNV genome, so it should produce a single blot of 7.9 kb after the hybridization; however, the DNA probe bound in additional locations showing that the RYNV is flanked by other DNA containing *Bam*HI restriction sites. Finally, according to the patterns observed after the Southern blotting, the integrated sequences are consistent among plants (Figure 4.3), which suggests that incorporation has been a single event that has been transmitted from generation to generation.

The assays to differentiate between exogenous and endogenous RYNVs can avoid unnecessary costs of false positives at virus-elimination stages and more importantly restriction on plant movement based on inaccurate virus diagnosis. Quarantine restrictions need to be developed to reflect this phenomenon (integration). In the case of eRYNV we did not find any evidence of the excision of the incorporated sequences that led to virus infection. The process of integration into and excision from host genomes of pararetroviruses should be studied. If there are excision signals that could be identified, then the risk of an EPRV in a plant genome could be assessed in terms of whether it might excise and cause disease.

Further studies are underway to determine the impact of endogenous RYNV sequences on disease (RMD) development. If incorporated sequences do not enhance symptoms of other viruses, while RYNV virions do enhance symptoms of other viruses, this would also support the idea that plant material with endogenous RYNV should not be regulated in the same manner as RYNV-infected plants.

Based on the NGS result, most of the viral genome is present in the red raspberry DNA, although, with missing regions (Figure 4.4). Because of the unavailability of the red

raspberry genome, we attempted to identify the specific integration sites using the black raspberry genome (VanBuren et al., 2016) as a reference. This attempt was unsuccessful, hypothesizing that black raspberry lacks incorporated sequences of RYNV. Discerning the location and structure of the integration sites is important, as some EPRVs have been reported to activate (assembly of virions and induce infection) after exposure to abiotic stress or tissue culture (Ndowora et al., 1999; Richert-Poggeler et al., 2003); the last such case described being *Banana streak virus*, sharing the same genus as RYNV. If endogenous RYNV sequences are shown to excise under various types of stress conditions and develop infectious virus particles or are shown to negatively impact plants co-infected with other viruses, then integrated RYNV sequences should be treated in much the same way as RYNV particles in plants.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Diaz-Lara, A., Mosier, N.J., Keller, K.E., Martin, R.R., 2015. A variant of *Rubus yellow net virus* with altered genomic organization. Virus Genes 50, 104–110.

Geering, A.D.W., Scharaschkin, T., Teycheney, P.-Y., 2010. The classification and nomenclature of endogenous viruses of the family *Caulimoviridae*. Arch. Virol. 155, 123–131.

Gregor, W., Mette, M.F., Staginnus, C., Matzke, M.A., Matzke, A.J.M., 2004. A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. Plant Physiol. 134, 1191–1199.

Hansen, C.N., Harper, G., Heslop-Harrison, J.S., 2005. Characterisation of pararetrovirus-like sequences in the genome of potato (*Solanum tuberosum*). Cytogenet. Genome Res. 110, 559–565.

Harper, G., Hull, R., Lockhart, B., Olszewski, N., 2002. Viral sequences integrated into plant genomes. Annu. Rev. Phytopathol. 40, 119–136.

Harper, G., Osuji, J.O., Heslop-Harrison, J.S. (Pat), Hull, R., 1999. Integration of *Banana streak Badnavirus* into the *Musa* genome: molecular and cytogenetic evidence. Virology 255, 207–213.

Jakowitsch, J., Mette, M.F., Winden, J. Van Der, Matzke, M.A., Matzke, A.J.M., 1999. Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. Proc. Natl. Acad. Sci. 96, 13241–13246.

Jones, A.T., McGavin, W.J., Geering, A.D.W., Lockhart, B.E.L., 2002. Identification of *Rubus yellow net virus* as a distinct badnavirus and its detection by PCR in *Rubus* species and in aphids. Ann. Appl. Biol. 141, 1–10.

Kalischuk, M.L., Fusaro, A.F., Waterhouse, P.M., Pappu, H.R., Kawchuk, L.M., 2013. Complete genomic sequence of a *Rubus yellow net virus* isolate and detection of genome-wide pararetrovirus-derived small RNAs. Virus Res. 178, 306–313.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kunii, M., Kanda, M., Nagano, H., Uyeda, I., Kishima, Y., Sano, Y., 2004. Reconstruction of putative DNA virus from endogenous *Rice tungro bacilliform virus*-like sequences in the rice genome: implications for integration and evolution. BMC Genomics 5, 80.

Laney, A.G., Hassan, M., Tzanetakis, I.E., 2012. An integrated badnavirus is prevalent in fig germplasm. Phytopathology 102, 1182–1189.

Lyttle, D.J., Orlovich, D.A., Guy, P.L., 2011. Detection and analysis of endogenous badnaviruses in the New Zealand flora. AoB Plants 2011, plr008.

Martin, R.R., MacFarlane, S., Sabanadzovic, S., Quito, D., Poudel, B., Tzanetakis, I.E., 2012. Viruses and virus diseases of *Rubus*. Plant Dis. 97, 168–182.

Ndowora, T., Dahal, G., LaFleur, D., Harper, G., Hull, R., Olszewski, N.E., Lockhart, B., 1999. Evidence that badnavirus infection in *Musa* can originate from integrated pararetroviral sequences. Virology 255, 214–220.

Quito-Avila, D.F., Martin, R.R., 2012. Real-time RT-PCR for detection of *Raspberry bushy dwarf virus*, *Raspberry leaf mottle virus* and characterizing synergistic interactions in mixed infections. J. Virol. Methods 179, 38–44.

Richert-Poggeler, K.R., Noreen, F., Schwarzacher, T., Harper, G., Hohn, T., 2003. Induction of infectious *Petunia vein clearing* (pararetro) *virus* from endogenous provirus in petunia. EMBO J. 22, 4836–4845.

Richert-Pöggeler, K.R., Shepherd, R.J., 1997. Petunia Vein-Clearing Virus: A plant pararetrovirus with the core sequences for an integrase function. Virology 236, 137–146.

Shahid, M.S., Aboughanem-Sabanadzovic, N., Sabanadzovic, S., Tzanetakis, I.E., 2016. Genomic characterization and population structure of a badnavirus infecting blackberry. Plant Dis. PDIS-04-16-0527-RE.

Stace-Smith, R., 1955. Studies on *Rubus* virus diseases in British Columbia: I. *Rubus* yellow-Net. Can. J. Bot. 33, 269–274.

Stace-Smith, R., Jones, A.T., 1987. *Rubus* yellow net. Agric. Handb. - U. S. Dep. Agric. Comb. For. Pest Res. Dev. Program USA.

Staginnus, C., Richert-Pöggeler, K.R., 2006. Endogenous pararetroviruses: two-faced travelers in the plant genome. Trends Plant Sci. 11, 485–491.

Umber, M., Filloux, D., Muller, E., Laboureau, N., Galzi, S., Roumagnac, P., Iskra-Caruana, M.-L., Pavis, C., Teycheney, P.-Y., Seal, S.E., 2014. The genome of African yam (*Dioscorea cayenensis-rotundata* complex) hosts endogenous sequences from four distinct badnavirus species. Mol. Plant Pathol. 15, 790–801.

VanBuren, R., Bryant, D., Bushakra, J.M., Vining, K.J., Edger, P.P., Rowley, E.R., Priest, H.D., Michael, T.P., Lyons, E., Filichkin, S.A., Dossett, M., Finn, C.E., Bassil, N.V., Mockler, T.C., 2016. The genome of black raspberry (*Rubus occidentalis*). Plant J. 87, 535–547.

Zerbino, D.R., 2002. Using the Velvet de novo assembler for short-read sequencing technologies, in: Current Protocols in Bioinformatics. John Wiley & Sons, Inc.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.

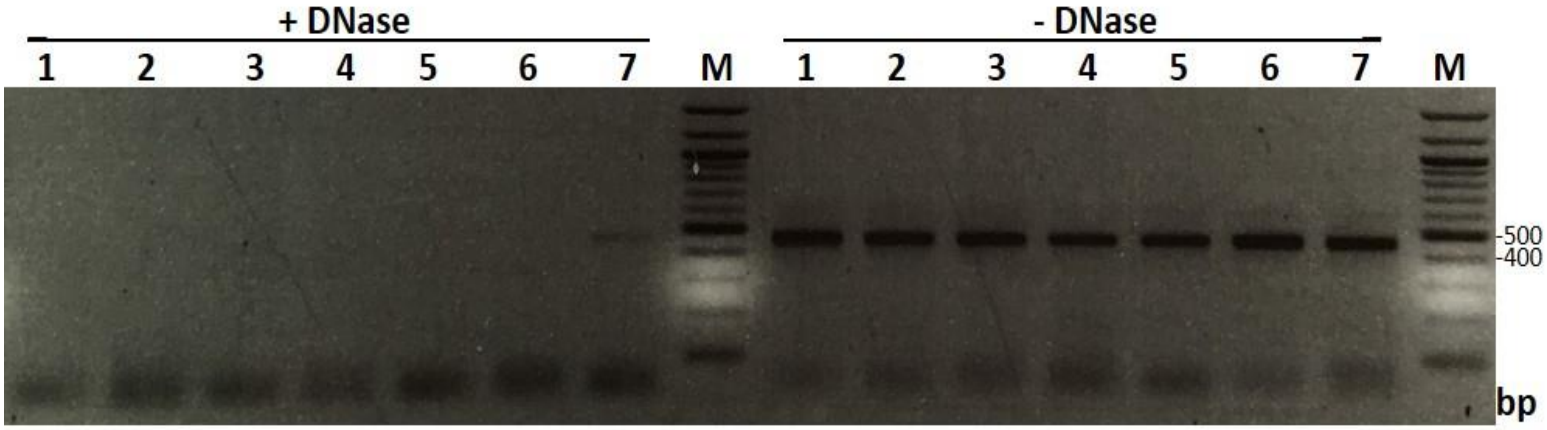

Figure 4.1. RT-PCR assay for RYNV in red raspberry samples. In order to determine whether RYNV is actively replicating or is incorporated into the raspberry genome a DNase treatment was included before the test. + DNase, total RNA extracts were DNase digested before RT-PCR; – DNase, total RNA extracts were used as template for RT-PCR without prior digestion. Lanes 1-7 represent different red raspberry cultivars: 1, 'Caroline'; 2, 'Comox'; 3, 'Glen Clova'; 4, 'Meeker'; 5, 'Willamette'; 6, 'Cascade Bounty'; 7, 'Baumforth's Seedling A'. Lane M contains 100 bp DNA molecular size marker. The expected size of the amplicon is 463 bp using the RYNV6F/RYNV6R oligonucleotide primers.

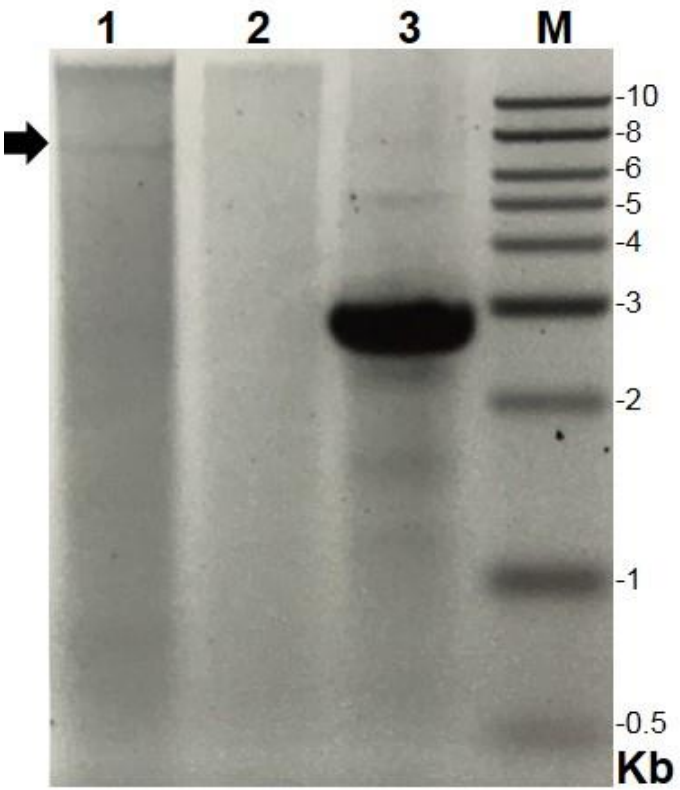

Figure 4.2. Agarose gel electrophoresis of RCA products previously digested with *Bam*HI restriction enzyme. Lane 1 and 2 contain products from reactions involving two different red raspberry samples: 1, 'Baumforth's Seedling A'; 2, 'Cascade Bounty'. Lane 3 is the positive control included in the commercial kit of RCA (pUC19 plasmid). Lane M contains 1 kb DNA molecular size marker. The arrow indicates the expected amplicon for episomal RYNV (7.9 kb).

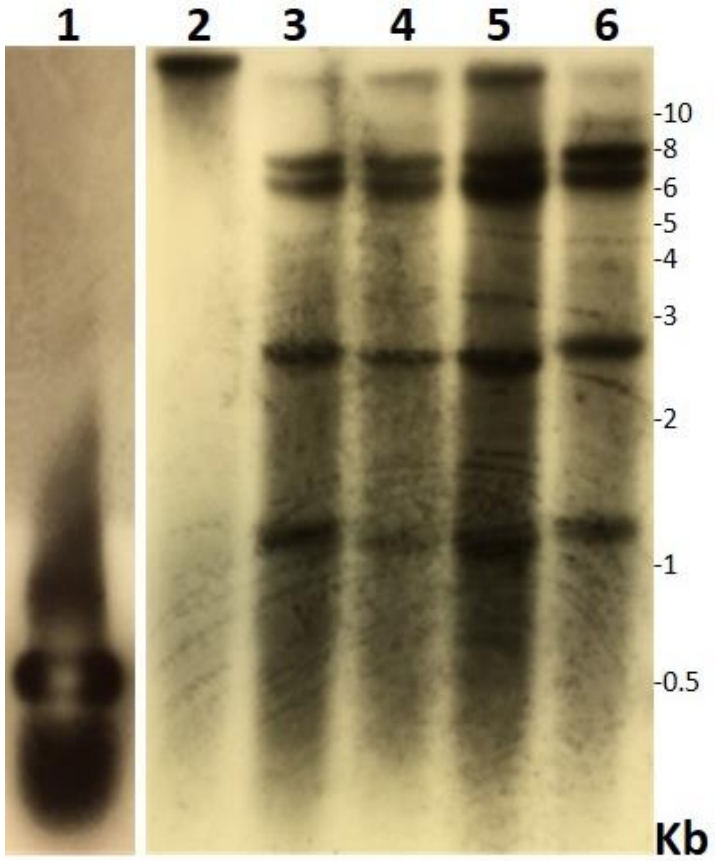

Figure 4.3. Southern blot analysis for RYNV using red raspberry DNA treated with *Bam*HI (lanes 3-6). Lane 1, amplicon produced by the RYNV6F/RYNV6R primers and employed to generate the hybridization probe; lane 2, undigested control of cultivar 'Cascade Bounty'; lane 3, digested cultivar 'Cascade Bounty'; lane 4, digested cultivar 'Comox'; lane 5, digested cultivar 'Glen Clova'; lane 6, digested cultivar 'Meeker'. Expected signal size for episomal RYNV cut with *Bam*HI is 7.9 kb.

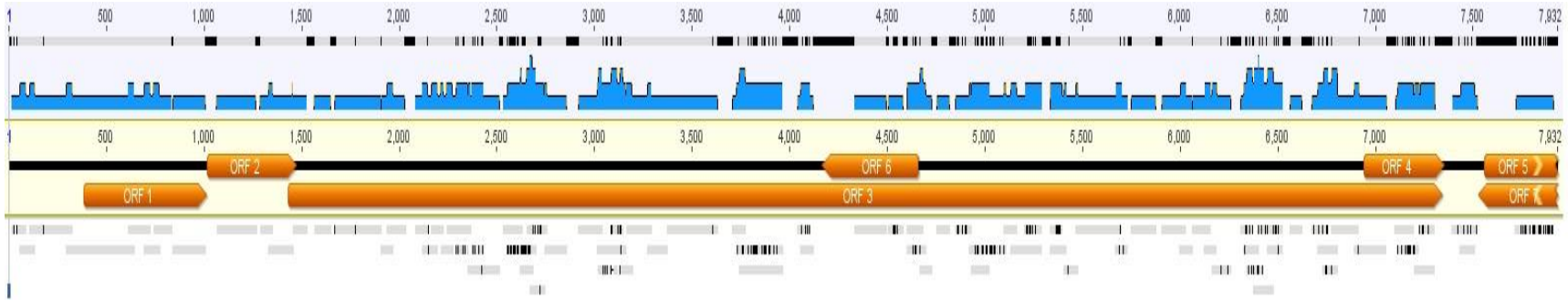

Figure 4.4. Diagram of RYNV-like contigs mapped against the RYNV genome. From top to bottom: consensus (gray-black); coverage (blue); RYNV genome (GenBank Accession no. KF241951), displaying open reading frames (black line and orange arrows); contigs (gray).

Table 4.1. *Rubus* cultivars involved in preliminary tests for RYNV presence. Result of PCR test using primers RYNV6F and RYNV6R. Cultivars were used as donor sources for grafting transmission over indicator plants, which were later analyzed for successful induction of symptoms. Plants were exposed to heat therapy and meristem tip culture protocols prior to a second RYNV-PCR test. N/A: not assessed.

Cultivar	PCR test	Graft transmission	Virus eradication
Baumforth's Seedling A	Positive	Successful	N/A
Caroline	Positive	Unsuccessful	N/A
Cascade Bounty	Positive	Unsuccessful	Unsuccessful
Cascade Delight	Positive	Unsuccessful	N/A
Cascade Harvest	Positive	Unsuccessful	Unsuccessful
Chilliwack	Positive	Unsuccessful	N/A
Comox	Positive	Unsuccessful	N/A
Cowichan	Positive	Unsuccessful	N/A
Fairview	Positive	Unsuccessful	N/A
Glen Clova	Positive	Unsuccessful	N/A
Glen Esk	Positive	Successful	N/A
Glen Moy	Positive	Successful	N/A
Malahat	Positive	Unsuccessful	N/A
Malling Landmark	Positive	Unsuccessful	N/A
Meeker	Positive	Unsuccessful	N/A
Prelude	Positive	Unsuccessful	N/A
Qualicum	Positive	Unsuccessful	N/A
Willamette	Positive	Unsuccessful	N/A

Chapter 5:

Conclusions

Alfredo Diaz-Lara

The family *Caulimoviridae* (also known as plant pararetroviruses) is the only plant virus group whose members have double-stranded DNA (dsDNA) genomes (Hull, 2013), with a mode of replication involving reverse transcription. *Cauliflower mosaic virus* is the type species of the family, this virus has been employed for understanding fundamental aspects of plant virology (Geering, 2007). On the other hand, less well known virus species in the family are serious pathogens in horticultural crops around the world.

The causal agent of the fruit-drop disorder in 'Bluecrop' blueberry plants was identified (Chapter 2). A novel virus, named *Blueberry fruit drop-associated virus* (BFDaV), was discovered in symptomatic material using rolling circle amplification (RCA). Further molecular characterization of BFDaV, concluded that the species belongs to the family *Caulimoviridae*, however, the genome organization and sequence differed considerably to any known genera in the group. Based on this information, it was proposed that BFDaV is a new genus in the family, which was supported by phylogenetic analyses (Diaz-Lara et al., 2016).

The correlation of BFDaV with the fruit-drop symptom was confirmed with a set of 66 plants in a field in northern Washington, where all symptomatic plants tested positive for the virus and all asymptomatic plants were negative for two growing seasons. BFDaV is known to occur in three fields in the Pacific Northwest, one in northern Washington and two in British Columbia, Canada. The grower in Washington is working to eradicate the virus from his farm. Recently, using the detection primers developed during this study, BFDaV was identified in 'Aron' blueberry samples that came from the Scandinavia region (R.R. Martin, unpublished), indicating that the virus distribution is more extensive than previously thought and is not restricted to the 'Bluecrop' cultivar.

RYNV was initially described in 1956 affecting Himalaya blackberry (*Rubus procerus*) (Stace-Smith, 1956). The first efforts at molecular characterization were carried out in 2002, when Jones et al. (2002) amplified a 1.7 kb fragment of the RYNV genome using badnavirus degenerate primers. This information led to the development of a PCR test for the virus and the subsequent observation that there were nontransmissible forms of the virus in some commercial red raspberry cultivars. The first sequenced genome of RYNV was reported in 2013 (Kalischuk et al., 2013). The complete characterization of a European strain of RYNV described here documented considerable variability in the species (Chapter 3), which differed from the Canadian isolate not only at genomic level but also in aphid transmission capacity. This novel strain, RYNV-BS, had a genome size of 7.8 kb containing five open reading frames rather than seven, as reported previously (Kalischuk et al., 2013). Furthermore, RYNV-BS was not aphid transmissible (RYNV is one of the few known badnaviruses transmitted by aphids), presumably as result of the altered genomic arrangement.

As mentioned previously, plant pararetroviruses replicate through an RNA intermediate, resembling retroviruses (RNA reverse-transcribing viruses), though it is believed that genome integration is not common during pararetrovirus replication (Bhat et al., 2016). Reports of illegitimate pararetroviral integrations are becoming more frequent, a few examples include: tobacco with integrated sequences of *Tobacco vein clearing virus* (Lockhart et al., 2000), rice with *Rice tungro bacilliform virus* (Kunii et al., 2004) and banana with different strains of *Banana streak virus* (Geering et al., 2005). Six out of eight genera (except for *Soymovirus* and *Rosadnavirus*) making up the family *Caulimoviridae*

have at least one species reported as able to incorporated in their host plant genomes (Bhat et al., 2016).

Using different methods, including graft indexing, reverse transcription-PCR with prior DNase digestion, Southern blot hybridization of the labeled viral probe to the digested DNA of the host, and RCA the occurrence of an endogenous form of RYNV into the red raspberry genome was demonstrated and provisionally named endogenous-*Rubus yellow net virus* (eRYNV) (Chapter 4). The integrated RYNV-like sequence in the plant DNA explains the positive results by PCR-based assays and the lack of induced symptoms in bioassays. The existence of eRYNV in the *Rubus* genome, supports the assertion of Geering et al. (2014), that endogenous plant pararetroviruses are widespread in flowering plants; in fact, they suggested a completely new genus (*Florendovirus*) inside the family *Caulimoviridae*, whose known members are only integrated (Geering et al., 2014). The above mentioned techniques, plus *in situ* hybridization and immunocapture (IC)-PCR have been employed previously to discriminate between incorporated and episomal forms of viruses (Bhat et al., 2016; Harper et al., 1999; Laney et al., 2012).

Next-generation sequencing was employed to elucidate the RYNV-like sequences present in the red raspberry genome, also putative virus-plant junctions were identified. The RYNV genome is fragmented and lost small regions inside the host DNA. The lack of a complete RYNV sequence could explain why the eRYNV never reactivated after tissue culture of the host plant.

Diagnostic testing, germplasm movement, disease management and taxonomy are problematic because of the occurrence of endogenous viruses (Bhat et al., 2016). Therefore, the use of a single diagnostic test (e.g. PCR) may not differentiate between

integrated and episomal virus sequences; this suggests the need for a more reliable combination of assays for virus detection. RCA is an innovative technique for the detection of caulimoviruses or any small virus with a circular DNA genome. Additionally, it has the advantage of identifying exogenous viruses specifically to be sequence-independent via random hexamer primers (Dean et al., 2001; Lizardi et al., 1998), thus overcoming the false positives generated with PCR.

Overall, the results obtained from this work highlight the importance of plant pararetroviruses for the agriculture, and describe a simple methodology for their identification as endogenous or exogenous forms. Also, they open the possibility for further studies; in the particular case of BFDaV, to identify the potential vector and alternative hosts should be a priority; and for RYNV, it is still necessary to unravel the implications of the incorporated virus sequence for the disease development.

Bibliography

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anonymous, 2016. International Committee on Taxonomy of Viruses, Virus Taxonomy: 2015 Release. http://www.ictvonline.org/virusTaxonomy.asp. Accessed 11.29.2016.

Bhat, A.I., Hohn, T., Selvarajan, R., 2016. Badnaviruses: the current global scenario. Viruses 8, 177.

Bouhida, M., Lockhart, B.E., Olszewski, N.E., 1993. An analysis of the complete sequence of a *Sugarcane bacilliform virus* genome infectious to banana and rice. J. Gen. Virol. 74, 15–22.

Dean, F.B., Nelson, J.R., Giesler, T.L., Lasken, R.S., 2001. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. 11, 1095–1099.

Diaz-Lara, A., Martin, R.R., 2016. *Blueberry fruit drop-associated virus*: a new member of the family *Caulimoviridae* isolated from blueberry exhibiting fruit-drop symptoms. Plant Dis. 100, 2211–2214.

Diaz-Lara, A., Mosier, N.J., Keller, K.E., Martin, R.R., 2015. A variant of *Rubus yellow net virus* with altered genomic organization. Virus Genes 50, 104–110.

Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Gibson, T.J., 2013. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. D259-D266.

Ellis, M.A., Nita, M., 2008. Organic small fruit disease management guidelines. Ohio State Univ. Ohio Agric. Res. Dev. Cent.

Geering, A.D., 2007. *Caulimoviridae* (plant pararetroviruses). eLS.

Geering, A.D.W., Hull, R., 2012. Family *Caulimoviridae*, Pp 429-443, In: Virus Taxonomy, 9th Report of the International Committee on Taxonomy of Viruses (Eds. King, A.M.Q, Adams, M.J., Carstens, E.B. and Lefkowitz, E.J.).

Geering, A.D.W., Maumus, F., Copetti, D., Choisne, N., Zwickl, D.J., Zytnicki, M., McTaggart, A.R., Scalabrin, S., Vezzulli, S., Wing, R.A., Quesneville, H., Teycheney, P.-Y., 2014. Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. Nat. Commun. 5, 5269.

Geering, A.D.W., Olszewski, N.E., Harper, G., Lockhart, B.E.L., Hull, R., Thomas, J.E., 2005. Banana contains a diverse array of endogenous badnaviruses. J. Gen. Virol. 86, 511–520.

Geering, A.D.W., Scharaschkin, T., Teycheney, P.-Y., 2010. The classification and nomenclature of endogenous viruses of the family *Caulimoviridae*. Arch. Virol. 155, 123–131.

Gish, W., States, D.J., 1993. Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

Gregor, W., Mette, M.F., Staginnus, C., Matzke, M.A., Matzke, A.J.M., 2004. A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. Plant Physiol. 134, 1191–1199.

Hansen, C.N., Harper, G., Heslop-Harrison, J.S., 2005. Characterisation of pararetrovirus-like sequences in the genome of potato (*Solanum tuberosum*). Cytogenet. Genome Res. 110, 559–565.

Harper, G., Hull, R., Lockhart, B., Olszewski, N., 2002. Viral sequences integrated into plant genomes. Annu. Rev. Phytopathol. 40, 119–136.

Harper, G., Osuji, J.O., Heslop-Harrison, J.S., Hull, R., 1999. Integration of *Banana streak Badnavirus* into the *Musa* genome: molecular and cytogenetic evidence. Virology 255, 207–213.

Higo, K., Ugawa, Y., Iwamoto, M., Korenaga, T., 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 27, 297–300.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Hull, R., 2013. Plant Virology. Academic Press.

Hull, R., Covey, S.N., 1995. Retroelements: propagation and adaptation. Virus Genes 11, 105–118.

Hull, R., Geering, A., 2012. Virus taxonomy. Family *Caulimoviridae.* IXth report of the ICTV. In: King, A.M.Q., Lefkowitz, E., Adams, M.J. Carstens, E.B. (eds.). Elsevier Academic Press. 429-443.

Jakowitsch, J., Mette, M.F., Winden, J. Van Der, Matzke, M.A., Matzke, A.J.M., 1999. Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. Proc. Natl. Acad. Sci. 96, 13241–13246.

James, A.P., Geijskes, R.J., Dale, J.L., Harding, R.M., 2010. Development of a novel rolling-circle amplification technique to detect *Banana streak virus* that also discriminates between integrated and episomal virus sequences. Plant Dis. 95, 57–62.

Jones, A.T., McGavin, W.J., Geering, A.D.W., Lockhart, B.E.L., 2002. Identification of *Rubus yellow net virus* as a distinct badnavirus and its detection by PCR in *Rubus* species and in aphids. Ann. Appl. Biol. 141, 1–10.

Jones, A.T., Roberts, I.M., Murant, A.F., 1974. Association of different kinds of bacilliform particle with vein chlorosis and mosaic diseases of raspberry (*Rubus idaeus*). Ann. Appl. Biol. 77, 283–288.

Julian, J., Strik, B., Yang, W., 2011. Blueberry economics: the costs of establishing and producing blueberries in the Willamette Valley. Or. State Univ. Ext.

Kalischuk, M.L., Fusaro, A.F., Waterhouse, P.M., Pappu, H.R., Kawchuk, L.M., 2013. Complete genomic sequence of a *Rubus yellow net virus* isolate and detection of genome-wide pararetrovirus-derived small RNAs. Virus Res. 178, 306–313.

Kasteel, D.T.J., Perbal, M.C., Boyer, J.C., Wellink, J., Goldbach, R.W., Maule, A.J., Van Lent, J.W.M., 1996. The movement proteins of *Cowpea mosaic virus* and *Cauliflower mosaic virus* induce tubular structures in plant and insect cells. J. Gen. Virol. 77, 2857–2864.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kunii, M., Kanda, M., Nagano, H., Uyeda, I., Kishima, Y., Sano, Y., 2004. Reconstruction of putative DNA virus from endogenous *Rice tungro bacilliform virus*-like sequences in the rice genome: implications for integration and evolution. BMC Genomics 5, 80.

La Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K., Brunak, S., 2004. Analysis and prediction of leucine-rich nuclear export signals. Protein Engineering Design and Selection 17, 527-536.

Lam, E., Chua, N.H., 1989. ASF-2: a factor that binds to the *Cauliflower mosaic virus* 35S promoter and a conserved GATA motif in Cab promoters. Plant Cell Online 1, 1147–1156.
Laney, A.G., Hassan, M., Tzanetakis, I.E., 2012. An integrated badnavirus is prevalent in fig germplasm. Phytopathology 102, 1182–1189.

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

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., Ward, D.C., 1998. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat. Genet. 19, 225–232.

Lockhart, B., 1990. Evidence for a double-stranded circular DNA genome in a second group of plant viruses. Phytopathology 80, 127–131.

Lockhart, B.E., Menke, J., Dahal, G., Olszewski, N.E., 2000. Characterization and genomic analysis of *Tobacco vein clearing virus*, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. J. Gen. Virol. 81, 1579–1585.

Lyttle, D.J., Orlovich, D.A., Guy, P.L., 2011. Detection and analysis of endogenous badnaviruses in the New Zealand flora. AoB Plants 2011, plr008.

Maddison, W.P., Maddison, D.R., 2001. Mesquite: a modular system for evolutionary analysis, version 0.98. http://mesquiteproject.org

Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32, W327–W331.

Martin, R., Tzanetakis, I., Caruso, F., Polashock, J., 2008. Emerging and reemerging virus diseases of blueberry and cranberry, in: IX International *Vaccinium* Symposium 810. pp. 299–304.

Martin, R.R., MacFarlane, S., Sabanadzovic, S., Quito, D., Poudel, B., Tzanetakis, I.E., 2012. Viruses and virus diseases of *Rubus*. Plant Dis. 97, 168–182.

Martin, R.R., Polashock, J.J., Tzanetakis, I.E., 2012b. New and emerging viruses of blueberry and cranberry. Viruses 4, 2831–2852.

Martin, R.R., Tzanetakis, I.E., Sweeney, M., Wegener, L., 2006. A virus associated with blueberry fruit drop disease. Acta Hortic. 715, 497–501.

Martin, R.R., Zhou, J., Tzanetakis, I.E., 2011. *Blueberry latent virus*: An amalgam of the *Partitiviridae* and *Totiviridae*. Virus Res. 155, 175–180.

McGavin, W. J., MacFarlane, S.A., 2010. Sequence similarities between *Raspberry leaf mottle virus*, *Raspberry leaf spot virus* and the closterovirus *Raspberry mottle virus*. Ann. Appl. Biol. 156, 439–448.

Medberry, S.L., Lockhart, B.E., Olszewski, N.E., 1992. The *Commelina yellow mottle virus* promoter is a strong promoter in vascular and reproductive tissues. Plant Cell Online 4, 185–192.

Medberry, S.L., Lockhart, B.E.L., Olszewski, N.E., 1990. Properties of *Commelina yellow mottle virus*'s complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. Nucleic Acids Res. 18, 5505–5513.

Ndowora, T., Dahal, G., LaFleur, D., Harper, G., Hull, R., Olszewski, N.E., Lockhart, B., 1999. Evidence that badnavirus infection in *Musa* can originate from integrated pararetroviral sequences. Virology 255, 214–220.

Quito-Avila, D.F., Lightle, D., Lee, J., Martin, R.R., 2012. Transmission biology of *Raspberry latent virus*, the first aphid-borne reovirus. Phytopathology 102, 547–553.

Quito-Avila, D.F., Martin, R.R., 2012. Real-time RT-PCR for detection of *Raspberry bushy dwarf virus*, *Raspberry leaf mottle virus* and characterizing synergistic interactions in mixed infections. J. Virol. Methods 179, 38–44.

Regenmortel, M.H.V. van, Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B., 2000. Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses.

Richert-Poggeler, K.R., Noreen, F., Schwarzacher, T., Harper, G., Hohn, T., 2003. Induction of infectious *Petunia vein clearing* (pararetro) *virus* from endogenous provirus in petunia. EMBO J. 22, 4836–4845.

Richert-Pöggeler, K.R., Shepherd, R.J., 1997. Petunia Vein-Clearing Virus: A plant pararetrovirus with the core sequences for an integrase function. Virology 236, 137–146.

Roy, A., Shao, J., Schneider, W.L., Hartung, J.S., Brlansky, R.H., 2014. Population of endogenous pararetrovirus genomes in Carrizo citrange. Genome Announc. 2, e01063-13.

Schmidt, I., Blanc, S., Esperandieu, P., Kuhl, G., Devauchelle, G., Louis, C., Cerutti, M., 1994. Interaction between the aphid transmission factor and virus particles is a part of the molecular mechanism of *Cauliflower mosaic virus* aphid transmission. Proc. Natl. Acad. Sci. 91, 8885–8889.

Sether, D.M., Melzer, M.J., Borth, W.B., Hu, J.S., 2012. *Pineapple bacilliform CO virus*: diversity, detection, distribution, and transmission. Plant Dis. 96, 1798–1804.

Shahid, M.S., Aboughanem-Sabanadzovic, N., Sabanadzovic, S., Tzanetakis, I.E., 2016. Genomic characterization and population structure of a badnavirus infecting blackberry. Plant Dis. PDIS-04-16-0527-RE.

Shahid, M.S., Tzanetakis, I.E., Aboughanem-Sabanadzovic, N., Sabanadzovic, S., 2016. An integrating Badnavirus infects blackberry. Acta Hortic. 507–510.

Song, J., Tan, H., Perry, A.J., Akutsu, T., Webb, G.I., Whisstock, J.C., Pike, R.N., 2012. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. PloS one 7, e50300.

Stace-Smith, R., 1955. Studies on *Rubus* virus diseases in British Columbia: I. *Rubus* yellow-Net. Can. J. Bot. 33, 269–274.

Stace-Smith, R., 1956. Studies on *Rubus* virus diseases in British Columbia: III. Separation of components of raspberry mosaic. Can. J. Bot. 34, 435–442.

Stace-Smith, R., Jones, A.T., 1987. *Rubus* yellow net. Agric. Handb. - U. S. Dep. Agric. Comb. For. Pest Res. Dev. Program USA.

Staginnus, C., Richert-Pöggeler, K.R., 2006. Endogenous pararetroviruses: two-faced travelers in the plant genome. Trends Plant Sci. 11, 485–491.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Umber, M., Filloux, D., Muller, E., Laboureau, N., Galzi, S., Roumagnac, P., Iskra-Caruana, M.-L., Pavis, C., Teycheney, P.-Y., Seal, S.E., 2014. The genome of African yam (*Dioscorea cayenensis-rotundata* complex) hosts endogenous sequences from four distinct badnavirus species. Mol. Plant Pathol. 15, 790–801.

VanBuren, R., Bryant, D., Bushakra, J.M., Vining, K.J., Edger, P.P., Rowley, E.R., Priest, H.D., Michael, T.P., Lyons, E., Filichkin, S.A., Dossett, M., Finn, C.E., Bassil, N.V., Mockler, T.C., 2016. The genome of black raspberry (*Rubus occidentalis*). Plant J. 87, 535–547.

Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353–3362.

Xu, D., Mock, R., Kinard, G., Li, R., 2011. Molecular analysis of the complete genomic sequences of four isolates of *Gooseberry vein banding associated virus*. Virus Genes 43, 130–137.

Zerbino, D.R., 2002. Using the Velvet de novo assembler for short-read sequencing technologies, in: Current Protocols in Bioinformatics. John Wiley & Sons, Inc.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.

Zhou, J., Bean, R.L., Vogt, V.M., Summers, M., 2007. Solution structure of the *Rous sarcoma virus* nucleocapsid protein: μΨ RNA packaging signal complex. J. Mol. Biol. 365, 453–467.

APPENDIX

Supplemental Table S 2.1. Primers used for BFDaV sequencing. Binding sites in reference to BFDaV genome sequence. Products correspond to size in base pairs.

Primer	Type	Sequence (5'-3')	Binding Site	Product
Preliminary sequencing: Primer Walking				
FDFr2F1158	Sense	CCAAGGATTGAACCCAATCCG	2763-2783	
FDFr2R1284	Antisense	CTTGAGCTCTTTGTTTCTTCCACC	3456-3479	
FDFr2Reverse	Antisense	GACAGGTCTTCTTGTTGCTCTAAG	4254-4277	
FDFr1For2	Sense	CGCACCATATACGCAATACCC	5087-5107	
FDFr1F1042	Sense	GTAAGAGTCCACATTCCAGCC	5884-5904	
FDFr1F1989	Sense	GAAGTCTTAGCCATCATCCGTAC	6831-6853	
FDFr1F2883	Sense	CTCAAACTCAAATGTCCTCCCATC	7725-7748	
FDFr2For1	Sense	GGAGGTTCGAAGCACTAGAGTC	8933-8954	
FDFr1R1489	Antisense	CAGACTCTAGTGCTTCGAACC	8936-8956	
FDFr1R208	Antisense	GATCTGGTAGTGCTTGAATGCG	9762-9783	
Closing gaps				
FD1For1	Sense	CCAGAAGGCACAACAGAAAGC	988-1009	269
FD2Rev1	Antisense	GAGCTGGATCACTGACAGGTTTG	1235-1257	
FD2For1	Sense	GTCCTACAGCAAATTCCACAGC	4596-4617	362
FD1Rev1	Antisense	CAAGTTTCTCCATTTGCCACTGTG	4935-4957	
Resequencing				
F5082	Sense	GATCTCCCACAGCATCACAG	74-93	1322
R6382	Antisense	CTCCTTTCACTGACTGCAATTCC	1374-1396	
F5989	Sense	GATGAGAACCAGAAGGCACAAC	981-1002	1318
R6270	Antisense	GCTTTGCTGTGACTACTTCCTTC	2277-2299	
F6914	Sense	CGTGCACAATTGAAGATGCC	1906-1925	1297
R7175	Antisense	GCCAAGGTTGTAGTATGCCAAG	3182-3203	
F6800	Sense	GCTCGACATCGACAACAACC	2807-2826	1273
R8052	Antisense	GATCGCTTCGTTGTATGACTGG	4059-4080	
F7675	Sense	GCTGGAACTGTAATCAGCAGG	3682-3702	1275
R176	Antisense	CAAGTTTCTCCATTTGCCACTGTG	4934-4957	
F8633	Sense	CATCACCGTAAAGAGTTCTCCG	4640-4661	1241
R1017	Antisense	GGCTGAATGAGTTCCAGTTTGAG	5859-5881	
F794	Sense	GCCCAACCTAAATGTGCTCG	5636-5655	1129
R1903	Antisense	CTTCTTCCTTCTTGTCAGGGC	6745-6765	
F1554	Sense	GGAGTCACCATATGCAAAGGAAAG	6396-6419	1219
R2751	Antisense	CAAGGAGTGTTGATGATCCGTTG	7593-7615	
F2428	Sense	CTCCACCGTTCGAATACAAGAC	7270-7291	1240
R3658	Antisense	GTAGCACATGGGATGGTGG	8500-8518	

Supplemental Table S 2.1. Primers used for BFDaV sequencing. Binding sites in reference to BFDaV genome sequence. Products correspond to size in base pairs (Continued).

Primer	Type	Sequence (5'-3')	Binding Site	Product
		Resequencing		
F3300	Sense	GTCAGCAAATGGAGAGCCTTTG	8142-8163	1221
R4502	Antisense	CCATACCAAGACTTGTGGCC	9344-9363	
F4288	Sense	CAGTGAGGTGTAGGCCTAGATC	9130-9151	1154
R5424	Antisense	CAGCGTCAAGAACACGCC	416-433	

Supplemental Table S 2.2. PCR detection of BFDaV from 'Bluecrop' blueberries in a row of 66 blueberry plants with approximately one half of plants exhibiting symptoms from a field in northern Washington state. Samples collected in spring and summer of 2015.

Lab Sample	Cultivar	Symptoms	BFDaV presence
1	Bluecrop	None	-
2	Bluecrop	None	-
3	Bluecrop	Fruit drop	+
4	Bluecrop	None	-
5	Bluecrop	None	-
6	Bluecrop	Fruit drop	+
7	Bluecrop	Fruit drop	+
8	Bluecrop	? Shock Symptoms	+
9	Bluecrop	Fruit drop	+
10	Bluecrop	Fruit drop	+
11	Bluecrop	Fruit drop	+
12	Bluecrop	Fruit drop	+
13	Plant missing		
14	Bluecrop	Fruit drop	+
15	Bluecrop	Fruit drop	+
16	Bluecrop	Fruit drop	+
17	Plant missing		
18	Bluecrop	Fruit drop	+
19	Bluecrop	Fruit drop	+
20	Bluecrop	Fruit drop	+
21	Bluecrop	Fruit drop	+
22	Bluecrop	Fruit drop	+
23	Bluecrop	Fruit drop	+
24	Bluecrop	Fruit drop	+
25	Bluecrop	Fruit drop	+
26	Bluecrop	Fruit drop	+
27	Bluecrop	Fruit drop	+
28	Bluecrop	Fruit drop	+
29	Bluecrop	None	-
30	Bluecrop	None	-
31	Bluecrop	None	-
32	Bluecrop	None	-
33	Bluecrop	Fruit drop	+
34	Bluecrop	None	-
35	Bluecrop	None	-
36	Bluecrop	None	-

Supplemental Table S 2.2. PCR detection of BFDaV from 'Bluecrop' blueberries in a row of 66 blueberry plants with approximately one half of plants exhibiting symptoms from a field in northern Washington state. Samples collected in spring and summer of 2015 (Continued).

Lab Sample	Cultivar	Symptoms	BFDaV presence
37	Bluecrop	None	-
38	Bluecrop	Fruit drop	+
39	Bluecrop	Fruit drop	+
40	Bluecrop	Fruit drop	+
41	Bluecrop	Fruit drop	+
42	Bluecrop	Fruit drop	+
43	Bluecrop	Fruit drop	+
44	Bluecrop	Fruit drop	+
45	Bluecrop	None	-
46	Bluecrop	Fruit drop	+
47	Bluecrop	None	-
48	Bluecrop	Fruit drop	+
49	Bluecrop	Fruit drop	+
50	Bluecrop	Fruit drop	+
51	Bluecrop	Fruit drop	+
52	Bluecrop	None	-
53	Bluecrop	None	-
54	Bluecrop	None	-
55	Bluecrop	None	-
56	Bluecrop	None	-
57	Bluecrop	None	-
58	Bluecrop	None	-
59	Bluecrop	None	-
60	Bluecrop	None	-
61	Bluecrop	None	-
62	Bluecrop	None	-
63	Bluecrop	None	-
64	Bluecrop	None	-
65	Bluecrop	None	-
66	Bluecrop	None	-
67	Bluecrop	None	-
68	Bluecrop	None	-

Supplemental Table S 3.1. Primers used for RYNV-BS sequencing. Binding sites in reference to RYNV-BS genome sequence. Products correspond to base pairs.

Primer	Type	Sequence (5'-3')	Binding Site	Product
Preliminary sequencing: Primer Walking				
ADL94-R838	Sense	AAGTCACTAGGCTCGAAAAA	1249-1269	
ADL94-R1	Sense	GGAAAGTCTCAAGCCGAGAC	1372-1392	
ADL113-F712	Sense	CAGAAGTGGTCCGTGGCAAAC	2219-2240	
ADL112-R112	Antisense	GGTATTAGGGTGAAGTCCTGTGG	3308-3331	
ADL94-F136	Antisense	TCTGGCCAAAACTCCAATG	3796-3815	
ADL97-F359	Antisense	GGTATGACGAACTCCACAAT	5015-5035	
ADL97-F697	Sense	CCACAATCCAAACAACCCTAGAGC	5373-5397	
ADL97-R179	Antisense	GACGTACCCGAGTGTGTGAT	5744-5764	
ADL97-R413	Sense	GATTCTATGAATGGAAGGTCA	5997-6018	
ADL108-F889	Sense	GCAGAACTAACCAGGAAAGACG	6983-7005	
Closing gaps				
Fr1For2	Sense	CACCCAGGAAGTGCCTATC	4402-4421	516
Fr2Rev1	Antisense	GCTCCGATGGTGTTACTGG	4899-4918	
Fr2For2	Sense	GAAGTCTCAACGACAATACCCAC	5823-5846	503
Fr3Rev1	Antisense	GGATTCGTCATCCACCTCG	6307-6326	
Fr3For2	Sense	GAGATCCGTGACTACTCACCAATC	7078-7102	1408
Fr1Rev1	Antisense	CGGTGATATTGAGATCGAACAAGG	625-650	
Resequencing				
F7682	Sense	CTCTCTTGAATTTCCGTGCTACC	519-542	1147
R970	Antisense	GCTCTTCTCCTTCGACTGAGG	1645-1666	
F757	Sense	GAGCAGAAGCCACACAAGG	1432-1451	1039
R1777	Antisense	CCTAGGGCTGGAGTGTTTG	2452-2471	
F1597	Sense	GATAGCAGCACCACTTGCAG	2272-2292	1161
R2735	Antisense	CCAAGAAGGTTCTCGACGTAGTC	3410-3433	
F2484	Sense	CATCCACTATGTCCACCATTGG	3159-3181	1154
R3622	Antisense	CTGTGTCCATGGCATCAGG	4294-4313	
F3331	Sense	CAAGTACAAGCATACGCAGGG	4003-4024	1110
R4420	Antisense	CTGTTAATGCACGTGCAGGTG	5092-5113	
F4234	Sense	CACCATCGGAGCACTTCTAG	4906-4926	1111
R5321	Antisense	GACCTTCCATTCATAGAATCCGAC	5993-6017	
F5106	Sense	CCAAGAAAGAAGTCAGAGGGAAAG	5778-5802	1039
R6121	Antisense	GAAGGTTTCTTTGCATTCAGCTTC	6793-6817	
F5941	Sense	CTCAGCAAGCAAGGAAGAAATCTG	6613-6637	1185
R7101	Antisense	GCGCTTTCGAACACTTAAGAACTC	7774-7798	
F6821	Sense	GCGCTGAGTCATAGTGATAAGG	7493-7515	1161
R121	Antisense	CCTTGATTTCCTTCAGATCTGCC	795-818	

Supplemental Table S 3.2. Viruses and sequences used in the phylogenetic analysis, including GenBank accession numbers.

Virus	Abbreviation	GenBank Accession Numbers				Reference
		ORF 1 (aa)	ORF 2 (aa)	ORF 3 (aa)	Genomic DNA	
Banana streak Mysore virus	BSV_MY	AAW80646.1	AAW80647.1	AAW80648.1	AY805074	Geering et al., 2005
Banana streak OL virus	BSV_OL	NP569148	NP569149.1	NP569150.1	NC003381	Harper et al., 1998
Bougainvillea spectabilis chlorotic vein-banding virus	BCVBV	YP002321511.1	YP002321512.1	YP002321513.1	NC011592	Wang et al., 2008
Cacao swollen shoot virus	CSSV	NP041732.1	NP041733.1	NP041734.1	NC001574	Hagen et al., 1993
Citrus yellow mosaic virus	CYMV	NP569151.1	NP569152.1	NP569153.1	NC003382	Huang et al., 2001
Commelina yellow mottle virus	ComYMV	NP039818.1	NP039819.1	NP039820.1	NC001343	Medberry et al., 1990
Dioscorea bacilliform virus	DBV	ABI47981.1	ABI47982.1	ABI47983.1	DQ822073	Seal et al., 2007
Dracaena mottle virus	DMV	ABE77342.1	ABE77343.1	ABE77344.1	DQ473478	Su et al., 2007
Fig badnavirus 1 isolate Arkansas	FBV_1	AEF56562.1	AEF56563.1	AEF56561.1	JF411989	Laney at al., 2012
Gooseberry vein banding virus BC isolate	GVBAV	AEE39274.1	AEE39275.1	AEE39276.1	HQ852250	Xu et al., 2011
Kalanchoe top-spotting virus	KTSV	NP777315.1	NP777316.1	NP777317.1	NC004540	Yang et al., 2003
Pineapple bacilliform comosus virus isolate HI1	PBV_CO	AEV42074.1	AEV42075.1	AEV42076.1	GQ398110	Sether et al., 2012
Rice tungro bacilliform virus	RTBV	CAA40995.1	CAA40996.1	CAA40997.1	X57924	Hay et al., 1991
Rubus yellow net virus isolate Canada	RYNV_CA	AHB61258	AHB61259	AHB61260	KF241951	Kalischuk et al., 2013
Sugarcane bacilliform IM virus	SCBV_IM	NP149411.1	NP149412.1	NP149413.1	NC003031	Geijskes et al., 2002
Taro bacilliform virus	TaBV	AAN75638.1	ANN75639.1	ANN75640.1	AF357836	Yang et al., 2003

Supplemental Table S 4.1. Contigs searched against the RYNV genome. Contigs generated after the extension process were compared with two different complete genomes of RYNV available in the GenBank (Accessions no. KF241951.1 and NC_026238.1). BLASTn analysis with word size=20.

Query id	Subject id	% identity	Alignment length	Mismatches	Query length	Subject length	Query start	Query end	Subject start	Subject end	E-value
NODE_64337_cov_517.076050	KF241951.1	100	201	0	201	7932	1	201	1267	1067	2.00E-106
NODE_114146_cov_225.792679	KF241951.1	100	112	0	112	7932	1	112	613	724	4.00E-57
NODE_309804_cov_217.711105	NC_026238.1	100	21	0	570	7836	188	208	6386	6366	9.00E-06
NODE_623305_cov_484.000000	KF241951.1	100	65	0	65	7932	1	65	4120	4056	3.00E-31
NODE_664769_cov_312.708344	KF241951.1	100	102	0	102	7932	1	102	6805	6704	1.00E-51
NODE_887024_cov_266.225800	KF241951.1	88.52	61	7	61	7932	1	61	5395	5335	2.00E-17
NODE_910113_cov_115.964073	NC_026238.1	100	20	0	197	7836	45	64	4186	4167	1.00E-05
NODE_1163691_cov_162.954544	KF241951.1	100	74	0	74	7932	1	74	7507	7434	3.00E-36
NODE_1178070_cov_238.064514	KF241951.1	100	61	0	61	7932	1	61	2213	2273	4.00E-29
NODE_1187696_cov_521.770508	KF241951.1	91.49	94	5	91	7932	1	91	3021	3114	9.00E-33
NODE_1202419_cov_667.698914	KF241951.1	100	123	0	123	7932	1	123	5906	6028	3.00E-63
NODE_1326839_cov_664.587646	KF241951.1	100	127	0	127	7932	1	127	3290	3164	2.00E-65
NODE_1326840_cov_679.484863	KF241951.1	98.41	63	1	63	7932	1	63	3193	3131	2.00E-28
NODE_1377895_cov_255.014359	KF241951.1	100	20	0	239	7932	43	62	2762	2781	1.00E-05
NODE_1385284_cov_250.978729	KF241951.1	98.7	77	1	77	7932	1	77	6523	6447	3.00E-36
NODE_1385285_cov_246.514282	KF241951.1	100	100	0	100	7932	1	100	6476	6377	2.00E-50
NODE_1623293_cov_770.725464	KF241951.1	100	81	0	81	7932	1	81	4600	4680	5.00E-40
NODE_1658296_cov_284.409760	KF241951.1	99.15	235	2	235	7932	1	235	1905	1671	8.00E-122
NODE_2369098_cov_455.353668	KF241951.1	89.18	194	21	194	7932	1	194	4540	4347	2.00E-67
NODE_2609319_cov_130.600006	KF241951.1	92.59	108	8	110	7932	3	110	7114	7221	1.00E-41
NODE_2609326_cov_272.847839	KF241951.1	94.74	76	4	76	7932	1	76	7227	7302	1.00E-30
NODE_2689219_cov_561.868408	KF241951.1	100	220	0	220	7932	1	220	3961	3742	8.00E-117

Supplemental Table S 4.1. Contigs searched against the RYNV genome. Contigs generated after the extension process were compared with two different complete genomes of RYNV available in the GenBank (Accessions no. KF241951.1 and NC_026238.1). BLASTn analysis with word size=20 (Continued).

NODE_2697558_cov_660.515137	KF241951.1	100	63	0	63	7932	1	63	4817	4755	4.00E-30
NODE_2912037_cov_591.354858	KF241951.1	91.8	61	5	61	7932	1	61	3145	3085	1.00E-20
NODE_2945355_cov_490.626099	KF241951.1	86.9	145	19	145	7932	1	145	2554	2698	1.00E-43
NODE_2984209_cov_330.962158	KF241951.1	100	347	0	347	7932	1	347	296	642	0
NODE_3131720_cov_237.666672	KF241951.1	100	96	0	96	7932	1	96	2533	2628	3.00E-48
NODE_3145452_cov_668.609741	KF241951.1	100	112	0	112	7932	1	112	2745	2856	4.00E-57
NODE_3193230_cov_285.109589	KF241951.1	100	103	0	103	7932	1	103	7302	7200	4.00E-52
NODE_3195454_cov_245.322586	KF241951.1	100	92	0	92	7932	1	92	834	743	4.00E-46
NODE_3195455_cov_292.958344	KF241951.1	100	78	0	78	7932	1	78	772	695	2.00E-38
NODE_3666317_cov_508.818848	KF241951.1	100	168	0	168	7932	1	168	1008	841	5.00E-88
NODE_3676197_cov_613.872986	KF241951.1	100	93	0	93	7932	1	93	6154	6062	1.00E-46
NODE_3755950_cov_119.179108	KF241951.1	93.81	97	6	97	7932	1	97	4600	4696	7.00E-39
NODE_4276735_cov_244.940002	KF241951.1	93.75	80	5	80	7932	1	80	4848	4927	4.00E-31
NODE_4964970_cov_384.207550	KF241951.1	95.06	81	4	83	7932	1	81	6673	6753	2.00E-33
NODE_4964972_cov_376.863647	KF241951.1	94.59	74	4	74	7932	1	74	6732	6805	2.00E-29
NODE_4990703_cov_120.125000	KF241951.1	100	118	0	118	7932	1	118	3037	2920	2.00E-60
NODE_5071015_cov_243.101700	KF241951.1	93.26	89	6	89	7932	1	89	6339	6427	2.00E-34
NODE_5071016_cov_340.979156	KF241951.1	93.65	126	8	126	7932	1	126	6398	6523	2.00E-51
NODE_6389488_cov_637.460510	KF241951.1	99.61	258	1	258	7932	1	258	3632	3375	3.00E-136
NODE_6506147_cov_249.787582	KF241951.1	100	20	0	336	7932	51	70	1101	1082	2.00E-05
NODE_6913866_cov_657.208740	KF241951.1	99.58	236	1	236	7932	1	236	5464	5699	5.00E-124
NODE_6913867_cov_689.806458	KF241951.1	96.72	61	2	61	7932	1	61	5670	5730	9.00E-26
NODE_7423605_cov_383.288452	KF241951.1	100	82	0	82	7932	1	82	5335	5416	1.00E-40
NODE_7707067_cov_354.106384	KF241951.1	98.7	77	1	77	7932	1	77	6402	6326	3.00E-36

Supplemental Table S 4.1. Contigs searched against the RYNV genome. Contigs generated after the extension process were compared with two different complete genomes of RYNV available in the GenBank (Accessions no. KF241951.1 and NC_026238.1). BLASTn analysis with word size=20 (Continued).

NODE_7805668_cov_421.154633	KF241951.1	95.12	123	6	127	7932	1	123	7393	7515	4.00E-53
NODE_8018477_cov_142.771423	KF241951.1	100	65	0	65	7932	1	65	5098	5162	3.00E-31
NODE_8920531_cov_670.635437	KF241951.1	100	126	0	126	7932	1	126	5747	5872	7.00E-65
NODE_9025221_cov_621.548401	KF241951.1	100	61	0	61	7932	1	61	6059	5999	4.00E-29
NODE_9188551_cov_477.482147	KF241951.1	91.86	86	7	86	7932	1	86	5290	5205	4.00E-31
NODE_9208949_cov_473.129028	KF241951.1	91.8	61	5	61	7932	1	61	2431	2371	1.00E-20
NODE_9307179_cov_344.843750	KF241951.1	91.94	62	5	62	7932	1	62	4042	4103	3.00E-21
NODE_10107233_cov_229.196976	KF241951.1	99.38	162	1	162	7932	1	162	2354	2515	4.00E-83
NODE_10369074_cov_441.585175	KF241951.1	100	165	0	165	7932	1	165	4495	4331	2.00E-86
NODE_10447828_cov_119.325584	KF241951.1	91.78	73	6	73	7932	1	73	4507	4579	1.00E-25
NODE_10841650_cov_302.500000	KF241951.1	100	94	0	94	7932	1	94	5022	4929	3.00E-47
NODE_10889242_cov_470.203125	KF241951.1	90.16	61	6	94	7932	34	94	1690	1630	7.00E-19
NODE_11761644_cov_281.826080	KF241951.1	100	76	0	76	7932	1	76	133	58	3.00E-37
NODE_11761645_cov_290.487793	KF241951.1	97.18	71	2	71	7932	1	71	87	17	3.00E-31
NODE_11804955_cov_234.326080	KF241951.1	100	122	0	122	7932	1	122	2365	2244	1.00E-62
NODE_11909079_cov_117.408699	KF241951.1	99.31	145	1	145	7932	1	145	3016	3160	1.00E-73
NODE_11969271_cov_209.808517	KF241951.1	100	20	0	77	7932	41	60	5990	5971	4.00E-06
NODE_12449265_cov_346.877014	KF241951.1	99.54	217	1	217	7932	1	217	325	109	2.00E-113
NODE_12768035_cov_798.047607	KF241951.1	100	72	0	72	7932	1	72	1455	1526	4.00E-35
NODE_12827152_cov_124.829269	KF241951.1	91.91	235	19	235	7932	1	235	3727	3961	2.00E-93
NODE_12828568_cov_590.567139	KF241951.1	97.94	97	2	97	7932	1	97	6259	6163	2.00E-45
NODE_12994603_cov_629.483887	KF241951.1	100	61	0	61	7932	1	61	6562	6622	4.00E-29
NODE_13157263_cov_666.129028	KF241951.1	100	123	0	123	7932	1	123	1333	1455	3.00E-63
NODE_13222004_cov_629.977783	KF241951.1	98.67	75	1	75	7932	1	75	2743	2669	4.00E-35

Supplemental Table S 4.1. Contigs searched against the RYNV genome. Contigs generated after the extension process were compared with two different complete genomes of RYNV available in the GenBank (Accessions no. KF241951.1 and NC_026238.1). BLASTn analysis with word size=20 (Continued).

NODE_13424623_cov_622.179077	KF241951.1	99.39	164	1	164	7932	1	164	6891	7054	3.00E-84
NODE_13682426_cov_130.709671	KF241951.1	94.92	59	3	61	7932	1	59	6308	6366	6.00E-23
NODE_14131442_cov_344.593750	KF241951.1	100	94	0	94	7932	1	94	7101	7194	3.00E-47
NODE_14257168_cov_640.108093	KF241951.1	100	104	0	104	7932	1	104	3374	3271	1.00E-52
NODE_14537390_cov_520.944458	KF241951.1	100	66	0	66	7932	1	66	3771	3706	8.00E-32
NODE_14550741_cov_379.786987	NC_026238.1	82.25	169	24	199	7836	12	176	7571	7737	8.00E-37
NODE_14996177_cov_614.648132	KF241951.1	100	138	0	138	7932	1	138	6913	6776	2.00E-71
NODE_16454517_cov_528.312500	KF241951.1	100	62	0	62	7932	1	62	1289	1350	1.00E-29
NODE_17208363_cov_653.435913	KF241951.1	98.55	69	1	69	7932	1	69	5473	5405	8.00E-32
NODE_17246797_cov_258.657135	KF241951.1	100	64	0	65	7932	1	64	2083	2146	1.00E-30
NODE_17246798_cov_258.234039	KF241951.1	98.7	77	1	77	7932	1	77	2118	2194	3.00E-36
NODE_17246799_cov_258.000000	KF241951.1	100	61	0	61	7932	1	61	2165	2225	4.00E-29
NODE_17429278_cov_612.967712	KF241951.1	100	61	0	61	7932	1	61	6125	6185	4.00E-29
NODE_17636147_cov_286.433136	KF241951.1	90.37	187	18	187	7932	1	187	4918	5104	1.00E-68
NODE_17665916_cov_314.419342	KF241951.1	100	61	0	61	7932	1	61	1907	1967	4.00E-29
NODE_17665917_cov_295.047607	KF241951.1	100	93	0	93	7932	1	93	1938	2030	1.00E-46
NODE_17688863_cov_139.492188	KF241951.1	100	158	0	158	7932	1	158	5290	5133	2.00E-82
NODE_18165492_cov_653.875000	KF241951.1	100	62	0	62	7932	1	62	4728	4667	1.00E-29
NODE_18346971_cov_480.000000	KF241951.1	95	60	3	61	7932	2	61	2293	2352	2.00E-23
NODE_18436161_cov_142.767120	KF241951.1	93.2	103	7	103	7932	1	103	2653	2755	2.00E-40
NODE_18824393_cov_759.905640	KF241951.1	100	83	0	83	7932	1	83	1649	1567	4.00E-41
NODE_18831076_cov_333.881989	NC_026238.1	100	21	0	191	7836	60	80	5198	5178	3.00E-06
NODE_18933524_cov_286.000000	KF241951.1	100	61	0	61	7932	1	61	2682	2622	4.00E-29