11-15-2017

Occurrence Of Putative Endornaviruses In Non-Cultivated Plant Species And Characterization Of A Novel Endornavirus In Geranium Carolinianum

Rachel Herschlag
Louisiana State University and Agricultural and Mechanical College, rachelherschlag@gmail.com

OCCURRENCE OF PUTATIVE ENDORNAVIRUSES IN NON-CULTIVATED PLANT SPECIES AND CHARACTERIZATION OF A NOVEL ENDORNAVIRUS IN *GERANIUM CAROLINIANUM*

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
In partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Plant Pathology and Crop Physiology

by
Rachel Herschlag
B.S., Tulane University, 2015
December 2017

To my parents Richard and Susan Herschlag and to my sister Elise Herschlag for their continued support and encouragement. And to my grandmother Marianne Mann, for introducing me to botany, gardening, and an overall love of nature.

ACKNOWLEDGEMENTS

I would like to thank my major advisor Dr. Rodrigo A. Valverde for his guidance and mentorship throughout my course of study, as well as my committee members Dr. Christopher Clark and Dr. Vinson Doyle. I would also like to thank my lab members Cesar Escalante and Surasak Khankhum for helping me in the lab during my time at LSU. Also, thank you to Ricardo Alcala-Briceño for conducting *de novo* genome assembly at University of Florida. Finally, I would like to thank Dr. Raymond Schneider for introducing me to the Department of Plant Pathology and Crop Physiology at LSU, and Dr. Richardson and the LSU AgCenter for providing me with the opportunity to conduct research in plant virology.

TABLE OF CONTENTS

ABSTRACT

Endornaviruses are RNA viruses, which can infect plants yet cause no apparent symptoms. To date, most descriptions of endornaviruses infecting plants have been in cultivated species. A survey for endornaviruses in non-cultivated plants was initiated in 2015 and continued through 2017 in Baton Rouge, Louisiana. Two hundred and seven plant species were tested for distinctive dsRNA profiles by selective extraction and gel electrophoresis, of which seven contained endornavirus-like dsRNA. RT-PCR amplification of an endornavirus-specific sequence supported the endornavirus nature of six of the seven samples. Of the six host species, one species, *Geranium carolinianum* was confirmed as being infected with a novel endornavirus. The endornavirus in *G. carolinianum* was characterized and named Geranium carolinianum endornavirus 1 (GcEV-1). The genome of GcEV-1 is approximately 14.7 kb and is related to other endornaviruses, some infecting plants and some infecting fungi. GcEV-1 is a unique plant endornavirus containing genes closely related to fungal and bacterial genes. A GcEV-1 seed transmission test conducted in the greenhouse resulted in a 100% transmission rate.

The occurrence of endornavirus-like dsRNA within *G. carolinianum* was evaluated at three different locations in Louisiana, two within Baton Rouge and one in Belle Chasse. Among the 184 individual plants tested, three individuals were dsRNA-free. There were no clear phenotypic differences in dsRNA-free individuals compared to those containing dsRNA. All three endornavirus-free *G. carolinianum* plants were collected from the same location. The discovery of only six putative endornaviruses after testing 207 plant species suggests that endornaviruses are not very common in non-cultivated plant species. The results of this study provide a foundation for future research investigating the origin of endornaviruses and the effect endornaviruses have on plants.

Chapter I. LITERATURE REVIEW

1.1 A Brief History of Plant Virology

Plant virology is a relatively new field. Although the first plant virus symptoms were described in 750 A.D. by a Japanese empress describing leaf yellowing in a *Eupatorium* species, modern plant virology was not born until the late nineteenth century (van der Want and Dijkstra 2006). Two scientists are credited with the birth of plant virology—Dmitri Ivanovsky and Martinus Beijerinck (van der want and Dijkstra 2006; Hull 2013). In the late 1800s, many tobacco fields were plagued with a mosaic disease, however the responsible pathogen was unknown (Mayer 1886). In 1892, Ivanovsky demonstrated that the pathogen responsible was smaller than bacteria, filtering inoculum through a filter that would not allow bacteria to pass (van der Want 2006; Ivanovski 1892). Ivanovsky hypothesized that the pathogen was possibly a very small bacterium or a substance excreted by bacteria (van der Want 2006; Ivanovski 1892). A few years later, in 1898, Beijerinck found that the pathogen responsible for mosaic disease was different from other microorganisms, calling it "*contagium vivum fluidum*" (van der Want and Dijkstra 2006). Soon after, the word "virus" was adopted as the preferred term.

With the description of tobacco mosaic virus (TMV), the "biological age" of plant virology began, spanning from roughly 1900-1935 (Hull 2013). During that time, plant virologists were mostly focused on describing new plant viruses (Hull 2013). The biological age was largely defined by Francis Holmes, who in 1929 conducted mechanical inoculations of TMV that resulted in the formation of local lesions on tobacco plants (Hull 2013).

The biological age was followed by the "biochemical/physical age," which began in 1935 with the crystallization of purified TMV by Wendell Stanley (Hull 2013; Roossinck 2016). Purification of a plant virus not only paved the way for purification of other viruses, but also

provided evidence that viruses are different from bacteria and fungi, due to viruses forming crystal structures, a characteristic associated with chemicals rather than microscopic organisms (Roossinck 2016). Stanley also demonstrated that TMV consisted of proteins and RNA (Roossinck 2016). Not long after Stanley's experiments, Bernal and Fankuchen used X-ray analysis to determine the shape and size of TMV (Hull 2013; Bernal and Fankuchen 1937).

The molecular age of plant virology began in 1960, as molecular techniques were developed and used to diagnose plant virus diseases and study plant viruses (Hull 2013). It was during the molecular age that the amino acid sequence in the coat protein of TMV was determined, as well as how viruses replicate (Hull 2013). Polymerase chain reaction (PCR) was developed in 1983, which allowed for the molecular detection of plant viruses (Roossinck 2016).

Advances in diagnostic techniques and sequencing paved the way for the current age in plant virology, referred to as the viromics age (Hull 2013). The viromics age is defined as an age where detailed interactions between virus, plant hosts and invertebrate vectors are being studied (Hull 2013). Popular topics of the viromics age include how viruses cause disease, and how plant hosts defend against virus infection (Hull 2013). Metagenomic studies, which analyze genetic material extracted from organisms, communities or the environment are increasingly popular (Hull 2013). Plant virus ecology is another emerging field in the viromics age, which aims to uncover the principles behind interactions between plants, viruses and vectors, assess the genetic and ecological characteristics of both established and novel plant viruses, and evaluate the effect of plant virus dynamics on ecosystems (Malmstrom et al. 2011). Studies of complex interactions between viruses, their plant hosts and vectors includes a growing interest in the distribution of persistent viruses in plant hosts, and the effect persistent viruses have on plants.

1.2 Persistent and Acute Viruses

Persistent plant viruses are defined as viruses that generally do not cause symptoms in their plant hosts and do not move from cell to cell, but are found in every cell including the meristem (Roossinck 2010). In addition to infecting plants, persistent viruses have also been described in fungal and oomycete species (Roossinck 2010; Osaki et al. 2006; Stielow et al. 2011; Shang et al. 2015; Hacker et al. 2005). There are currently five accepted families of persistent plant viruses: *Amalgaviridae*, *Chrysoviridae*, *Endornaviridae*, *Partitiviridae*, and *Totiviridae* (ICTV 2017).

Acute plant viruses include well-studied viruses such as TMV, cucumber mosaic virus (CMV), and barley yellow dwarf virus (BYDV). In contrast to persistent plant viruses, acute plant viruses generally cause symptoms in their plant hosts, move from cell to cell, and are usually transmitted horizontally, although some viruses can be transmitted vertically (Roossinck 2010). Because acute viruses can be horizontally transmitted in most cases, many acute viruses have the ability to infect more than one plant species, whereas persistent viruses tend to be host-specific (Roosinck 2010).

1.3 Endornaviruses

Within the persistent viruses is the family *Endornaviridae*, genus *Endornavirus,* which includes all endornaviruses (King et al. 2011). Endornaviruses differ from other persistent viruses in that they lack both coat and movement proteins, being comprised solely of naked RNA (Roossinck et al. 2011). Currently, endornaviruses are classified into two genera, *Alphaendornavirus,* which includes viruses that infect plants, fungi, and oomycetes, and *Betaendornavirus,* which includes viruses of ascomycete fungi (Adams et al. 2017). Endornavirus genomes are relatively large in size, ranging from 9.8 kb to 20.3 kb, however endornavirus species infecting crops are

approximately 13-17 kb in size (Fukuhara et al. 2006). Like other persistent viruses, endornaviruses are transmitted vertically to progeny at a very high rate, nearly 100% when measured in infected common bean, as well as in rice and bell pepper (Moriyama et al. 1996; Valverde and Gutierrez 2007; Okada et al. 2013). Endornaviruses have a single open reading frame with a nick in the positive-sense strand of the replicative form (dsRNA; Roossinck et al. 2011).

Most endornaviruses have been described in crop species. To date, endornaviruses have been described in approximately eleven crop species, with some species infected with more than one endornavirus (Khankhum et al. 2015). Crops infected with endornaviruses include avocado (Villanueva et al. 2012), barley (Candresse et al. 2016), broad bean (Pfeiffer 1998), common bean (Okada et al. 2013), cucurbits (Kwon et al. 2014; Sabanadzovic *et al*., 2016), bottle gourd (Kwon et al. 2014), pepper (Okada et al. 2011), rice (Fukuhara 1999), and spinach (Okada et al. 2014). In most cases, only some cultivars of these crops have been shown to be endornavirus-infected. Nevertheless, in the United States, infection rates of endornaviruses in commercial cultivars of bell pepper and melon have been reported to be nearly 100% (Okada et al. 2011; Sabanadzovic et al. 2016; Valverde et al. 1990).

In addition to being found in several plant species, endornaviruses have also been described in several species of oomycetes and fungi. Infected fungi include *Alternaria brassicola*, *Helicobasidium mompa*, and *Tuber aestivum* (Osaki et al. 2006; Stielow et al. 2011; Shang et al. 2015). Endornaviruses infecting oomycete species include Phytophthora endornavirus-1, which was found in a *Phytophthora* isolate collected from Douglas fir (Hacker et al. 2005). Endornaviruses have not yet been described infecting bacteria, although two endornaviruses, bell pepper endornavirus (BPEV) and Oryza sativa endornavirus (OsEV) share

genes with marine bacteria (Song et al. 2013). It is thought that the genes shared with bacteria in BPEV and OSeV were acquired from marine bacteria as the result of horizontal gene transfer (Song et al. 2013).

1.4 Molecular Properties of Plant Endornaviruses

Although Vicia faba endornavirus (VfEV) dsRNA has been found to be associated with membranous vesicles in the cytoplasm, endornaviruses are not associated with virus-like particles and therefore do not have a coat protein. Endornaviruses encode a single polypeptide, which is presumed to be processed by virus-encoded proteases. Based on conserved domain database comparison, the genome of all completely sequenced endornaviruses contains conserved motifs of an RNA-dependent RNA polymerase (RdRp) similar to the alpha-like virus superfamily of positive-stranded RNA viruses, although other domains are not conserved and have various origins (Roossinck et al. 2011). Evolution of endornaviruses appears to be congruent with the host group only in the short term but not in the long term (Roossinck et al. 2011). Moreover, some plant endornaviruses are more closely related to fungal endornaviruses than their plant counterparts. It is possible that endornavirus infection in some plant hosts may increase tolerance of environmental stressors (Roossinck et al. 2011).

In addition to the RdRp, the polyprotein of some plant endornaviruses contains conserved motifs of putative viral methyltransferase (MTR), helicase 1 (Hel-1), capsular polysaccharide synthase, and UDP-glycosyltransferase (Fukuhara et al. 2006; Okada et al. 2011 and 2013; Sabanadzovic et al. 2016). Moreover, it has been shown that some plant endornaviruses contain a discontinuity near the 5'end in the plus strand of the replicative form (Okada et al. 2011; Okada et al. 2013; Okada et al. 2014). The function of the nick is unknown but it is thought to be involved in virus replication (Horiuchi and Fukuhara 2004).

1.5 Detection of Endornaviruses

Large dsRNAs, also called high molecular weight dsRNA (>1 kb) has been recognized as genetic material in many plant, animal, fungal, and bacterial viruses (Libonati et al. 1980). In most plants and fungi infected with RNA viruses, dsRNAs can be found most commonly as genomic segments of dsRNA viruses or replicative forms of single-stranded RNA viruses (Buck 1999; Derrick 1978; Dodds et al. 1984; Morris and Dodds 1979).The extraction and electrophoretic analyses of high molecular weight dsRNA from plants is a technique that has been shown to be reliable to detect RNA viral infections in plants (Morris and Dodds 1979; Khankhum et al. 2017; Valverde et al. 1986; 1990; Bar-Joseph et al. 1993; Tzanetakis and Martin 2008; Dodds et al. 1984).

Because of the lack of coat protein, detection of endornaviruses relies mainly on the properties of the viral RNA. Plant endornaviruses reported to date contain a single RNA genome that ranges from 13- 17 kb in size (Fukuhara and Gibbs 2012). Indirect evidence suggests that the genome consists of ssRNA. However, the replicative form (dsRNA) of the genome is the most commonly detected and used for identification purposes. Detection of endornavirus dsRNA has often been conducted by dsRNA extraction and electrophoresis (Fukuhara et al. 2006; Okada et al. 2011, 2013, 2014, 2017; Sabanadzovic et al. 2016; Valverde and Gutierrez 2007). Alternatively, detection of endornavirus ssRNA can be achieved by reverse transcription PCR using endornavirus-specific or degenerate primers (Okada et al. 2011, 2012; Sabanadzovic et al. 2016).

1.6 Endornavirus Interactions with Plants

Because endornaviruses do not cause any apparent symptoms in their hosts, determining the interaction between endornaviruses and their plant hosts has been a priority in endornavirus

research. It is thought that endornaviruses interact with their plant hosts in one of three ways. The first possibility is that endornaviruses, like acute viruses, are parasitic, with the rationale being that all viruses must use host resources for replication. The second possibility is that the interaction between endornaviruses and their host is mutualistic, the reasoning being that if endornaviruses are maintained at a high rate from parent to progeny, there must be selection for endornavirus infection. The third possibility is that the interaction between endornaviruses and their host is commensalistic, meaning that the endornavirus benefits from the host, but the host is not affected by the presence of the virus, whether positively or negatively. Assuming commensalism can be problematic because it is often the default interaction when no clear mutualistic or parasitic relationship is observed (Zapalski 2011). As a result, the chances of type II error increase, as there may be an effect on the host, whether positive or negative, but the interaction needs to be more closely observed (Zapalski 2011). Although "endophyte" typically refers to fungi or bacteria that may not cause symptoms in their hosts, rather than viruses that cause no symptoms in their hosts, there are an increasing number of studies on endophytic plant fungi once thought to have no effect on their plant host having either an antagonistic or facilitative effect (Busby et al. 2016). Studies demonstrating endophytic fungi actually benefitting or harming plant hosts suggests that upon further observation, plant viruses thought to have zero effect on plant hosts may actually be mutualists or parasites.

Several studies have already demonstrated that plant viruses can affect hosts in ways more complex than a typical host-parasite interaction. A 2008 study by Xu et al. observed the effect acute virus infection has on a plant's response to abiotic stress (Xu et al. 2008). After infecting several plant species with CMV, Xu et al. observed increased drought tolerance in beets (*Beta vulgaris*), pepper (*Capsicum annuum*), watermelon (*Citrullus lanatus*), cucumber

(*Cucumis sativus*), zucchini (*Cucurbita pepo*), and tomato (*Solanum lycopersicum*; Xu et al. 2008). A more in-depth study on beets infected with CMV found that infected beets had higher drought recovery rate compared to mock-inoculated beets, as well as an increased recovery rate from cold stress and a higher average water content (Xu et al. 2008). It was also found that rice infected with brome mosaic virus had a higher recovery rate than mock-inoculated rice (Xu et al. 2008).

Persistent viruses have also been demonstrated as having complex interactions with plant hosts. In persistent viruses other than endornaviruses, yield, gene regulation, and thermal tolerance have been investigated. In 1994, Xie et al. associated beet cryptic virus 1 (BCV1) or beet cryptic virus 2 (BCV2) infection with reduced root yield of up to 17% or 21%, respectively (Xiet et al. 1994). With co-infection of BCV1 and BCV2 root yield decreased by up to 23% (Xie et al. 1994). Another persistent virus, white clover cryptic virus 1 (WCCV1), may indirectly play a role in regulation of root nodulation in white clover, although the exact mechanism is unclear (Nakatsukasa-Akune et al. 2005). A study by Nakatsukasa-Akune et al. demonstrated that white clover produces a gene TrEnodDR1 that encodes the coat protein of WCCV1, and that the artificial expression of TrEnodDR1 suppresses nodulation formation (Nakatsukasa-Akune et al. 2005). More recently, a mycovirus of a fungus infecting panic grass growing in geothermal soils at Yellowstone National Park suggested that the both the virus and fungus are required for thermal tolerance (Marquez et al. 2007).

Few studies have been published on the interaction between endornaviruses and their plant hosts. In 1981, male cytoplasmic sterility in broad bean was associated with the presence of double-stranded RNA, later determined to be Vicia faba endornavirus (Grill and

Garger 1981; Pfeiffer 1998). More recently, co-infection of Phaseolus vulgaris endornavirus 1 and Phaseolus vulgaris endornavirus 2 in common bean has been associated with an increased yield and faster germination (Khankhum, 2016). Somewhat contrastingly, infection of bell pepper endornavirus in bell pepper has been associated with a decrease in yield (measured by fruit mass) and a decrease in percent germination Escalante et al. 2016). Finally, although not a virus-plant interaction, an endornavirus infecting the fungus *Helicobasidium mompa* was associated with hypovirulence in infected strains (Osaki et al. 2006). Stobbe and Roossinck 2014 have suggested that endornavirus-plant interactions are thought to be mutualistic, but note that there is currently no definitive evidence, as endornavirus research is a relatively new field.

The few studies on endornavirus-plant interactions have been limited to crop species, with the interaction between endornaviruses and non-cultivated plant species being completely unknown. The only non-cultivated plant species confirmed as being infected with an endornavirus are wild common bean (*Phaseolus vulgaris*) and wild rice (*Oryza rufipogon*; Moriyama et al. 1999; Khankhum et al. 2015). Endornavirus-like dsRNA has also been isolated from eelgrass (*Zostera marina*) and Korean dandelion (*Taraxacum platycarpum*), although the endornavirus nature of these dsRNAs has not been confirmed (Fukuhara et al. 2006). It is unknown if the lack of endornaviruses described in non-cultivated plant species is due to the possibility that endornaviruses are uncommon in non-cultivated plant species, or if the lack of endornavirus descriptions is a result of endornaviruses in non-cultivated plant species being understudied compared to those of crop species.

1.7 Endornavirus in Non-Cultivated Plants

Based on the presence of viruses both acute and persistent in cultivated plant species, it seems likely that endornaviruses are present in more non-cultivated plant species than described. A 2009 survey for plant viruses in the Tallgrass Prairie Preserve of Northeastern Oklahoma screened for virus-like sequences in fifty-two plant species and found virus-like sequences in nineteen percent of the plant species sampled (Muthukumar et al. 2009). Metagenomics, which uses the sequence analysis of environmental samples containing an unknown mixture of diverse microbes, including those that cannot be cultured, is beginning to be used for the detection of persistent viruses in non-cultivated plants (Roossinck 2012). Preliminary data show that endornaviruses, as well as other persistent viruses are fairly common in wild plant species, although only limited formal data have been published (Roossinck 2012; 2017). More specifically, it is unknown if endornaviruses are present in more than a few non-cultivated plant species, including many of the non-cultivated plant species most closely related to crop species infected with endornavirus. Plant viral ecologists have concluded that "the full extent of plant-virus interactions cannot be fully studied until we have a better understanding of the ecology of plant viruses," which includes understanding endornavirus-plant interactions (Stobbe and Roossinck 2014).

Understanding endornavirus-plant interactions first requires determining how commonly endornavirus infection occurs in non-cultivated plant species. However, only one detailed survey has yet been published on the occurrence of endornaviruses in non-cultivated plant species (Thapa et al. 2015). Furthermore, investigations into interactions between endornaviruses and plant hosts will not only require detailed surveys of non-cultivated plant species for endornaviruses, but also surveys for endornaviruses or endornavirus-like dsRNA in individual

plants within a single plant species, in order to compare infected individual plants with other individuals that are endornavirus-free. As found in common bean (*P. vulgaris*) from the Andean region and the Mesoamerican region, not every non-cultivated individual within *P. vulgaris* was infected with endornavirus, and infection appeared to be somewhat location dependent, with a higher percentage of non-cultivated Mesoamerican common bean infected compared to common bean from the Andean region (Khankhum et al. 2015). It is presently unknown if endornavirus infection varies by location in all non-cultivated plant host species, as well as how much distance is required between locations in order to see a difference in the percentage of plants infected. Determining how location plays a role in endornavirus infection in non-cultivated plant species requires the description of more non-cultivated plant species infected with endornaviruses. Novel endornaviruses in non-cultivated plant species may be discovered as more surveys for persistent viruses in non-cultivated plant species are completed.

1.7 Objectives of the Investigation

As previously described, there is a major lack of knowledge of the occurrence of endornaviruses in non-cultivated plants species, and the interaction between non-cultivated plant species and endornaviruses. The lack of understanding of endornaviruses in non-cultivated plant species highlights a need for surveys of non-cultivated plant species and characterization of novel endornaviruses found in non-cultivated plant species. Therefore, the objectives of this investigation were:

a. Survey non-cultivated plant species for presence of endornaviruses
b. Characterize a novel endornavirus of *G. carolinianum*
c. Determine the occurrence of endornavirus-like dsRNA in *G. carolinianum* at three distinct locations

CHAPTER II. SURVEY FOR THE OCCURRENCE OF PUTATIVE ENDORNAVIRUSES IN NON-CULTIVATED PLANT SPECIES

2.1 Introduction

The association between endornaviruses and plant hosts is thought to have pre-dated the advent of agriculture, due to endornaviruses being vertically transmitted over many generations from parent to progeny. Tracing vertical transmission back over many generations, endornaviruses would have been present in the non-cultivated ancestors of crop species, assuming that endornaviruses could not be horizontally transmitted at any point in the past ten thousand years or so of plant cultivation. A long-term association with the host is also thought to be the case for plasmids, which have been compared to endornaviruses by some, due to both endornaviruses and plasmids being un-encapsidated genetic material separate from host chromosomal DNA (Kado 1998, Fukuhara et al. 2006). It is thought that much like plasmids, endornaviruses have developed an association with their hosts over evolutionary time, rather than within the past few thousand years (Kado 1998; Fukuhara et al. 2006).

Fukuhara et al. (2006), showed that the phylogeny of several endornaviruses does not mirror the phylogeny of their plant hosts. Endornaviruses of broad bean and kidney bean are not grouped together, although both hosts belong to Fabaceae (Fukuhara et al. 2006). Additionally, endornaviruses of monocots do not form a monophyletic group (Fukuhara et al. 2006). Since endornaviruses can only be transmitted vertically, researchers hypothesize the ancestors of endornaviruses at one point had the ability to be horizontally transmitted (Fukuhara et al. 2006). Endornaviruses may have been originally horizontally transmitted to plants via fungi, supported by descriptions of several mycovirus members of *Endornaviridae* (Osaki et al. 2006; Khalifa et al. 2014; Shang et al. 2015). Research showing evidence of horizontal gene transfer between *Endornavirus* and marine bacteria also suggest that ancient endornaviruses may have infected

marine algae, and co-evolved with their hosts, infecting land plants during the evolution of higher plants (Song et al. 2013; Sabanadzovic et al. 2016).

If the association between endornaviruses and crop species does in fact pre-date agriculture, endornavirus infection would be expected in non-cultivated plants as well as crops of the same species. To address questions regarding endornavirus infection in plants at different stages of domestication, as well as broader questions regarding patterns of infection with respect to centers of domestication, Khankhum et al. (2015) tested common bean cultivars, breeding lines, landraces and wild *P.* vulgaris for the presence of Phaseolus vulgaris endornavirus 1 (PvEV-1) and Phaseolus vulgaris endornavirus 2 (PvEV-2) from the Mesoamerican region and the Andes region (Khankhum et al. 2015). Wild *P. vulgaris* from Mesoamerica was infected, however wild beans from the Andes were not, which was attributed to Mesoamerican *P. vulgaris* being the original source of PvEV-1 and PvEV-2 (Khankhum et al. 2015).

Another discovery was that in tracing percent infection from wild common bean to landraces to cultivars and finally breeding lines, infection increased significantly in Mesoamerican beans, but only very slightly in Andean beans, suggesting that endornavirus may have been more beneficial to the host in Mesoamerica than the Andes region (Khankhum et al. 2015). Although it is impossible to know what early domesticators were selecting for when taking *P. vulgaris* seeds from the wild, it does potentially suggest that endornavirus-infected *P. vulgaris* was being selected for in the Mesoamerican region, possibly due to traits that may be more beneficial in the Mesoamerican climate, compared to the Andean region.

With a limited number of surveys for the presence of endornaviruses in non-cultivated plant species, and with so few endornavirus-infected non-cultivated plant species to study, the interaction between endornaviruses and their plant hosts, especially between plant hosts under

natural selection, remains unknown. Ideally, all non-cultivated plant species need to be tested for endornavirus, starting by testing within specific locations.

2.2 Objectives

In 2015, a survey was initiated to determine the occurrence of endornaviruses in non-cultivated plant species in the city of Baton Rouge, Louisiana. The survey was continued in 2016 and 2017 and the overall findings reported in this chapter.

2.3 Materials and Methods

The extraction and electrophoretic analysis of viral dsRNA technique has been valuable in the initial stages of the discovery of most plant endornaviruses reported to date (Valverde et al. 1990; Fukuhara 1999; Fukuhara et al. 2006; Okada et al. 2011; 2013, 2015, 2017; Sabanadzovic et al. 2016). Therefore, it was used as a primary tool to detect putative endornaviruses in this investigation.

2.3.1 Selection of Survey Location

The city limits of Baton Rouge, Louisiana were chosen as the survey area for endornaviruses in non-cultivated plant species. Being located in a transition weather zone 8b, tropical and subtropical plants often grow in many locations within the city limits. They include a variety of non-cultivated, native and introduced as well as invasive plant species. East Baton Rouge Parish is estimated to have over 1600 plant species, which was used as a reference to determine how representative sampling of plant species was of the total number of plant species within the area (Thomas and Allen 1993).

2.3.2 Collection of Plant Species Samples

Non-cultivated plant species were collected and tested for the presence of endornavirus-like dsRNA by gel electrophoresis. Locations included home gardens, wetlands, undeveloped land,

parks, Louisiana State University campus, LSU Agricultural Center Botanic Gardens, Louisiana State University Central Research Station at Ben Hur, as well as roadsides throughout the city. When feasible, multiple individuals were collected of each plant species at each location.

Each plant sample collected was identified using The Manual of the Vascular Flora of the Carolinas by Radford et al. 2010, the USDA Plants Database, and The Atlas of the Vascular Flora of Louisiana by Thomas and Allen, the Shirley C. Tucker Herbarium at Louisiana State University, and the Louisiana Plant Identification and Interactive Ecosystem Virtual Tours (rnr.lsu.edu/plantid/default.htm). The origin of each plant species, whether introduced or native to Louisiana, was also recorded.

2.3.3 Testing Plant Species for Endornavirus-like dsRNA

The presence or absence of endornavirus-like dsRNA (dsRNAs of approximately 13-17 kb) in non-cultivated plant species was determined by electrophoretic analyses of extracted dsRNAs reported by Khankhum et al. 2017). Briefly, foliar tissue was finely chopped and desiccated in silica gel at 4°C overnight. Tissue was finely ground in a mortar and pestle and 0.07g used for dsRNA extraction. DsRNA was phenol-extracted and purified using fibrous cellulose (Sigma-Aldrich, St. Louis, MO). DsRNA was ethanol precipitated, suspended in nuclease-free water and treated with of RNase-free DNase I (Fisher Scientific, Waltham, MA). The presence or absence of dsRNA was determined in 1.2% agarose gel at 70 V for 2 h. Tissues from plants known to have endornavirus or be endornavirus-free were used for positive and negative controls, respectively.

2.3.4 Reverse Transcription PCR

To further investigate the possible endornavirus nature of samples containing endornavirus-like dsRNA, samples were tested by reverse transcription (RT)-PCR using degenerate endornavirus

primers. Total RNA was extracted from plants consistently showing large dsRNAs using the Plant Total RNA Kit (Spectrum ^{TM}, Sigma-Aldrich). To determine the RNA concentration of the samples (ng/ µl), 2µl of total RNA were measured in a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE). Extracted RNA samples were stored at -70°C for RT-PCR analysis. Alternatively, endornavirus-like dsRNAs were used as templates after denaturation for 5 min at 95°C. RNA was amplified in RT-PCR reactions using Super-Script One-Step RT-PCR with Platinum Taq. cDNA amplification consisted of 50°C for 30 min and 94°C for 2 minutes. A pair of degenerate endornavirus primers, endo-F (5'AAGSGAGAATWATHGTRTGGCA 3'), and endo-R (5' CTAGWGCKGTBGTAGCTTGWCC 3'), designed to amplify a 381-nucleotide (nt) region of the RdRp of plant endornaviruses were used (Valverde et al. 2011).

2.4 Results

2.4.1 Collection of Plant Species Samples

The three-year survey identified 207 plant species, 197 of which were to the species level, and ten to the genus level (Appendix 1). One hundred twenty plant species were native to Louisiana, eighty-three were introduced and four were not determined (Appendix 1). Seventy-eight plant families were represented (Appendix 1).

Seven plant species contained endornavirus-like dsRNA and three plant species contained other, smaller dsRNAs that may represent the genome of other persistent viruses (Fig. 2.1). DsRNA extractions from all plant species with endornavirus-like dsRNA were repeated and the presence of dsRNA confirmed. Similarly, repeated extractions of plant species lacking dsRNA did not yield dsRNAs. Plant species with endornavirus-like dsRNA were: *Alternanthera philoxeroides* (alligator weed), *Dracopis amplexicaulis* (clasping cone flower), *Geranium*

carolinianum (Carolina geranium), *Hydrocotyle umbellata* (dollar weed), *Hydrocotyle prolifera* (whorled pennywort), *Sonchus asper* (sow thistle) and *Sorghum halepense* (Johnson grass) (Fig. 2.2; Table 2.1; Appendix 2). Plant species that contained dsRNA smaller than endorna-like dsRNA included, *D. amplexicaulis, Erythrina herbacea* (coral bean), *H. prolifera*, and *Phyla lanceolata* (lanceleaf fogfruit); Table 2.1; Fig. 2.1. To confirm the results, plants species that yielded dsRNAs were sampled again from the same original locations and GPS coordinates recorded (Appendix 2).

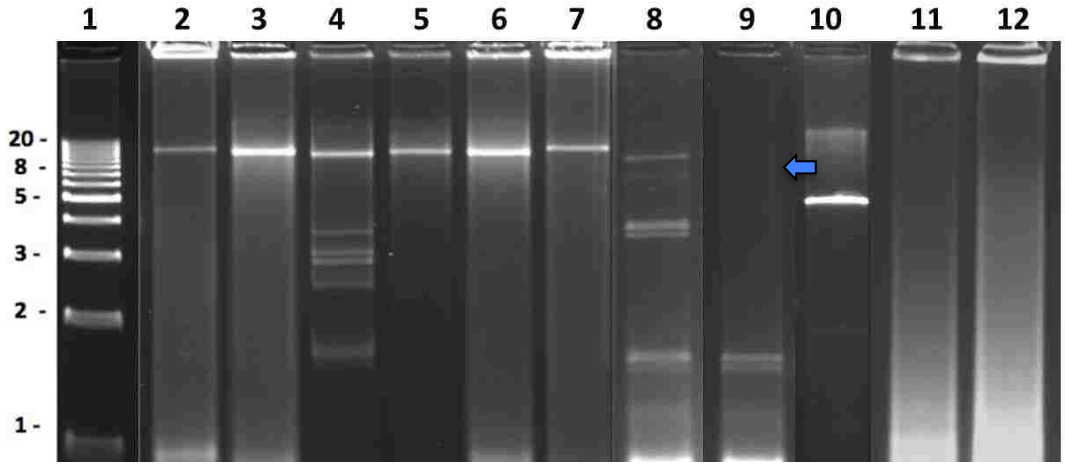

Figure 2.1 Composite illustration of agarose gel electrophoresis (1.2%) showing endornavirus-like dsRNAs detected in non-cultivated plants. 1, 1kb ladder; 2, *S. halepense*; 3, *G. carolinianum*; 4, *H. prolifera*; 5, *A. philoxeroides*; 6, *H. umbellata*; 7, *S. asper*; 8, *D. amplexicaulis*; 9, *E. herbacea*; 10, *P. lanceolata*; and 11 and 12, dsRNA typical results of dsRNA-negative plants. Lanes 4, 8, 9, and 10 contain dsRNAs of smaller size than edornavirus-like dsRNA. Arrow points at the endornavirus-like dsRNAs.

Table 2.1 Plant species infected with putative endornaviruses

Plant Species	Common Name	Family	Origin
Alternanthera philoxeroides	Alligator weed	Amaranthaceae	Introduced
Dracopis amplexicaulis	Clasping coneflower	Asteraceae	Introduced
Geranium carolinianum	Carolina geranium	Geraniaceae	Native
Hydrocotyle prolifera	Whorled pennywort	Araliaceae	Native
Hydrocotyle umbellata	Dollar weed	Araliaceae	Native
Sonchus asper	Sow thistle	Asteraceae	Introduced
Sorghum halepense	Johnson grass	Poaceae	Introduced

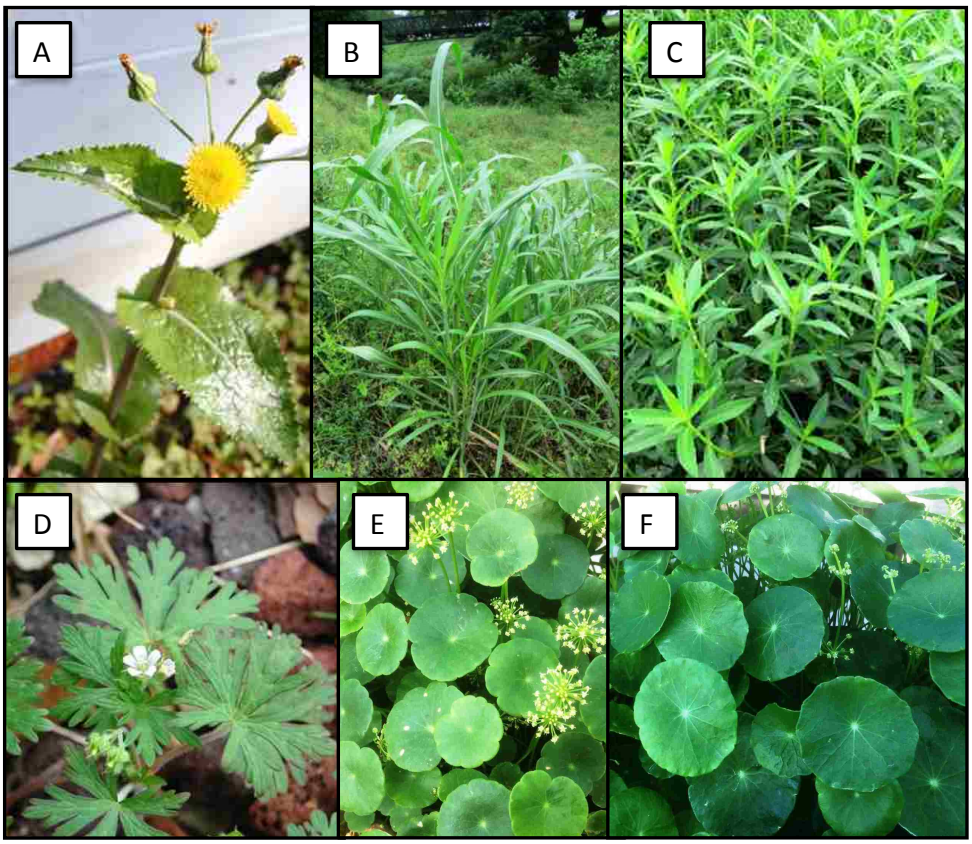

Figure 2.2 Plant species found infected with putative endornaviruses. A, *S. asper*; B, *S. halepense;* C, *A. philoxeroides*; D, *G. carolinianum*; E, *H. umbellata*; and F, *H. prolifera.*

RT-PCR using degenerate endornavirus primers consistently yielded amplicons with RNA extracted from six of the seven plant species. The amplicons ranged from approximately 380 bp to 700 bp. *Hydrocotyle umbellata,* and *H. prolifera* each yielded two amplicons of 400 and 500 bp, *S. halepense* one of 450 bp, *S. asper* two of 500 bp and 700 bp, *G. carolinianum* one of 400 bp, and *A. philoxeroides* one of 380 bp (Fig. 2.3 and 2.4) *Dracopis amplexicaulis* consistently did not yield amplicons. These results support the endornavirus nature of all endornavirus-like dsRNA except for the endornavirus-like dsRNA extracted from *D. amplexicaulis*.

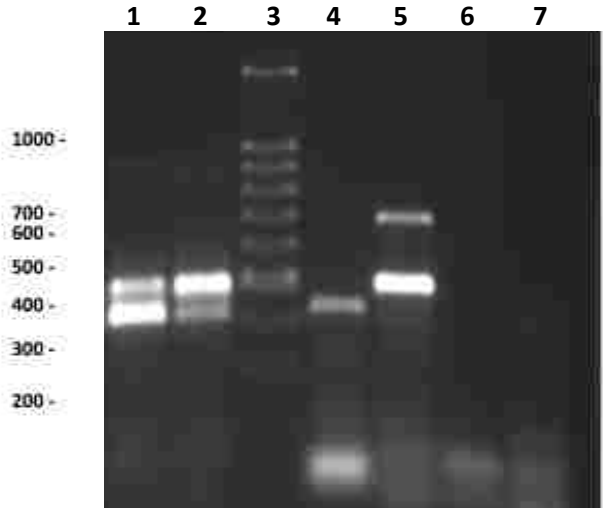

Figure 2.3 Agarose gel electrophoresis (1.2%) showing RT-PCR amplicons obtained from endornavirus-like dsRNA templates extracted from four plant species. 1, *H. umbellata*; 2, *H. prolifera*; 3, 100bp ladder; 4, *S. halepense*; 5, *S. asper*; 6, negative control (*P. vulgaris* cv. Black Turtle Soup, endornavirus-free); and 7, water.

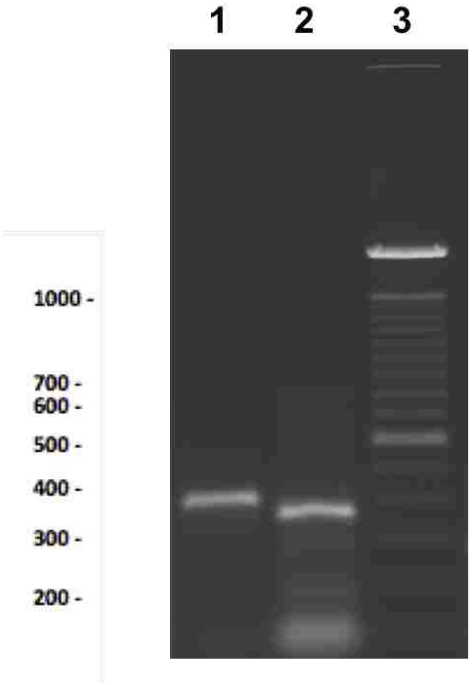

Figure 2.4. Agarose gel electrophoresis (1.2%) showing RT-PCR amplicons obtained from endornavirus-like dsRNAs templates extracted from 1, *G. carolinianum*; 2, *A. philoxeroides*, 3, 100bp ladder

2.5 Discussion

The extraction of dsRNA and subsequent electrophoretic analysis of viral dsRNA has been a reliable method to detect endornavirus-like dsRNAs from many plant species, particularly because of the unique size (13-17 kb) of the dsRNA of plant endornaviruses, and the fact that healthy plants do not contain large molecular weight (larger than 1 kb) dsRNAs. (Fukuhara et al. 2006; Khankhum et al. 2015; Okada et al. 2011).

Previously, it was largely unknown how commonly endornavirus infection occurs in non-cultivated plant species, and more specifically, how many plant species are infected within a given geographical location. The finding of six putative endornavirus-infected plant species in Baton Rouge using both dsRNA extraction and amplification by RT-PCR with degenerate endornavirus primers suggests that there are more endornaviruses of non-cultivated plant species yet to be described.

The seven species with endornavirus-like dsRNA show no pattern of endornavirus infection with respect to habitat, plant family, or whether they are introduced or native species. However, more plant species in more locations will need to be collected in order to determine if there is a potential pattern of infection with respect to such characteristics. As expected, there was no evidence of typical viral symptoms commonly associated with viral infections.

In addition to the endornavirus-infected plant species described, the finding of several plant species with other putative persistent viruses also suggests that other persistent viruses can be found in non-cultivated plant species, in some cases as mixed infection with a putative endornavirus. There was no apparent pattern in species infected with other putative persistent viruses with respect to habitat, family, and whether plants are native or introduced.

Regarding endornavirus infection and its potential impact on the domestication process, more non-cultivated plant species will need to be tested for endornaviruses with special attention to wild relatives or infected crop species as well as non-cultivated plant species in origins of domestication. Because the survey sampled a small subset of all non-cultivated plant species, it cannot yet be determined whether endornaviruses are more or less common in non-cultivated plant species compared to crop species.

The survey for endornavirus in non-cultivated plants of Baton Rouge was a necessary first step in determining how common endornaviruses occur in plants subject to natural selection and minimal human intervention compared to cultivated crops. In addition to finding seven plant species with endornavirus-like dsRNA, the survey also confirmed that plants infected with endornavirus one year remained infected the following year (Appendix 2), providing further evidence that endornavirus infection remains high from one generation to the next. Once surveys of more plant species are conducted, hopefully questions about the effect endornaviruses have on non-cultivated plant species can be addressed. Together with studies of endornaviruses infecting crop species, determining the interactions between endornaviruses and their plants becomes increasingly likely.

CHAPTER III: CHARACTERIZATION OF A NOVEL ENDORNAVIRUS FROM *GERANIUM CAROLINIANUM*

3.1 Introduction

Geranium carolinianum (Geraniaceae) is a common weed native to North America, and is found in nearly every U.S. state (USDA NRCS). Within Louisiana, *G. carolinianum* has been collected in every parish (USDA NRCS). Although the common name for *G. carolinianum* is Carolina geranium, *G. carolinianum* is distinct from ornamental geraniums, which are also members of Geraniaceae, but are in the genus *Pelargonium*. The most closely related genus to *Geranium* is *Erodium* (Price and Palmer 1993).

Carolina geraniums are annuals or biennials that typically grow in cooler weather, and typically bloom between the months of March and July, but can bloom as early as February in Louisiana. Carolina geraniums are typically no taller than 0.5 m and have leaves that are palmately five-parted, with leaf divisions being cleft or lobed (Radford et al. 2010). Flowers are five-petaled and pale pink, with a pistil of five carpels forming a long beak (Radford et al. 2010). Each carpel is single seeded, and at maturity, each carpel separates from the pistil, forming a long tail that aids in dispersing seeds in a catapult-like motion (Fig. 3.1; Radford et al. 2010). *Geranium carolinianum* grows well in disturbed habitats, such as roadsides, lawns, pastures, and near railroad tracks (Baskin and Baskin 1974). *Geranium carolinianum* seeds have physical dormancy, meaning that dormancy is caused by a hard seed coat impermeable to water (Baskin and Baskin 1974).

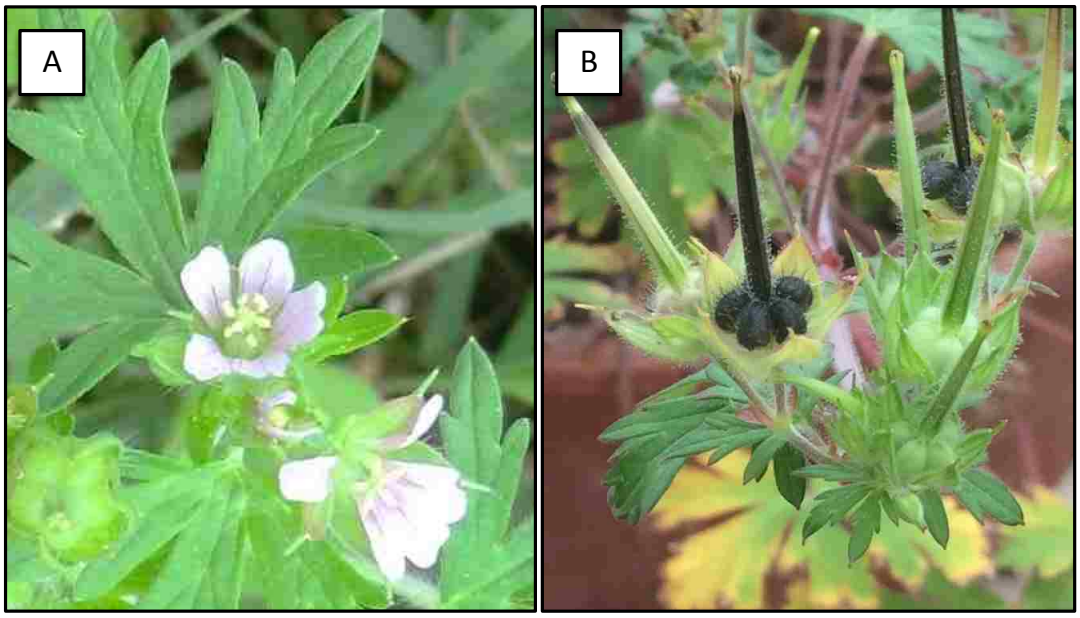

Figure 3.1. Flower morphology of *G. carolinianum*. A) Pinkish white flower typical of *G. carolinianum*. B) Mature flowers with mature black seeds or immature green seeds.

Studies on endornavirus infection in non-cultivated plant species are lacking. Endornaviruses have been described infecting wild rice (*Oryza rufipogon*) and wild common bean (*Phaseolus vulgaris*; Moriyama et al. 1999; Khankhum et al. 2015). Other species have been described as containing endornavirus-like dsRNA, however they have not been confirmed as being infected with endornavirus. Beyond the two non-cultivated plant species infected with endornavirus, little is known about endornaviruses in non-cultivated plant species.

Of the six plant species in Louisiana with putative endornaviruses (see previous chapter), all putative endornaviruses will need to be characterized if they are in fact novel endornaviruses. A putative endornavirus in *G. carolinianum* is chosen here as the first of six putative endornaviruses to be characterized because the host *G. carolinianum* is the most common species of the six plant species, and is considered a common weed of Louisiana (Miller 1969). Because *G. carolinianum* is a common weedy species, individuals can be easily sampled from for

endornavirus characterization. Additionally, *G. carolinianum* individuals can be distinguished from one another, compared to other plant species with putative endornaviruses such as *H. prolifera* and *H. umbellata*. *Hydrocotyle prolifera* and *H. umbellata* grow in dense mats and it can be difficult to determine where one individual plant ends and another begins. Additionally, *Hydrocotyle* species can be difficult to identify when not flowering, presenting further challenges in characterizing and sampling endornaviruses in both *Hydrocotyle* species.

The genome of plant endornaviruses typically range from approximately 13-17 kb. (Fukuhara et al. 2006; Okada et al. 2011, Okada et al. 2013). Endornaviruses have a single long open reading frame with a nick in the positive sense strand, and often conserved motifs for methyltransferase, helicase, UDP-glycosyltransferase, RNA-dependent RNA polymerase, and in some cases methyltransferase (Okada et al. 2011; Okada et al. 2013). Some endornaviruses also show evidence of horizontal gene transfer with bacteria (Song et al. 2013). Endornaviruses of plants are not monophyletic, and often share most recent common ancestors with endornaviruses infecting fungi (Fukuhara et al. 2006).

3.2 Objective

The objective of this investigation was to conduct the characterization of a putative endornavirus obtained from *G. carolinianum*.

3.3 Materials and Methods

3.3.1 DsRNA Purification and Sequencing

DsRNA was extracted following the method of Khankhum et al. (2017), separated in 1% agarose gels, gel purified using the Qiagen gel purification kit (Qiagen, Valencia, CA), and used for a library preparation. Sequencing was conducted at the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign by Illumina MiSeq (pair-end 2 x 250). dsRNA was

denatured at 95°C for 5 min and used to prepare RNAseq libraries. The libraries were prepared with Illumina's 'TruSeq RNA Sample Prep kit' with two modifications: RNA was not polyA selected. RNA was randomly primed but not chemically fragmented. The libraries were pooled and the pool was quantitated by qPCR and sequenced on one MiSeqNano flowcell for 251 cycles from each end of the fragments using a MiSeq sequencing kit version 2. Fastq files were generated and demultiplexed with the bcl2fastq v2.17.1.14 Conversion Software (Illumina). Adaptors were trimmed from the reads. Reads were 250 nt (nucleotides) in length. The number of reads was 164,891.

Genome assembly was conducted by a collaborator, Mr. Ricardo Alcala-Briceño, Department of Plant Pathology, University of Florida. The viral genome was assembled via *de novo* assembly using Spades 3.7.1.2 (Bankevich et al. 2012), mapped and reconstructed with Bowtie2, and elongation and redundancy of contigs determined with CAP2. The minimum and maximum length of contigs was 150 and 15,000 nt, respectively.

3.3.2 Sequence Analysis

From the sequence data, a contig of 14,638 nt was obtained. The entire length of the contig (14,638 nt) was translated into protein using the Expasy tool of the Swiss Institute of Bioinformatics (https://web.expasy.org/translate/). The sequence contained a single open reading frame and the conserved protein domains determined using the Conserved Domain Database available in the National Center for Biotechnology Information (NCBI). Blastx was conducted to determine sequence similarities. A BLAST search was conducted using the conserved domains found in GcEV-1, and domains were compared with other conserved domains from endornaviruses found in GenBank. Percentage of amino acid sequence identity was determined and compared among selected endornaviruses.

3.3.3 Construction of a Phylogenetic Tree

The amino acid sequences of several endornavirus genomes were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/; Appendix 3). Sequences, including the sequence of GcEV-1, were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE) in Molecular Evolutionary Genetics Analysis (MEGA) version 7.0 (Edgar 2004; Kumar et al. 2016). The evolutionary history was inferred using the Neighbor-Joining method and the RdRp domain of GcEV-1(Saitou and Nei 1987; Fig. 3.4). The optimal tree with the sum of branch length= 14.82620109. A bootstrap test was used with 500 replicates. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site (Zuckerland and Pauling 1965). The analysis involved twenty-one amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 1405 positions in the final dataset.

3.3.4 Testing Selected Species within the Genus

To determine if endornavirus-like dsRNA was present in foliar samples of other members of the genus *Geranium*, *G. dissectum* (Louisiana)*, G. lucidum* (California and Oregon)*, G. macrorrhizum* (Washington, D.C.) and *G. maculatum* (Maine) were tested using the dsRNA extraction method (Khankhum et al. 2016). Geranium samples were collected from undeveloped land at each location.

3.3.5 Testing the Seed Coat and Progeny for Endornavirus-like dsRNA

Some seeds of *G. carolinianum* plants are often aborted, being comprised solely of the seed coat and lacking an embryo (Fig. 3.2). To determine if the presence of GcEV-1 was potentially required for embryogenesis, aborted seeds were tested for the presence of GcEV-1 via dsRNA

extraction and compared to the presence of dsRNA in fully formed seeds. 0.05g of seed coat tissue or fully formed seed tissue was ground using a mortar and pestle, and dsRNA was extracted following the dsRNA extraction method (Khankhum et al. 2016).

The transmission rate of GcEV-1 to progeny in a greenhouse from one individual parent plant was determined. A *G. carolinianum* plant known to be infected with GcEV-1was collected, planted in a greenhouse, and allowed to set seed. Because seeds are discharged at maturity, a bag was placed over the flowers at maturity, and all seeds from the plant collected. Seeds were removed from mature carpels and planted in potting soil under a 16/8h photoperiod. Plants were harvested just before individuals formed flowers, and had enough tissue to extract dsRNA. DsRNA was extracted using the method developed by Khankhum et al. 2016. The presence of endornavirus-like dsRNA was visualized in 1.2% agarose gel at 70 V for 2 h.

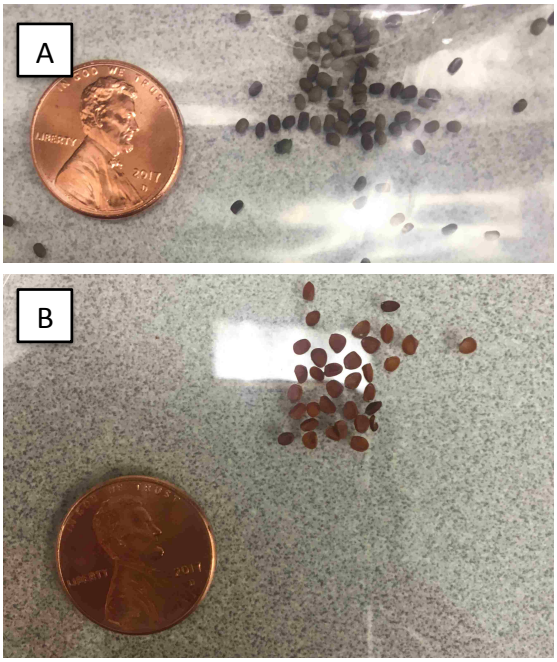

Figure 3.2. *G. carolinianum* seeds. A) Fully formed seeds comprised of seed coat and embryo. B) Aborted seeds comprised of seed coat only.

3.4 Results

3.4.1 DsRNA Purification and Sequencing

DsRNA was readily purified in relatively large quantities from *G. carolinianum* tissues. Sequencing the *G. carolinianum* RNA yielded a contig of 14,638 nt (Appendix 4) containing a single open reading frame that coded for a polyprotein of 4,815 aa (amino acid; Appendix 5). The polyprotein contained conserved domains for RNA-dependent RNA polymerase (RdRp), helicase-1, Peptidase, and a glycosyltransferase (Fig. 3.3). The 5'end consisted of 171 nt while the 3'end of 54 nt ending in 9 Cs, which is typical of several plant endornaviruses (Appendix 4). Other smaller contigs were obtained as well but they were fragments of the large 14,638 nt contig.

Conserved protein domain analysis found four putative conserved domains: viral helicase (Hel), UDP-glycosyltransferase (UDP), peptidase C97 (PEP) and RNA-dependent RNA polymerase (RdRp) (Table 3.1). Positions of the Hel, UDP, and RdRP domains were typical of endornaviruses and the C97 domain unique to GcEV-1.

Table 3.1 Conserved domains of Geranium carolinianum endornavirus 1

Name	Accession	Description	Interval (aa)	E-value
RdRp	cd01699	RNA-dependent RNA polymerase (RdRp)	4499-4633	1.20E-04
Peptidase_C97	pfam05903	Putative peptidase domain; The PPPDE superfamily	2708-2812	1.09E-03
YjiC	COG1819	UDP:flavonoid glycosyltransferase YjiC, YdhE family [Carbohydrate transport and metabolism]	3398-3566	3.57E-12
Viral_helicase1	pfam01443	Viral (Superfamily 1) RNA helicase	1434-1691	2.07E-09

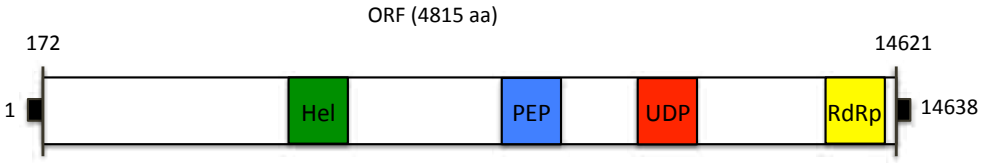

Figure 3.3 Genome organization of a novel endornavirus Geranium carolinianum endornavirus - 1 (GcEV-1): Viral helicase 1 (Hel), Peptidase C97 (PEP) Flavonoid glycosyltransferase (UDP), and RNA dependent RNA Polymerase (RdRp). ORF=open reading frame of the 4,815 polyprotein.

The aminoacid sequence of the RdRp and viral helicase revealed amino acid sequence homology to fungal and plant endornaviruses (Tables 3.2, 3.3 & 3.4). The multiprotein domain YjiC of a glycosyltransferase superfamily had various degrees of identity with bacterial glycosyltransferases, but not with those of endornaviruses (Table 3.3). Peptidase had various degrees of identity with proteins of fungal and algal species (Table 3.4).

Table 3.2 Percent sequence identity of GcEV-1 motifs to those of other endornaviruses. ND= none detected

Virus	**Hel%**	**Pep%**	**RdRp%**	**Polyprotein%**	**Accession No.**
Hordeum vulgare endornavirus	27	ND	32	31	YP_009212849.1
Rhizoctonia cerealis endornavirus 1	24	ND	33	33	YP_008719905.1
Soybean leaf-associated endornavirus 1	ND	ND	28	29	ALM62234.1
Phaseolus vulgaris endornavirus 1	ND	ND	29	29	ATB20096.1
Cucumis melo aendornavirus	ND	ND	31	31	ARI71634.1
Lagenaria siceraria endornavirus-Hubei	ND	ND	31	30	YP_009351891.1
Lagenaria siceraria endornavirus-California	ND	ND	31	30	YP_009010973.1
Phaeolus vulgaris endornavirus 2	ND	ND	30	30	ATB20098.1
Persea americana endornavirus 1	ND	ND	31	30	YP_005086952.1
Grapevine endophyte endornavirus	26	ND	ND	23	YP_007003829.1
Ceratobasidium endornavirus H	29	ND	30	30	AOV81686.1
Erysiphe cichoracearum endornavirus	26	ND	ND	25	YP_009225663.1
Chalara endornavirus CeEV1	24	ND	ND	ND	ADN43901.1
Phytophthora endornavirus 1	24	ND	27	26	YP_241110.1
Ceratobasidium endornavirus D	24	ND	30	30	YP_009310051.1
Yellow head virus	28	ND	ND	ND	ACU52714.1

Table 3.3 Percent sequence identity of GcEV-1 UDP-glucuronosyltransferase to related UDPs of bacteria species

Description	Identity (%)	Accession No.
Thalassiosira oceanica	29	EJK74790.1
Mycobacterium rhodesiae	27	WP_014209917.1
Rhodococcus kroppenstedtii	25	WP_068366619.1
Arthrobacter sp. Br18	31	WP_051476964.1
Amycolatopsis pretoriensis	30	SEF21459.1
Rhodococcus sp. PBTS 1	25	WP_068101951.1
Enterobacter aerogenes	32	WP_043865424.1
Atlantibacter hermannii	32	WP_002437160.1
Microbacterium sp. SCN 70-18	41	ODT11531.1

Table 3.4 Percent sequence identity of GcEV-1 Peptidase C97 to related peptidases of fungi and algae species

Description	Identity (%)	Accession No.
Chlorella variabilis	36	XP_005846835.1
Scleroderma citrinum	32	KIM66795.1
Lichtheimia ramosa	35	CDS12093.1
Macrophomina phaseolina MS6	33	EKG17000.1
Diplodia seriata	34	KKY23539.1
Exophiala dermatitidis	36	XP_009153741.1
Pestalotiopsis fici	37	XP_007834194.1
Phialophora attae	33	XP_017996049.1
Moniliophthora roreri	35	KTB44590.1

3.4.3 Phylogenetic Tree

The phylogenetic tree revealed that GcEV-1 was not very closely related to any of the endornaviruses selected for analysis, and shared a most recent common ancestor with endornaviruses infecting both plant and fungal species (Fig. 3.4).

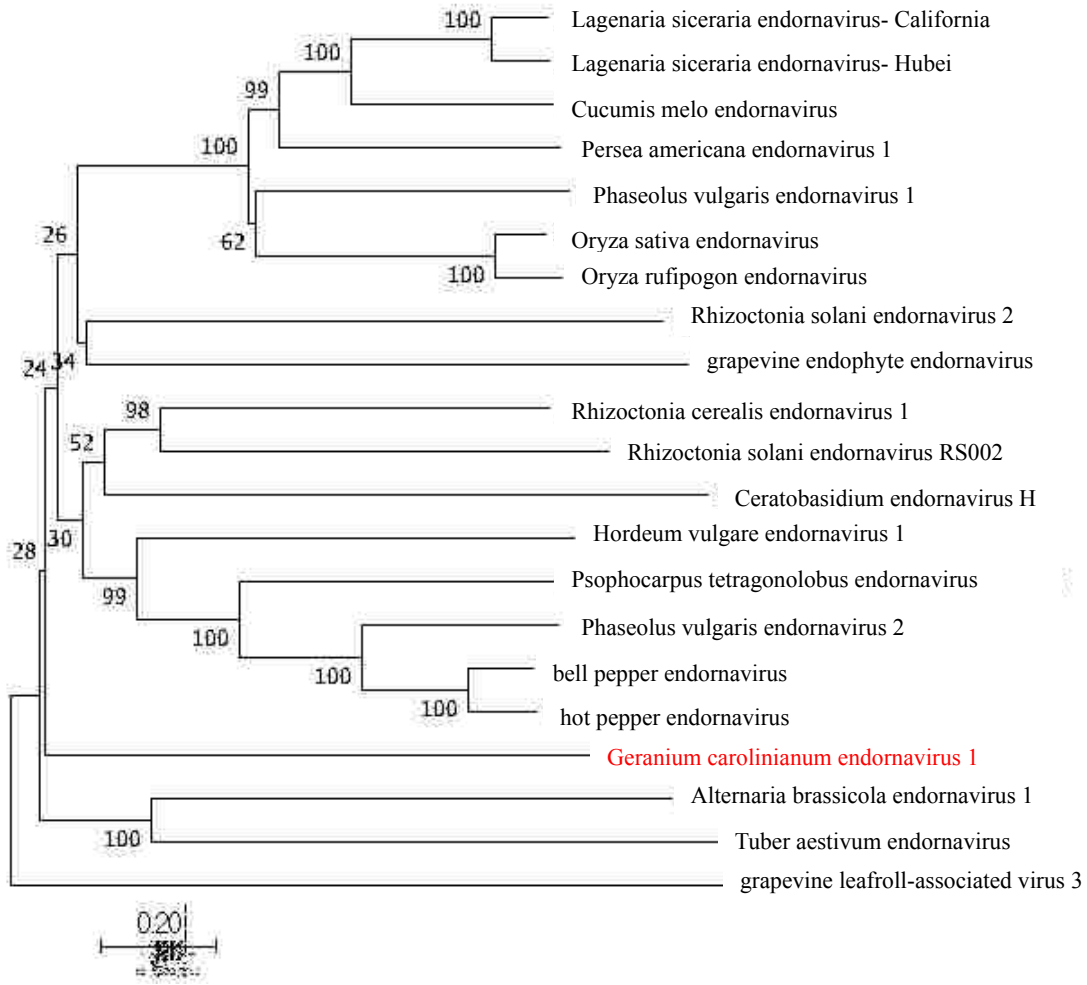

Figure 3.4 Evolutionary relationships of Geranium carolinianum endornavirus 1 to other taxa inferred using the Neighbor-Joining method (Saitou and Nei 1987).

3.4.4 Testing Selected Species within the Genus for Endornavirus-like dsRNA

All other *Geranium* species tested (*G. dissectum*, *G. lucidum*, *G. macrorrhizum* and *G. maculatum*) appeared to be endornavirus-like dsRNA free, based on the electrophoretic assays conducted.

3.4.5 Testing the Seed Coat and Progeny for Endornavirus-Like dsRNA

Both the aborted seeds (seeds comprised of seed coat only), and seeds comprised of both the seed coat and the embryo contained endornavirus-like dsRNA. Of the fifteen progeny from the parent plant, all fifteen had endornavirus-like dsRNA suggesting the presence of GcEV-1.

3.5 Discussion

The characterization of a novel endornavirus GcEV-1 adds to a very small group of endornaviruses described infecting non-cultivated plants, but suggests that there are possibly many other endornaviruses in non-cultivated plants that have not yet been described. GcEV-1 was similar to other endornaviruses, given the presence and position of a helicase (which mediates the unwinding of nucleic acid), a UDP-glycosyltransferase (which attaches sugar residues to small lipophilic chemicals and is required for the pathogenicity of plant pathogenic fungi such as *Colletotrichum gloeosporiodes* and *Magnaporthe grisea*), and RdRP (which is required for virus replication; Gorbalenya and Koonin 1993; Hansen et al. 1997; Hacker et al. 2005; Mackenzie et al. 2008, Espach et al. 2017). GcEV-1 also seems to be a unique endornavirus, given its lack of a MTR domain, presence of peptidase (an enzyme that cleaves peptide bonds via hydrolysis) and its high percent identity of UDP-glycosyltransferase to a wide range of bacterial species (Binkley et al. 1954; Hacker et al. 2005; Espach et al. 2017). Being that the UDP-glycosyltransferase of GcEV-1 is not conserved among other endornavirus species, but rather in bacterial species further supports the possibility that there was at some point horizontal gene transfer between bacteria species and endornaviruses (Song et al. 2013). Additionally, the GcEV-1 genome contained a peptidase related to peptidases in fungi. The dissimilarity of the UDP of GcEV-1 to the UDP of other endornaviruses may also suggest that there are other endornaviruses of non-cultivated plant species with genes that are more closely

related to bacteria. Similarly, the unique Peptidase C97 represents the first report of a peptidase gene in an endornavirus, and suggests that there are other endornaviruses of non-cultivated plant species with genes more closely related to fungi.

The percent identity of the RdRP and polyprotein GcEV-1 to other sequences being the highest for endornaviruses supports GcEV-1 being an endornavirus. However, the relatively low percent identity overall of GcEV-1 to other endornaviruses supports that GcEV-1 is a novel endornavirus. The percent identity of RdRP and the polyprotein of GcEV-1 to other endornaviruses were fairly similar, with the highest percent identity being to endornaviruses of plants and fungi. The high percent identity of GcEV-1 to endornaviruses of both plants and fungi suggests that more endornaviruses of both plants and fungi may need to be reported in order to better resolve the relationships between endornaviruses of different hosts. GcEV-1 also appears to be a unique endornavirus in that it did not cluster with any other endornaviruses in the neighbor-joining tree, although bootstrap values were low near the base of the tree.

Another goal of the study was to determine the transmission rate to progeny of endornavirus in a non-cultivated plant species. Although the high transmission rate of nearly 100% in *G. carolinianum* is similar to the transmission rate seen in crop species, more transmission tests will be needed to further confirm that the transmission rate in a non-cultivated plant species does not differ with respect to variables such as location and year. Endornavirus-like dsRNA was not detected in four other species of *Geranium.* This finding is not surprising, because there have been reports of endornaviruses occurring only in a single species of a genus (Sabanadzovic et al. 2016; Khankhum et al. 2015). It has been shown that the same or a closely related endornavirus can occur in different species of a genus (Okada et al. 201; Sabanadzovic et al. 2016; Moriyama et al. 1999).

Additionally, the description of GcEV-1 represents the first endornavirus described in a non-cultivated plant species that has not been domesticated. The other two non-cultivated plant species reported as having endornaviruses, *P. vulgaris* and *O. rufipogon* are both economically important crops, with *P. vulgaris* being domesticated and *O. rufipogon* representing the wild ancestor of *Oryza sativa* or cultivated rice (Londo et al. 2006). Due to GcEV-1 being the only endornavirus reported thus far in a plant species that has never been domesticated, the relationship between endornaviruses of non-cultivated plants and crop species remains unknown. Presently, it seems that evolutionary relationships of endornaviruses does not mirror the evolutionary relationships of endornavirus hosts. However, as more endornaviruses of non-cultivated plant species are likely described, the relationship between endornaviruses, as well as the reason for the relatively distant relationship between endornaviruses of closely related hosts (in some cases) may become clear.

CHAPTER IV: OCCURRENCE OF ENDORNAVIRUS-LIKE DSRNA IN CAROLINA GERANIUM IN THREE LOCATIONS

4.1 Introduction

Carolina Geranium (*Geranium carolinianum*, Geraniaceae) is a common weed native to North America, and is found in nearly every U.S. state (USDA NRCS). Within Louisiana, *G. carolinianum* has been collected in every parish (USDA NRCS). Although the common name for *G. carolinianum* is Carolina geranium, *G. carolinianum* is distinct from ornamental geraniums, which are also members of Geraniaceae, but are in the genus *Pelargonium*. The most closely related genus to *Geranium* is *Erodium* (Price and Palmer 1993).

Geranium carolinianum is an annual or biennial that typically grows in cooler weather, and blooms between the months of March and July, although in Louisiana it can bloom as early as February. Carolina geraniums are typically no taller than 0.5m and have leaves that are palmately five-parted, with leaf divisions being cleft or lobed (Radford et al. 2010). Flowers are five-petaled and pale pink, with a pistil of five carpels forming a long beak (Radford et al. 2010). Each carpel is single seeded, and at maturity, each carpel separates from the pistil, forming a long tail that aids in dispersing seeds in a catapult-like motion (Radford et al. 2010). *Geranium carolinianum* grows well in disturbed habitats, such as roadsides, lawns, pastures and near railroad tracks (Baskin and Baskin 1974). *Geranium carolinianum* seeds have physical dormancy, meaning that dormancy is caused by a hard seed coat impermeable to water (Baskin and Baskin 1974).

Studies on endornavirus infection in non-cultivated plant species are lacking. Endornaviruses have been described infecting wild rice (*Oryza rufipogon*) and wild common bean (*Phaseolus vulgaris*) (Moriyama et al. 1999; Khankhum et al. 2015). Other species have been described as containing endornavirus-like dsRNA, however they have not been confirmed

as being infected with endornavirus. Beyond the two non-cultivated plant species infected with endornavirus, nothing is known about endornaviruses in non-cultivated plant species. Although studies have described viruses in non-cultivated plant species, there have been no surveys to determine the occurrence of endornavirus in individual plants within a single non-cultivated plant species.

Previously, it was assumed that if one individual plant within a non-cultivated plant species was found to be infected with endornavirus, then the entire plant species would likely be infected. However, assuming endornavirus infection in all *G. carolinianum* can be problematic due to the documented variation in endornavirus infection between different cultivars of a crop species, such as the increased prevalence of endornavirus in *P. vulgaris* from Mesoamerica vs. the lesser prevalence in beans from the Andean region (Khankhum et al. 2015). The existence of ecotypes also makes it possible that endornavirus infection might vary between locations. The concept of ecotypes is generally defined as the genetic variation between local populations, with each population possessing heritable traits that make them better adapted to their particular environment (Solbrig 1970; Taylor and Murdey 1975). Ecotypes are considered to be a universal phenomenon among all non-cultivated plant species, including *G. carolinianum*, which has populations adapted to increased sulfur dioxide in locations with increased SO_2 pollution (Davis and Heywood 1963; Taylor and Murdy 1975). *Geranium carolinianum* populations growing in environments with increased SO_2 were shown to be better adapted to increased SO_2 levels compared to *G. carolinianum* populations that did not grow in areas with a high level of SO_2, but were then subjected to increased levels of the pollutant (Taylor and Murdy 1975). Endornavirus infection may also vary between individual plant species, especially in non-cultivated plant

species, which have more genetic variation compared to crops and are under increased natural selection pressures compared to cultivated plants.

4.2 Objective

A novel endornavirus, Geranium carolinianum endornavirus 1 (GcEV-1), infecting *G. carolininanum* has been identified and characterized (Chapter III). The objective of this study was to determine the occurrence of endornavirus-like dsRNA within populations of *G. carolinianum* in three distinct locations in Louisiana.

4.3 Materials and Methods

4.3.1 Location Selection

To test individual *G. carolinianum* plants for the presence of dsRNA, three locations in Louisiana were chosen for sampling: 1) The LSU AgCenter Central Research Station located at 2310 Ben Hur Road, Baton Rouge 2) The LSU AgCenter Botanic Gardens located at 4560 Essen Ln, Baton Rouge and 3) The Tulane University Biodiversity Research Institute (TUBRI) located at 3705 Main St., Belle Chasse. Each of the three locations was chosen for its difference in habitat, level of disturbance, and proximity to cultivated plant species. The LSU AgCenter Central Research Station has many areas where *G. carolinianum* is likely to grow, including roadsides, field edges, and drainage systems. Most of the area is dedicated to row crop cultivation and livestock. The LSU AgCenter Botanic Gardens has mostly field edges as potential habitat for *G. carolinianum* and is a site for cultivation of some row crops, horticultural crops, and ornamentals. TUBRI is located roughly 120 km from the other two sampling locations, is less disturbed habitat compared to the other two locations, has mostly roadsides and open fields as potential habitat for *G. carolinianum*, and is not the site for any plant cultivation,

with the grounds being mostly dedicated to fish specimen storage, and the surrounding area being open fields that are lightly maintained.

4.3.2 Sampling *G. carolinianum*

Collections of *G. carolinianum* plants were conducted from February 9th to April 27th of 2017. A modified hoop sampling strategy was used for collection of *G. carolinianum* samples (Cavieres et al. 2005). Because sampling was for a single, common weed species, *G. carolinianum*, a hoop 86cm in diameter was used to limit the number of *G. carolinianum* samples collected, rather than to control for bias when collecting plant species in a given area (Cavieres et al. 2005). At each collection site (LSU AgCenter Central Research Station, LSU Agricultural Center Botanic Gardens, and Tulane University Biodiversity Research Institute), *G. carolinianum* specimens were collected by placing a hoop around a cluster of *G. carolinianum* and collecting all individuals within the hoop (Fig. 4.1). One hoop represented a sampling area, and ten areas were sampled per site. Each site was sampled twice, representing sixty sampling areas in total. For each sampling area, GPS location was recorded using My GPS Coordinates Pro™ (Appendix 6).

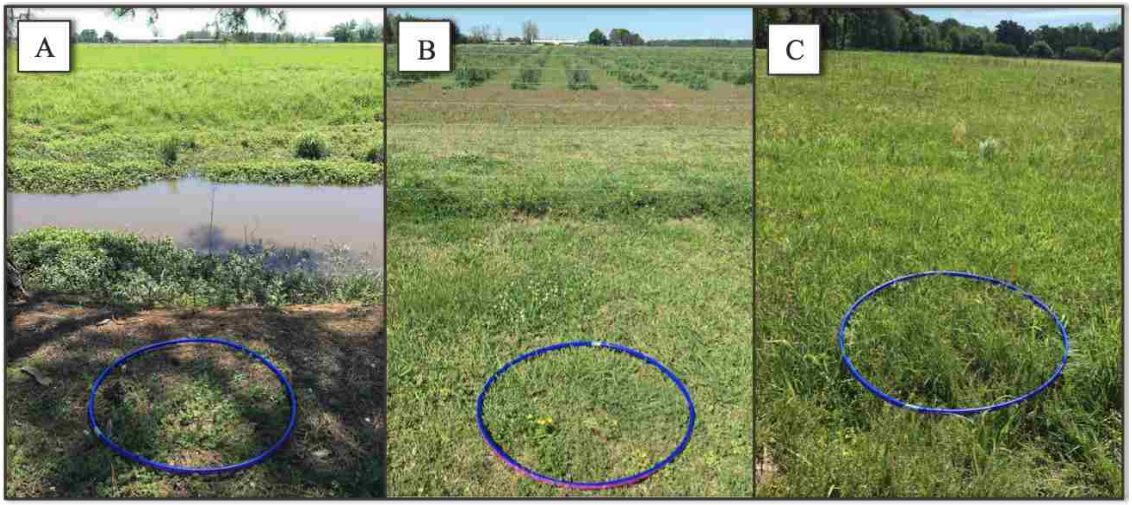

Figure 4.1 Representation of hoop sampling for *G. carolinianum* at three locations in Louisiana. A) An example of sampling at the LSU Central Research Station on Ben Hur Rd., Baton Rouge, B) Sampling at LSU Agricultural Center Botanic Gardens, Baton Rouge; C) Sampling at Tulane University Biodiversity research institute, Belle Chasse.

4.3.3 Plant Identification and Recording of Phenotypic Traits

Each individual plant was photographed and its phenotypic traits recorded, including stage of maturity, flower color, stem and petiole color and leaf shape.

4.3.4 Testing for Presence or Absence of Endornavirus-like dsRNA by Electrophoretic Analysis

The dsRNA technique was used as a practical and reliable tool to detect GcEV-1 in this investigation. The presence or absence of endornavirus-like dsRNAwas determined using a modified version of the dsRNA extraction method developed by Khankhum et al. (2016) as described in Chapter II. Six dsRNA samples from *G. carolinianum* from each collection were tested by RT-PCR using degenerate primers as described in Chapter II.

4.3.5 Testing *G. carolinianum* Infected with Pathogens

During the sample collection, some individual *G. carolinianum* plants were found naturally infected with *Synchytrium* sp., and unidentified fungi and oomycetes causing powdery mildew and downy mildew. The presence of pathogens was confirmed through light microscopy. To

determine if the infection of these pathogens had an effect on the presence of endornavirus-like dsRNA, samples were tested, using the dsRNA extraction method (Khankhum et al. 2016). The dsRNA from pathogen-infected plants was resolved in 1.2% agarose gel at 70 V for 2 h and visually compared to dsRNA obtained from healthy *G. carolinianum*.

4.4 Results

4.4.1 Testing for Presence or Absence of Endornavirus-like dsRNA by Electrophoretic Analyses

In total, 184 plants were tested for the presence of endornavirus-like dsRNA. One hundred eighty-one plants contained endornavirus-like dsRNA, and three plants were be endornavirus free, representing over 98% occurrence. RT-PCR testing confirmed the endornavirus presence in six selected samples from each location. The absence in the three dsRNA-free plants was confirmed by RT-PCR (see Chapter II).

There was no clear phenotypic difference between individuals with endornavirus-like dsRNA and those that were dsRNA-free (Fig. 4.3). The three *G. carolinianum* plants that were dsRNA-free had similar leaf color, stem color, and leaf shape to many other *G. carolinianum* plants collected. One of the dsRNA-free individual plants was infected with *Synchytrium sp*., however all other *G. carolinianum* infected with *Synchytrium sp.* contained endornavirus-like dsRNA. Therefore no phenotypic variation could be associated with the presence of the endornavirus-like dsRNA.

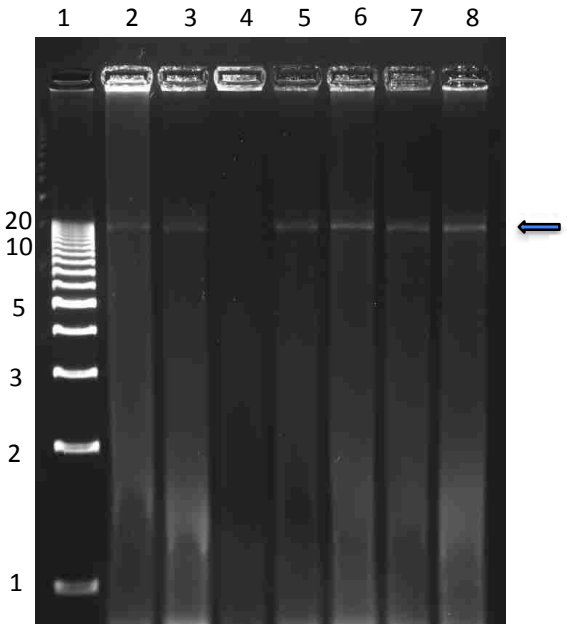

Figure 4.2 Agarose gel (1.2%) with dsRNA extracted from *G. carolinianum* samples collected in Louisiana. Lane 1, 1kb ladder; lanes 2, 3, 5, 6, 7, 8 *G. carolinianum* plants with endornavirus-like dsRNA; lane 4, dsRNA-free *G. carolinianum* collected from the LSU Central Research Station.

Figure 4.3. Examples of phenotypic plasticity observed in *G. carolinianum* collected in Louisiana. A) Individuals represent four of the 181 individuals infected with endornavirus. B) The three endornavirus-free plants collected.

4.4.2 Testing *G. carolinianum* Infected with Pathogens

G. carolinianum individuals infected with *Synchytrium sp.* as well as individuals infected with a powdery mildew and a downy mildew were also contained endornavirus-like dsRNA. There was no clear difference in the intensity of the bands that resulted from gel electrophoresis when inspected visually, whether plants were infected with a known pathogen in addition to endornavirus, or infected solely with endornavirus.

4.5 Discussion

In this survey, the endornavirus-nature of selected dsRNAs from the three locations was confirmed by positive RT-PCR tests. Therefore it is assumed that the endornavirus-like dsRNA detected in this survey consists of the replicative form of an endornavirus.

The survey of *G. carolinianum* plants for the presence of endornavirus-like dsRNA represents the first survey that tests for their occurrence within a single non-cultivated plant species by collecting plants directly from their natural habitat and testing each plant individually. More *G. carolinianum* ecotypes from other geographic locations will need to be surveyed in order to further confirm the widespread nature of endornavirus infection within Carolina geranium. The results suggest that at least in the case of *G. carolinianum*, individuals with endornavirus-like dsRNA are not being selected against, and therefore may provide some unknown benefit to the species. Of course, determining any potential benefit will require the identification and collection of more endornavirus-like dsRNA-free individual *G. carolinianum* plants.

The percentage of *G. carolinianum* plants containing endornavirus-like dsRNA (over 98%) is similar to reported percentages of endornaviruses in crops species (Horiuchi et al. 2003, Valverde and Gutierrez 2007; Okada et al. 2011, 2013, 2014, 2017). However more non-cultivated plant species infected with endornavirus will need to be surveyed. Additional surveys

will need to be conducted on infected wild relatives of crop species to further compare the percent of endornavirus-like dsRNA infection in non-cultivated plants vs. cultivated. This may help to determine why in some cases endornavirus infection seems to be widespread in non-cultivated plants (as in *G. carolinianum*) and why in other cases endornavirus infection is less common (as seen in non-cultivated *P. vulgaris* from the Andean region) (Khankhum et al. 2015).

Observing both the occurrence of endornavirus-like dsRNA in *G. carolinianum* where it naturally grows, as well as the transmission rate to progeny in a controlled environment (Chapter III) is one way of determining whether the occurrence of endornavirus differs in a plant species when the selection pressures/ abiotic stressors are reduced (natural habitat vs. greenhouse setting). At this point in time, there seems to be no difference, although more samples in both the greenhouse and the field would need to be taken to definitively answer the question. The relationship between endornavirus infection and *G. carolinianum* survival was also investigated in the previous chapter by testing aborted *G. carolinianum* seeds for the presence of endornavirus. The presence of endornavirus in both aborted and fully formed seeds provides evidence that the presence of the endornavirus is not required for embryogenesis, but endornavirus may play a more complex role in *G. carolinianum* survival.

Endornavirus infection in a non-cultivated plant species was also similar to endornavirus infection in crop species with respect to intensity of dsRNA bands in the presence of a known pathogen, as well as occurrence of infected plant species within a genus. The presence of endornavirus in *G. carolinianum* plants both healthy (containing only endornavirus) and plants infected with *Synchytrium sp.,* powdery mildew or downy mildew demonstrates that the presence of the virus does not prevent plants from being infected with these plant pathogens. However, more complex interactions between the virus and known pathogens are certainly possible. Co-

infection of endornaviruses and known pathogens such as acute viruses has been observed in crop species, and is therefore not unique to *G. carolinianum* (Khankhum 2016). Also as generally seen in crop species, endornavirus infection is not widely distributed across a genus, and only one or a few species may have the virus. Such appears to be the case in *Geranium,* with all other species tested appearing to be endornavirus-free, however more species within the genus and individuals within a species will need to be tested.

One question that remains is why three individuals were dsRNA-free. Although in all cases where infection has been tested within a plant species, some samples have been endornavirus-free, never has the absence of endornavirus been connected to a single location with a relatively small area such as the LSU AgCenter Central Research Station. Rather, studies have either connected the absence of endornavirus to either an origin of domestication or a certain cultivar of a crop species. While there is always the possibility of a false negative, the association of endornavius-free individuals with one specific location leaves the possibility that endornavirus infection could be more location-dependent than previously thought.

With the *G. carolinianum* endornavirus survey being the first survey for endornavirus-like dsRNA presence in a non-cultivated plant species, testing plants collected directly from their natural habitat, it appears that endornavirus has similar patterns of infection in non-cultivated plant species as seen in some crops. However, as more endornaviruses of non-cultivated plant species are discovered, more detailed surveys will need to be completed in order to gain a complete picture of how endornavirus infection varies within a non-cultivated plant species, and if endornavirus occurrence within *G. carolinianum* is typical. With a main goal of endornavirus studies being to determine the interaction endornaviruses have with their plant hosts, studying non-cultivated plant species more subject to abiotic stress and interspecific competition has the

potential to better elucidate complex interactions between the virus and hosts that may not be easily observable in crop species.

CHAPTER V: CONCLUSIONS

The main goal of this investigation was to determine the occurrence of endornavirus in non-cultivated plant species. Another objective was to determine endornavirus occurrence within a single non-cultivated plant species in the hopes of laying the groundwork for future comparative studies that will elucidate the interactions between endornaviruses and their non-cultivated plant hosts. Previously, endornaviruses were viruses mostly associated with crop species, and their occurrence in non-cultivated plant species was largely unknown. For a virus that is only transmitted vertically to progeny through seed and is thought to have pre-dated agriculture, only knowing the distribution in crops represents a major gap in the current knowledge of plant endornaviruses. Although some surveys had found that endornaviruses are present in non-cultivated plants, no survey had yet looked into what specific species are infected, as well as the occurrence within a non-cultivated plant species.

From the survey of non-cultivated plant species within Baton Rouge, it is now known that endornaviruses do occur in several non-cultivated plant species and that they are not viruses unique to non-cultivated plant species, or species that represent the ancestors of infected crop species (*Oryza rufipogon* and *Phaseolus vulgaris*). The fact that a novel endornavirus and five putative endornaviruses were found suggests that there may be endornaviruses in non-cultivated plants yet to be discovered. Determining how many non-cultivated plant species are infected with endornavirus will provide better insight into how and why endornavirus was introduced into many crop species, with the survey in this study providing a starting point for future plant surveys.

The description of a novel endornavirus detected in *Geranium carolinianum* represents the first description of an endornavirus in a non-cultivated plant species that has never been

domesticated, and provides further evidence that there may have been early horizontal gene transfer between endornaviruses and fungi, as well as bacteria. The genome organization of Geranium carolinianum endornavirus 1 (GcEV-1), which includes a viral helicase and an RNA-dependent RNA polymerase is typical of other endornaviruses. However the similarity of the UDP of GcEV-1 to the UDP of bacteria species, as well as the similarity of Peptidase C97 to peptidases of fungi suggests endornavirus genes may have more diverse origins than previously thought. It is possible that as more endornaviruses of non-cultivated plant species are discovered, more genes similar to those found in fungal and bacterial species will be described.

When comparing endornavirus-free non-cultivated plants with infected plants, the first step is to find endornavirus-free individuals. The *G. carolinianum* survey laid the groundwork for finding endornavirus- free individuals for future studies, both by determining that endornavirus- free individuals exist, but that they are rare, at less than (1%) of the plant species tested, and no specific phenotype can be associated with endornavirus- free individuals at this point in time. Future comparative studies between infected *G. carolinianum* and endornavirus-free *G. carolinianum* will most likely need to sample hundreds of individuals from multiple locations, or test all progeny from a few parent plants in order to successfully obtain endornavirus-free plants that can be used to develop near-isogenic lines.

As for questions regarding how endornaviruses interact with their plant hosts, the method of transmission of endornaviruses poses a challenge for comparative studies. Until a method of transmission other than vertical transmission to progeny through seed is discovered, comparative studies can only be conducted by discovering a naturally occurring endornavirus-free plant (whether a crop species or non-cultivated plant species), then creating near-isogenic lines of infected and endornavirus-free plants for comparison.

The *G. carolinianum* survey for endornavirus-like dsRNA opens a new area of research, looking at endornaviruses in non-cultivated plant species in order to address questions about endornaviruses in crop species. Finding that endornaviruses are likely present in many non-cultivated plant species, and that nearly every *G. carolinianum* plant was infected establishes that non-cultivated plants are worth further investigating and poses further questions of how representative endornavirus-like dsRNA infection of *G. carolinianum* plants is of other non-cultivated plant hosts. As more endornavirus plant hosts are inevitably discovered along with endornavirus-free individuals, determining the interaction between endornaviruses and their plant hosts will likely be accomplished.

REFERENCES

Adams, M.J., Lefkowitz, E.J., King, A.M., Harrach, B., Harrison, R.L., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Mushegian, A.R. and Nibert, M. 2017. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses. Archives of Virology 162: 2505-2538.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Altschul, S. F., Wootton, J.C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A., and Yu, Y. K. 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS Journal 272: 5101-5109.

Bankevich, A, Nurk, S, and Antipov, D. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455-477.

Bar-Joseph M, Rosner A, Moskovitz M and Hull R. 1983. A simple procedure for the extraction of double-stranded RNA from viral infected plants. Journal of Virological Methods 6:1-8.

Baskin, J. M. and Baskin, C. C. 1974. Some eco-physiological aspects of seed dormancy in Geranium carolinianum L. from central Tennessee. Oecologia 16: 209-219.

Bernal, J. D., and Fankuchen, I. 1937. Structure types of protein 'crystals' from virus-infected plants. Nature 139: 923-924.

Binkley, F. 1954. Organization of Enzymes in the Synthesis of Peptides. Proceedings of the Royal Society of London. Series B, Biological Sciences: 170-174.

Buck, K. W. 1999. Replication of tobacco mosaic virus RNA. Philosophical Transactions of the Royal Society B: Biological Sciences 354: 613-627.

Busby, P. E., Ridout, M., and Newcombe, G. 2016. Fungal endophytes: modifiers of plant disease. Plant Molecular Biology 90: 645-655.

Davis, P.H., and Heywood, V.H. 1963. *Principles of angiosperm taxonomy*. Van Nostrand, New York.

Derrick, K. S. 1978. Double-stranded RNA is present in extracts of tobacco plants infected with tobacco mosaic virus. Science 199: 538-539.

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

Escalante, C., Sela, N., Dombrovsky, A., and Valverde, R. A. 2016. Interactions between bell pepper endornavirus, bell pepper, and acute plant viruses. Procceedings XVIth EUCARPIA Capsicum and Eggplant Working Group Meeting, 12 -14 September 2016, Kecskemét, Hungary.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-97.

Espach, Y., Maree, H.J. and Burger, J.T., 2012. Complete genome of a novel endornavirus assembled from next-generation sequence data. Journal of Virology 86: 13142-13142.

Fukuhara, T. 1999. Double-stranded RNA in rice. Journal of Plant Research 112: 131-138.

Fukuhara, T., and Gibbs, M. J. 2012. Family Endornaviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, pp 855-880

Fukuhara, T., Koga, R., Aoki, N., Yuki, C., Yamamoto, N., Oyama, N., Udagawa, T., Horiuchi, H., Miyazaki, S., Higashi, Y. and Takeshita, M. 2006. The wide distribution of endornaviruses, large double-stranded RNA replicons with plasmid-like properties. Archives of Virology 151: 995-1002.

Grill, L. K., and Garger, S. J. 1981. Identification and characterization of double-stranded RNA associated with cytoplasmic male sterility in Vicia faba. Genetics 78: 7043-7046.

Gorbalenya, A.E. and Koonin, E.V. 1993. Helicases: amino acid sequence comparisons and structure-function relationships. Current opinion in structural biology 3: 419-429.

Hacker, C. V., Brasier, C. M., and Buck, K. W. 2005. A double-stranded RNA from a Phytophthora species is related to the plant endornaviruses and contains a putative UDP glycosyltransferase gene. Journal of General Virology 86: 1561-1570.

Hansen, J.L., Long, A.M. and Schultz, S.C., 1997. Structure of the RNA-dependent RNA polymerase of poliovirus. Structure, 5(8), pp.1109-1122.

Horiuchi, H. and Fukuhara, T. 2004. Putative replication intermediates in Endornavirus, a novel genus of plant dsRNA viruses. Virus Genes 29: 365-375.

Hull, R. 2013. *Plant Virology: Fifth Edition*. Academic Press.

ICTV. 2017. Virus taxonomy: the classification and nomenclature of viruses. The online (10th) report of the ICTV. Obtained from https://talk.ictvonline.org/ ictv-reports/ictv_online_report/.

Iwanowski, D. 1892. Über die Mosaikkrankheit der Tabakspflanze. Bulletin Scientifique publié par l'Académie Impériale des Sciences de Saint-Pétersbourg / Nouvelle Serie III (in German and Russian). St. Petersburg. 35: 67–70. Translated into English in Johnson, J., Ed. 1942 Phytopathological classics (St. Paul, Minnesota: American Phytopathological Society) No. 7, pp. 27-30.

Kado, C. I. 1998. Origin and evolution of plasmids. In Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology 73: 117-126.

Khalifa, M. E. and Pearson, M. N. 2014. Molecular characterisation of an endornavirus infecting the phytopathogen *Sclerotinia sclerotiorum*. Virus Research 189: 303-309.

Khankhum, S. 2016. Persistent RNA viruses of common bean (*Phaseolus vulgaris*): distribution and interaction with the host and acute plant viruses. *LSU Doctoral Dissertations*. 838.

Khankhum, S., Valverde, R. A., Pastor, M. A., Osorno, J. M., and Sabanadzovic, S. 2015. Two endornaviruses show differential infection patterns between gene pools of *Phaseolus vulgaris*. Archives of Virology 160: 1131-1137.

Khankhum, S., Escalante, C., de Souto, E. R. and Valverde, R. A. 2017. Extraction and electrophoretic analysis of large dsRNAs from desiccated plant tissues infected with plant viruses and biotrophic fungi. European Journal of Plant Pathology 147: 431-441.

King, A.M., Lefkowitz, E., Adams, M.J. and Carstens, E.B. 2011. Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier.

Kumar S., Stecher G., and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and Evolution 33: 1870-1874.

Kwon, S.,J., Tan, S. H. and Vidalakis, G. 2014. Complete nucleotide sequence and genome organization of an endornavirus from bottle gourd (Lagenaria siceraria) in California, USA. Virus Genes 49: 163-168.

Libonati, M. Casarna, N. and Furia, A. 1980. Double-stranded RNA. Molecular and Cellular Biochemistry 31: 147-164.

Londo, J.P., Chiang, Y.C., Hung, K.H., Chiang, T.Y. and Schaal, B.A. 2006. Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa. Proceedings of the National Academy of Sciences 25:9578-9583.

Mackenzie, P.I., Rogers, A., Treloar, J., Jorgensen, B.R., Miners, J.O. and Meech, R., 2008. Identification of UDP glycosyltransferase 3A1 as a UDP N-acetylglucosaminyltransferase. Journal of Biological Chemistry, 283(52), pp.36205-36210.

Malmstrom, C. M., Melcher, U., and Bosque-Pérez, N. A. 2011. The expanding field of plant virus ecology: Historical foundations, knowledge gaps, and research directions. Virus Research 159: 89–94.

Mayer, A. 1886. Ueber die Mosaikkrankheit des Tabaks. Die Landwirtschaftlichen Versuchs-Stationen 32: 451-467.

Marquez, L. M., Redman, R. S., Rodriguez, R. J., and Roossinck, M. J. 2007. A Virus in a Fungus in a Plant: Three-Way Symbiosis Required for Thermal Tolerance. Science 315: 513-515.

Miller, J.F., 1969. Weeds of the southern United States. Georgia Cooperative Extension Service, U.S. Department of Agriculture.

Moriyama, H., Kanaya, K., Wang, J. Z., Nitta, T., and Fukuhara, T. 1996. Stringently and developmentally regulated levels of a cytoplasmic double-stranded RNA and its high-efficiency transmission via egg and pollen stranded RNA and its high-efficiency transmission via egg and pollen in rice. Plant Molecular Biology 31: 713-719.

Moriyama, H., Horiuchi, H., Nitta, T. and Fukuhara, T. 1999. Unusual inheritance of evolutionarily-related double-stranded RNAs in interspecific hybrid between rice plants *Oryza sativa* and *Oryza rufipogon*. Plant Molecular Biology, 39: 1127-1136.

Morris, T. J. and Dodds, J. A. 1979. Isolation and analysis of double-stranded RNA from virus infected plant and fungal tissue. Phytopathology 69: 854-858.

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

Nakatsukasa-Akune, M., Yamashita, K., Shimoda, Y., Uchiumi, T., Abe, M., Aoki, T., Kamizawa, A., Ayabe, S.I., Higashi, S. and Suzuki, A. 2005. Suppression of root nodule formation by artificial expression of the TrEnodDR1 (Coat protein of White

clover cryptic virus 1) gene in *Lotus japonicas*. Molecular Plant-Microbe Interactions 18: 1069-1080.

Okada, R., Kiyota, E., Sabanadzovic, S., Moriyama, H., Fukuhara, T., Saha, P., Roossinck, R., Severin, A., and Valverde, R. A. 2011. Bell pepper endornavirus: molecular and biological properties, and occurrence in the genus *Capsicum*. Journal of General Virology 92: 2664-2673.

Okada, R., Kiyota, K., Moriyama, H. Fukuhara, T., and Valverde, R. A. 2014. A new endornavirus species infecting Malabar spinach (*Basella alba* L.). Archives of Virology 159: 807-809.

Okada, R., Yong, C.K., Valverde, R.A., Sabanadzovic, S., Aoki, N., Hotate, S., Kiyota, E., Moriyama, H. and Fukuhara, T. 2013. Molecular characterization of two evolutionarily distinct endornaviruses co-infecting common bean (*Phaseolus vulgaris*). Journal of General Virology 94: 220-229.

Okada, R., Kiyota, E., Moriyama, H., Fukuhara, T. and Valverde, R. A. 2017. Molecular and biological properties of an endornavirus infecting winged bean (*Psophocarpus tetragonolobus*). Virus Genes 53: 141-145.

Osaki, H., Nakamura, H., Sasaki, A., Matsumoto, N., and Yoshida, K. 2006. An endornavirus from a hypovirulent strain of the violet root rot fungus, *Helicobasidium mompa*. Virus Research 118: 143-149.

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

Price, R. A. and Palmer, J. D. 1993. Phylogenetic relationships of the Geraniaceae and Geraniales from rbcL sequence comparisons. Annals of the Missouri Botanical Garden: 661-671.

Radford, A. E., Ahles, H. E. and Bell, C.R. 2010. *Manual of the vascular flora of the Carolinas*. University of North Carolina Press.

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

Roossinck, M. J., Sabanadzovic, S., Okada, R., and Valverde, R. A. 2011. The remarkable evolutionary history of endornaviruses. Journal of General Virology 92: 2674-2678.

Roossinck, M. J. 2012. Plant Virus Metagenomics: Biodiversity and Ecology. Annual Review of Genetics 46: 359-369.

Roossinck, M. J. 2016. *Virus: An Illustrated Guide to 101 Incredible Microbes*. Princeton University Press.

Sabanadzovic, S., Wintermantel, W. M., Valverde, R. A, McCreight, J. D., and Aboughanem-Sabanadzovic, N. 2016. Cucumis melo endornavirus: genome organization, host range and co-divergence with the host. Virus Research 214:49-58.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Shang, H. H., Zhong, J., Zhang, R. J., Chen, C. Y., Gao, B. D., and Zhu, H. J. 2015. Genome sequence of a novel endornavirus from the phytopathogenic fungus *Alternaria brassicicola.* Archives of Virology 160: 1827-1830.

Song, D., Cho, W. K., Park, S. H., Jo, Y., and Kim, K. H. 2013. Evolution of and horizontal gene transfer in the Endornavirus genus. PLoS One 8: e64270.

Solbrig, O.T. 1970. *Principles and methods of plant biosystematics*. Macmillan, London.

Stielow, B., Klenk, H. P., and Menzel, W. 2011. Complete genome sequence of the first endornavirus from the ascocarp of the ectomycorrhizal fungus *Tuber aestivum* Vittad. Archives of Virology 156: 343–345.

Stobbe, A. H., and Roossinck, M. J. 2014. Plant virus metagenomics: what we know and why we need to know more. Frontiers in Plant Science 5.

Taylor, G. E., and Murdy, W. H. 1975. Population differentiation of an annual plant species, Geranium carolinianum, in response to sulfur dioxide. Botanical Gazette 136: 212-215.

Thomas, R. D. and Allen, C.M., 1993. *Atlas of the vascular flora of Louisiana*, *V.* I-III. Louisiana Department of Wildlife and Fisheries, National Heritage Program.

Tzanetakis, I. E. and Martin, R. R. 2008. A new method for extraction of double-stranded RNA from plants. Journal of Virological Methods 149: 167-70.

USDA NRCS. United States Department of Agriculture Natural Resources Conservation Service Plants Database. Obtained from https://plants.usda. gov

Valverde, R. A., Dodds, J. A. and Heick, J. A. 1986. Double stranded RNAs from plants infected with viruses having elongated particles and undivided genomes. Phytopathology 76: 459-465.

Valverde, R. A., Nameth, S., Abdallha, O., Al-Musa, O., Desjardins, P. and Dodds, A., 1990. Indigenous double-stranded RNA from pepper (*Capsicum annuum*). Plant Science 67: 195-201.

Valverde, R. A., Nameth, S. T., and Jordan, R. L. 1990. Analysis of double-stranded RNA for plant virus diagnosis. Plant Disease 74: 255-258.

Valverde, R. A., and Gutierrez, D. L. 2007. Transmission of a dsRNA in bell pepper andevidence that it consists of the genome of an endornavirus. Virus Genes 35: 399-403.

Valverde, R. A., Sabanadzovic, S., and Rush, M. C. 2011. Identification of Oryza sativa endornavirus in rice genotypes from breeding programmes in the United States. Plant Breeding 1302: 271-274.

Van Der Want, J. P. H., and Dijkstra, J. 2006. A history of plant virology. Archives of Virology 151: 1147-1498.

Villanueva, F., Sabanadzovic, S., Valverde, R. A., Navas-Castillo, J. 2012. Complete genome sequence of a double-stranded RNA virus from avocado. Journal of Virology 86: 1282-1283

Wakarchuk, D. A., and Hamilton, R.J. 1985. Cellular double-stranded RNA in Phaseolus vulgaris. Plant molecular biology 5: 55-63.

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

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

Zabalgogeazcoa, I., Cox-Foster, D., and Gildow, F. 1993. Pedigree analysis of the transmission of a double-stranded RNA in barley cultivars. Plant Science 91: 45-53.

Zapalski, M. K., 2011. Is absence of proof a proof of absence? Comments on commensalism. Palaeogeography, Palaeoclimatology, Palaeoecology 302: 484-488.

Zuckerlandl, E. and Pauling, L. 1965. Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.

APPENDIX

Appendix 1 Plant species tested for endorna-like dsRNA. A + indicates the presence of endorna-like dsRNA, - indicates the absence of endorna-like dsRNA, and a P indicates the presence of dsRNA potentially from other persistent viruses. N= native plant species; I= introduced plant species; ?= unknown.

Family	Species	dsRNA	Origin	Common Name	Year Tested
Acanthaceae	*Ruellia brittoniana*	-	I	Mexican petunia	2015
Acanthaceae	*Ruellia caroliniensis*	-	N	Carolina wild petunia	2016
Alismataceae	*Echinodorus cordifolius*	-	N	Creeping burhead	2016
Alismataceae	*Saggitaria lancifolia*	-	N	Bulltongue arrowhead	2015
Amaranthaceae	***Alternanthera philoxeroides***	+	I	Alligator weed	2015
Amaranthaceae	*Amaranthus retroflexus*	-	N	Common amaranth	2015
Amaryllidaceae	*Alium canadense* var. *canadense*	-	N	Meadow garlic	2016
Anacardiaceae	*Rhus glabra*	-	N	Smooth sumac	2015
Annonaceae	*Asimina triloba*	-	N	Pawpaw	2015
Apiaceae	*Chaerophyllum tainturieri*	-	N	Hairyfruit chervil	2016
Apiaceae	*Cyclospermum leptophyllum*	-	I	Fir-leaved celery	2016
Apiaceae	*Hydrocotyle bonariensis*	-	N	Largeleaf pennywort	2015; 2016
Apiaceae	***Hydrocotyle prolifera***	+	N	Whorled pennywort	2015; 2016
Apiaceae	***Hydrocotyle umbellata***	+	N	Dollar-weed	2015; 2016
Apiaceae	*Hydrocotyle verticillata*	-	N	Whorled pennywort	2015; 2016
Apiaceae	*Ptilimnium cappilaceum*	-	N	herbwilliam	2016
Araceae	*Colocasia eculenta*	-	I	Elephant ear	2015
Araceae	*Lemna minor*	-	N	Duckweed	2015
Araceae	*Syngonium podophyllum*	-	I	Arrowhead vine	2015
Araliaceae	*Heredera helix*	-	I	English ivy	2015
Arecaceae	*Sabal minor*	-	N	Dwarf palmetto	2015; 2016
Asclepiadaceae	*Asclepias tuberosa*	-	I	Milkweed	2015
Asparagaceae	*Aspidistra elatior*	-	I	Cast Iron plant	2015
Asteraceae	*Acmella oppositifolia*	-	N	Oppositeleaf spotflower	2016
Asteraceae	*Cirsium horridulum*	-	N	Bristle thistle	2015; 2016
Asteraceae	*Crepis tectorum*	-	I	Narrow-leaf hawksbeard	2016
Asteraceae	***Dracopis amplexicaulis***	**+ P**	N	Clasping coneflower	2015; 2016

Asteraceae	*Erigeron annuus*	-	N	Annual fleabane	2016
Asteraceae	*Helianthus* sp.		N	Sunflower	2015; 2016
Asteraceae	*Krigia caespitosa*	-	N	Weedy dwarfdandelion	2016
Asteraceae	*Pyrrhopappus carolinianus*	-	N	Carolina desert chicory	2015; 2016
Asteraceae	***Sonchus asper***	+	I	Sow thistle	2015; 2016
Astereaceae	*Ambrosia artemisiifolia*	-	N	Ragweed	2015
Astereaceae	*Calyptocarpus vialis*	-	N	Horseherb	2015
Astereaceae	*Cirsium horridulum*	-	N	Spiny thistle	2015; 2016
Astereaceae	*Coreopsis basalis*	-	N	Goldenname coreopsis	2015
Astereaceae	*Echinacea purpurea*	-	N	Purple coneflower	2015
Astereaceae	*Eupatorium capillifolium*	-	N	Dog fennel	2015
Astereaceae	*Gaillardia pulchella*	-	N	Firewheel	2015
Astereaceae	*Gnaphalium* sp.	-	N	Cudweed	2015
Astereaceae	*Parthenium hysterophorus*	-	N	Santa Maria feverfew	2015
Astereaceae	*Pyrrhopappus carolinianum*	-	N	False dandelion	2015
Astereaceae	*Ratibida columnifera*	-	N	Mexican hat	2015
Astereaceae	*Rudbeckia fulgida*	-	N	Goldstrum rudbeckia	2015
Astereaceae	*Soliva sessilis*	-	I	Field burrowed	2015
Astereaceae	*Taraxacum officinalis*	-	I	Dandelion	2015
Astereaceae	*Wedelia trilobata*	-	N	Wedelia	2015
Astereaceae	*Xanthium strumarium*	-	I	Cocklebur	2015
Astereaceae	*Youngia japonica*	-	I	False hawkbeard	2015
Berberidaceae	*Nandina domestica*	-	I	Nandina	2015
Betulaceae	*Carpinus caroliniana*	-	N	American hornbean	2016
Boraginaceae	*Heliotropium indicum*	-	I	Indian heliotrope	2015
Brassicaceae	*Coronopus didymus*	-	I	Lesser swinecress	2016
Brassicaceae	*Coronopus* sp.	-	I	Swinecress	2015
Brassicaceae	*Lepidium virginicum*	-	N	Pepperwort	2016
Bromeliaceae	*Tillandsia usneoides*	-	N	Spanish moss	2015
Campanulaceae	*Lobelia cardinalis*	-	N	Cardinal flower	2016
Campanulaceae	*Triodanis perfoliata*	-	N	Clasping venus's looking glass	2016
Caprifoliaceae	*Lonicera japonica*	-	I	Japanese honeysuckle	2015
Caprifoliaceae	*Sambucus canadensis*	-	N	Elderberry	2015
Caryophyllaceae	*Cerastium fontanum*	-	I	Mouse-ear chickweed	2015

Caryophyllaceae	*Stellaria media*	-	I	Common chickweed	2015
Commelinaceae	*Commelina communis*	-	I	Dayflower	2015
Commelinaceae	*Commelina* sp.	-	?	Dayflower	2016
Commelinaceae	*Gibasis pellicida*	-	I	Dotter bridal veil	2015
Commelinaceae	*Tradescantia fluminensis*	-	I	small-leaf spiderwort	2016
Commelinaceae	*Tradescantia hirsutiflora*	-	N	Hairyflower spiderwort	2015
Convolvulaceae	*Cuscuta japonica*	-	I	Japanese dodder	2015
Convolvulaceae	*Jacquemontia tamnifolia*	-	I	Smallflower morninglory	2015
Cupressaceae	*Taxodium distichum*	-	I	bald cypress	2015; 2016
Cyperaceae	*Cyperus entrerianus*	-	I	Deep-rooted sedge	2015
Cyperaceae	*Cyperus rotundus*	-	I	Sedge	2015
Cyperaceae	*Eleocharis albida*	-	N	White spikerush	2016
Dryopteridaceae	*Dryopterism ludoviciana*	-	N	Louisiana wood fern	2015
Ebenaceae	*Diospyros virginiana*	-	N	Persimmon	2015
Equisetaceae	*Equisetum hyemale*	-	N	Horsetail	2015
Euphorbiaceae	*Euphorbia nutans*	-	I	Eyebane sandmat	2015
Euphorbiaceae	*Triadica sebifera*	-	N	Chinese tallow	2015
Fabaceae	*Albizia julibrissin*	-	I	Persian silk tree	2016
Fabaceae	*Amorpha fruticosa*	-	N	False indigo	2015
Fabaceae	*Apios Americana*	-	N	Wild groundnut	2015
Fabaceae	*Centrosema virginianum*	-	N	Butterfly pea	2015
Fabaceae	*Cercis canadensis*	-	N	Red bud	2015
Fabaceae	*Desmodium perplexum*	-	N	Desmodium	2015
Fabaceae	***Erythrina herbacea***	**P**	N	Mamu, coral bean	2015
Fabaceae	*Indigofera tinctoria*	-	I	True indigo	2015
Fabaceae	*Lotus* sp.	-	I	Lotus	2015
Fabaceae	*Medicago sativa*	-	I	Alfalfa	2015; 2016
Fabaceae	*Medicago sp*	-	I	Alfalfa	2015
Fabaceae	*Mimosa púdica*	-	I	Touch-me-not	2015
Fabaceae	*Mimosa strigillosa*	-	N	Powderpuff	2015; 2016
Fabaceae	*Pueraria lobata*	-	I	Kudzu	2015
Fabaceae	*Sesbania drummondii*	-	N	Rattlebox	2015
Fabaceae	*Sesbania herbacea*	-	N	Coffeeweed	2015
Fabaceae	*Trifolium repens*	-	I	White clover	2015; 2016

Fabaceae	*Vicia angustifolia*	-	I	Narrowleaf vetch	2015
Fabaceae	*Vicia sativa*	-	I	Garden vetch	2016
Fabaceae	*Wisteria sinensis*	-	I	Wisteria	2015
Fagaceae	*Quercus acutissima*	-	I	Sawtooth oak	2016
Fagaceae	*Quercus nigra*	-	N	Water oak	2015; 2016
Fagaceae	*Quercus virginiana*	-	N	Live oak	2015
Gelceminaceae	*Gelsemium sempervirens*	-	N	Carolina jasmine	2015
Geraniaceae	***Geranium carolinianum***	+	N	Carolina geranium	2015; 2016
Geraniaceae	*Geranium dissectum*	-	I	Cutleaf geranium	2015; 2016
Hamamelidaceae	*Liquidambar styraciflua*	-	N	Sweet gum	2015
Iridaceae	*Iris pseudacorus*	-	I	Yellow flag iris	2015
Iridaceae	*Sisyrinchium* sp.	-	?	Small yellow eyed grass	2015
Iridaceae	*Sisyrinchium minus*	-	N	Dwarf blue-eyed grass	2016
Iridaceae	*Sisyrinchium rosulatum*	-	N	Annual blue-eyed grass	2016
Juglandaceae	*Carya glabra*	-	N	Pignut hickory	2016
Juncaceae	*Juncus biflorus*	-	N	Bog rush	2016
Juncaceae	*Juncus effusus*	-	N	Common rush	2016
Lamiaceae	*Lamium amplexicaule*	-	I	Henbit dead-nettle	2015
Lamiaceae	*Lamium purpureum*	-	I	Purple deadnettle	2015
Lamiaceae	*Salvia coccinea*	-	N	Scarlet sage	2015
Lamiaceae	*Salvia lyrata*	-	N	Lyreleaf sage	2015; 2016
Lamiaceae	*Stachys floridana*	-	I	Florida betony	2015
Lamiaceae	*Teucrium canadense*	-	N	Canada germander	2015
Lauraceae	*Cinnamomum camphora*	-	I	Camphor tree	2015
Liliaceae	*Nothoscordum bivalve*	-	N	Crowpoison	2016
Lygodiaceae	*Lygodium japonicum*	-	I	Japanese climbing fern	2015
Lythraceae	*Cuphea carthagenensis*	-	I	Colombian waxweed	2016
Lythraceae	*Lagerstroemia indica*	-	I	Crape myrtle	2015
Magnoliaceae	*Liriodendron tulipifera*	-	N	Tulip tree	2015
Magnoliaceae	*Magnolia grandiflora*	-	N	Southern magnolia	2015; 2016
Malvaceae	*Modiola caroliniana*	-	N	Carolina bristlemallow	2015
Malvaceae	*Sida rhombifolia*	-	N	Cuban jute	2016
Moraceae	*Broussonetia papyrifera*	-	I	Paper mulberry	2015; 2016
Moraceae	*Maclura pomifera*	-	N	Osage-orange	2016

Oleaceae	*Ligustrum japonicum*	-	I	Japanese ligustrum	2015
Oleaceae	*Ligustrum lucidum*	-	I	Glossy privet	2016
Oleaceae	*Ligustrum sinense*	-	I	Chinese privet	2015; 2016
Onagraceae	*Ludwigia decurrens*	-	N	Wingleaf primrose-willow	2016
Onagraceae	*Ludwigia glandulosa*	-	N	Cylindricfruit primrose-willow	2016
Onagraceae	*Oenothera biennis*	-	N	Common evening primrose	2016
Onagraceae	*Oenothera drummondii*	-	N	Beach evening primrose	2015
Onagraceae	*Oenothera speciose*	-	N	Pink evening primrose	2015
Onagraceae	*Oenothera. pilosella*	-	N	Yellow evening primrose	2015
Oxalidaceae	*Oenothera stricta*	-	N	Common yellow oxalis	2015
Oxalidaceae	*Oenothera triangularis*	-	I	Purple oxalis	2015
Oxalidaceae	*Oxalis debilis*	-	N	Pink oxalis	2015
Passifloraceae	*Passiflora incarnata*	-	N	Purple passionflower	2015
Phyllanthaceae	*Phyllantus urinaria*	-	I	Chamber bitter	2015
Phytolaccaceae	*Phytolacca americana*	-	N	Pokeweed	2015; 2016
Pinaceae	*Pinus echinata*	-	N	Shortleaf pine	2015
Pinaceae	*Pinus glabra*	-	N	Spruce pine	2016
Pinaceae	*Pinus palustris*	-	N	Long leaf pine	2015
Pittosporaceae	*Pittosporum tobira*	-	I	Pittosporum	2015
Plantaginaceae	*Plantago lanceolata*	-	I	narrow leaf plantain	2015; 2016
Plantaginaceae	*Plantago major*	-	I	Broad leaf plantain	2015
Plantaginaceae	*Plantago virginica*	-	N	Virginia plantain	2016
Plantanaceae	*Platanus occidentalis*	-	N	American sycamore	2015; 2016
Poaceae	*Alopecurus carolinianus*	-	N	Carolina foxtail	2016
Poaceae	*Arundinaria gigantea*	-	N	Giant cane	2016
Poaceae	*Briza minor*	-	I	Little quakinggrass	2016
Poaceae	*Chasmanthium latifolium*	-	N	Indian wood oats	2015
Poaceae	*Cortaderia selloana*	-	I	Pampas grass	2015
Poaceae	*Cynodon dactylon*	-	I	Bermuda grass	2015
Poaceae	*Equinochloa* sp.	-	I	Water grass	2015
Poaceae	*Paspalum distichum*	-	*N*	Knot grass	2015
Poaceae	*Poa* sp.	-	?	Bluegrass	2016
Poaceae	*Setaria pumila*	-	I	Yellow foxtail	2015; 2016

Poaceae	***Sorghum halepense***	+	I	Johnson grass	2015
Poaceae	*Stenotaphrum secundatum*	-	N	St.Augustine grass	2015
Polemoniaceae	*Phlox divaricata*	-	N	Woodland phlox	2016
Polygonaceae	*Persicaria hydropiperoides*	-	N	Swamp smartweed	2016
Polygonaceae	*Polygonum punctatum*	-	N	Dotted smartweed	2016
Polygonaceae	*Rumex* sp.	-	?	Dock	2015; 2016
Polygoniaceae	*Rumex crispus*	-	I	Curly dock	2015
Polygoniaceae	*Rumex verticillatus*	-	N	Swamp dock	2015
Ranunculaceae	*Ranunculus pusillus*	-	N	low spearwort	2016
Ranunculaceae	*Ranunculus sardous*	-	I	Buttercup	2015
Rosaceae	*Dushesnea indica*	-	I	Shrubby Cinquefoil	2015
Rosaceae	*Eriobotrya japonica*	-	I	Loquat	2015
Rosaceae	*Rubus* sp.	-	N	Black berry	2015; 2016
Rubiaceae	*Diodia virginiana*	-	N	Buttonweed	2015
Rubiaceae	*Galium aparine*	-	I	Stickywilly	2015
Rubiaceae	*Galium obtusum*	-	N	Bluntleaf bedstraw	2016
Rubiaceae	*Galium tinctorium*	-	N	Stiff marsh bedstraw	2016
Rubiaceae	*Galium uniflorum*	-	N	Oneflower bedstraw	2016
Rubiaceae	*Sherardia arvensis*	-	I	Blue fieldmadder	2016
Salicaceae	*Populus deltoides*	-	N	Eastern cottonwood	2015
Salicaceae	*Salix exiqua*	-	N	White willow	2015
Salviniaceae	*Salvinia molesta*	-	I	Giant salvinia	2015
Sapindaceae	*Acer rubrum*	-	N	Red maple	2015
Scrophulariaceae	*Mazus pumilus*	-	I	Japanese mazus	2015
Scrophulariaceae	*Nuttallanthus canadensis*	-	N	Canada toadflax	2016
Smilacaceae	*Smilax bona-nox*	-	N	Saw greenbrier	2015
Solanaceae	*Datura stramonium*	-	N	Jimsonweed	2015
Solanaceae	*Physalis minima*	-	I	Pigmy ground cherry	2015
Solanaceae	*Solanum carolinense*	-	N	Horse nettle	2015
Solanaceae	*Solanum nigrum*	-	I	Black nightshade	2015
Ulmaceae	*Celtis laevigata*	-	N	Sugarberry	2015
Valerianaceae	*Valerianella radiata*	-	N	beaked cornsalad	2016
Verbenaceae	*Callicarpa Americana*	-	I	Beautyberry	2015
Verbenaceae	*Clerodendrum bungei*	-	I	Mexican hydrangea	2015

Verbenaceae	*Lantana cámara*	-	I	Lantana	2015
Verbenaceae	***Phyla lanceolata***	**P**	N	lanceleaf fogfruit	2015
Verbenaceae	*Verbena brasiliensis*	-	I	Brasilian verbena	2015
Verbenaceae	*Verbena rigida*	-	I	Prostraste verbain	2015
Viscaceae	*Phoradendron leucarpum*	-	N	Oak mistletoe	2016
Vitaceae	*Ampelopsis arborea*	-	N	Pepper vine	2015
Vitaceae	Broussonetia papyrifera	-	I	Paper mulberry	2015
Vitaceae	*Cayratia japonica*	-	I	Bush killer	2015
Vitaceae	*Parthenocissus quinquefolia*	-	N	Virginia creeper	2015
Vitaceae	*Vitis rotundifolia*	-	N	Muscadine	both

Appendix 2 Location of collections of plant Species with endornavirus-Like dsRNA within Baton Rouge, LA

Species	**Coordinates Where Collected**
Alternanthera philoxeroides	30.411083, -91.172222
Dracopis amplexicaulis	30.413306, -91.171417
Geranium carolinianum	30.407556, -91.169722
Hydrocotyle prolifera	30.411833, -91.171167
Hydrocotyle umbellata	30.410722, -91.172194
Sonchus asper	30.409861, -91.153694

Appendix 3 Accession number of endornavirus sequences used in phylogenetic tree

Virus	**GenBank Accession No.**
Hordeum vulgare endornavirus	YP_009212849.1
Rhizoctonia cerealis endornavirus 1	YP_008719905.1
Phaseolus vulgaris endornavirus 2	BAM68540.1
Cucumis melo endornavirus	ARI71634.1
Lagenaria siceraria endornavirus-Hubei	YP_009351891.1
Lagenaria siceraria endornavirus-California	YP_009010973.1
Phaseolus vulgaris endornavirus 1	YP_009011062.1
bell pepper endornavirus	AKP92841.1
hot pepper endornavirus	YP_009165596.1
Psophocarpus tetragonolobus endornavirus	YP_009305414.1
Oryza sativa endornavirus	YP_438200.1
Oryza rufipogon endornavirus	YP_438202.1
Ceratobasidium endornavirus H	AOV81686.1
Rhizoctonia solani endornavirus	AHL25280.1
Tuber aestivum endornavirus	YP_004123950.1
Persea americana endornavirus 1	YP_005086952.1
Alternaria brassicicola endornavirus 1	YP_009115493.1
Rhizoctonia solani endornavirus 2	AMM45288.1
Grapevine endophyte endornavirus	YP_007003829.1
Grapevine leafroll-associated virus	AFH35871.1

Appendix 4 Nucleotide sequence of Geranium carolinianum endornavirus-1 (14,638nt, 4815 aa).

TATGGCAAATTCGGTTAATTCCCTCATTAGTTACAATTCAGCTAATTTATACAACTATATCAACAAGGTGCTCT
TAGAAATAGAGTCGGGAGTGCCAAGAGAGAAAAAAATTTTAGAAAAAGGAAGCAAAATGACTACCAAGTGCCTA
GCACGCACCCCGGTGACTAAAAATTATGTTAGTTTAATATCCAAAAACAAAAAGGAGAGCCGAAAAGGATGCTT
GACTGCCAAATTGCCAATTTTCAAACCAGATAATATAAAGGAAATCTTAAAGTGTGTTAGCAACACTTACATGA
TATCTAGACATAAAACCAAAATAATACCAGTCACCCCTTCCATAGGGATGGGTGAGTTTTTTTCCATCATGATG
GACGGGTTAGCGTGGGATTGCGCGAACAATTTTGGTCTAACTACATTCCTACTAGGGGTTATACAGGGAGGCAA
GGTATTGAGTAATCAAGTGATTGGCAACCACAATTTTAAAAGTTCCTACCGTTCTTTCAATGAGAAGGTTACGT
CTCTAACTGCAGACAATTACAAAGGTAACCACTCTGCAGAGCGGCTTAGCCATCAACTAACTAGGAATATAAAC
CTTAGACTGGACTATACGGGAATCCGGGAATTAGCTGTGTATATTGGATTTGATTACTTGTTATTTCCTGAAGA
AGACTATAATGCTTCGTTAGGCGACAACTTCCCCTTGGATCAAACGGAAAATTTTCTTAAGAAAAGTTTAGAGT
GGTGCCGTCCTTCAAGATATGCTGAACTACTAGGAGAAGGCTATTACACTGAATTTGTGACAGTTGGAAAGGGC
TCAGCGATCCATAATCTAAATAATAACGTTTTTATTTGTGTTGCCTGTGGTTGTTTAAACTCTAAGTGCGACGA
TGAAGTGCATACTTGTGGGAAATGCCTTAGTATAAATGCCACCGTTGACATTGACGTCACGGAACTTAAAAATT
ACAAGTGTCTAGTACTAGTACCTTCATGTTATGAGTTCGATAGGAGCATAAATGGCCCACTAGCAGCCCTTAGC
CTTAAAAATTATTTGGATGGAGGGCTTTTAAACCTGCCTAAATTAAATAGTGAACTCATAGCGCAGGACGTCCT
GCGTTATGTCAAAACACGATCTAGACTTGATGTTAAAAATCAATATGTATGCGCCAATTTAAGCAGAGACGAAT
TGAATTATTTGAGAGAAGAATTCAAGAACCTAGAAATAATAGTAAGGAATACTTGGCTCGATATGCAAGGAATG
CTCATGACGGAAAGTCATTGCCACTTGACCACCTTATTAAATTCGGAAAAAAAAAATAACAGAAGTTCAAGGGTT
CAATAGATCATATACTAATGTAGGGACTAGTGATACTAAGTTTAGTGCCGCCAACCACTTGACTAACTGGCAAG
AGAGCAACGGAGAGATGGCGAAAAATCCACAGGTGATGAACTACAATTCCATCCCGAAGATGGAAAAATTGCTA
ACCATGGCATGTGCGAGAAAGATTTACGTGGTAGTTCCTAACGTGTTCGACCAATATAACGGATTCTCAAACGG
ATTTAGTTTCGAAATAAGCAATGAACGGGGGCTAGTTAAGATTCTAATAAATGGCACCACGACAGTATTGGAAT
ATACGTTGGAGCAAATGAAGCTGTTACTAGAGTATGATTATATTTCTTGTGACGACAAACTATTCGAGGTTAAA
CTTATAAAGAAAACTTCCAACTGCTCCCTAATCTCAATTTCGAAACTCAGGAAAAATTCCTTAAACGTGGATGA
ACTAGGATATAAATCCGTCCAACCAAACAAATTTAAGTTGTTCACCCTAGAAATTCCTGACTGGCAGCAAAATG
TGATGGGACTGCAGTTGGGACCTATGATTAAACGTAGAATAAAGTTTAATATGAGATTCCTTAAACACCTAATT
ACCAGATGCGAGGCATGGCCTGTCAGTTTTCATGGGCTTAGGGAGTATGCAATAGTTTCTAGCTTTTCTAGGTC
AGAGAGTAATGACATAGTCAAAAACGTATTTGAACTAAACTTTGAGGACATACCCGACCACGTTTTCTGCGCAT
ATAATATCTACCTACGACAACAATTAAGCACACAATTCACCCACTGGCTAACCACAGAGAGGAGCCTGGGGCTT
GATAAACTAATACAGGGCTTTGCAGGGGGAATAGTCGGCCTGATAACAAATTTGATGAATGACCACCAAAACGA
CAAGTCTTTTATAAGCCAACTTCTTGACAAATGTAGTGATCTCTCGTGGCTTGTCAGTGCACAGTGGGATAAGA
TAGAAGAATCAATCAAAACATGGGAAACCGACACTGTCAAGCTGGCTGACTTCAAAGGAAAGATCGTCGAATCA
TTCGTAGACAAGTTGGAAACATGTCAACATGGAATGGAAAAAGATCTAATATCTAATGGTTGCAATTGTTGTGG
AAGGAAAACAAATGAATTAAGTGGTTATTGCTCAACGTGCAACTTAGAAGGATCGTGTTCACATCCTTGCCTGC
ATAGATGTAACAGTAAAATTGAACACTTTTGTGAGGGATCTGTTACCAGACCAGATGTGGCCGTGGGAGACAAC
CTAATATGTGGCCACATGGTAGTCACTTGTAAGTGCTGCAGAAAGCCATCTTGTCAAGAGCTATGTTACAAATG
CTTTGAATGGGGTCAATATGAAGACAACTTAACCAGAATGGCTGTAGTTGATCAAACAATATACAAAGGGGAAG
GAGCTGAATCTGCTGTAAGAAGAGTGATCAAACAAGCCAAAGGGAAACAGAGCTACCCCAAATCCATTGAAAGA
AAACGTGTTTTGTACAAGAAAAACACAAAACAAGCAAACAAGCAGCCCCAAAAGGGAACGACAACAAGCGTGAT
CACTGACAAACCTGGAGATAGTAGTCAGAAAATACAGGTACAGATAACTAAGAGGCAAGAAAGGCAATATTTCG
GATTAACTTCCGACAGTTCGGAGGAAGACACAAAGTCTGATCCTTTCAAGGAGAGAGAAGTAAATCTGAATTTA
GAAGACGAGGACCACGACAAATGCAGTGAACCCGAGGAACCCGTAGACGACACCAACGTGGAGAAACTTAAACA
GCACATAATGATGAATGAACTAAGGACAATAACCAATGACGAATATTGCAACATAATATGTCAAAGATCCAAAC
CAATGGATGAATTACTAGGAGAGTCAATAAAATTTAAATTCCTATACCAAGGGGAAGTCTTCGTTGATCCTACC
AGCATTAATTCAATACGACCTATAATAGTGCCTTACACAATTGGATTTTGCCTAAGGGATACTATGGCATACTA
TAATCCAATTATAAATGATATTACATGGGCTGACGCTTGTTCCGATACGGGGCTTGACGAATCCTGGTGTATAT
TGAACGACGTTTGGAAATATGCAGAATATTTCGAAATGAACGTTTTAATAATTCACGAAATGGAAGGCGAAGTA
GCCGGAGTTTATTGCTATATTCACGATAAGTATGAACAGGTTAACATGATAAGGTACTCTTCTGAAAAACCCGG
GAAAGAGGATTATGAAGAATTGAGGGGCCATTATGAACCCTGTGAAGTTGGGTTTAATAAGGAACCGAGTCTCC
CGCCAGTTTATGCCCCTGACATAACTTGGGAAGACGTGAACCAGGTATATTTCAATGTAACAGGGGGAGGAGAT

TTAGGAGAGTTTTACGACAAAAACATAGAAGATAGGCTAACCATAGCGTTGGCATTAAATGAGAACAAGGGATT
ACAGTACGCTAGTTCCGGCTTCCCTAAGATAGAGTTGAGTAACAGAAAATTAGGGTATATGATTTTCAACAACG
GTAATAATCTGCATGAACCCAGAAAGGGAAAGCTAGCAACATTTATCAAGGAGTCCACCATAGGTTCACAGCTG
AATATACCTAGCCTTCTCACTAGAGCCGAACTCCAAGAATGGTACCTGGATAATCTAGTTAAAGAAGAAGACCT
CAGCAATGATAAGGATTGCGTAAGGAATGCCATAACCATGATAATTAGCCAATATCTAGATCTTAAGATGAGCT
GCGAGGAAACATTTAATAATCAAAATGCGGACAATTCCAATATAGCTTTGCTCAAGAAGGGAATAAAGATAATA
GACAAGGGCAATTATTCGGTCATATTGGCTGTAGATGGACTAGAAAGGCTCAAGACAGGCGACGTAATAATGAT
ACGTCGTGGCCCTAATAGATTTTGCTGCCAAGTTGAAAGAAACATCAAAAGAATAATGATTCCTAAGATGCCAG
GACACGGAACCAAGTTACTAATAGACATTGCGTTGTTCAAGATTAGTTATACTTCCTTGATAATACAATTAGCA
TCGGTGAGCCGTCCTGGAATAAGTTTGGAAAGAGCTAAAGAACTATTGGGCAAAGCTACCTGCATCTTAGGATA
CCCAGGAACCGGGAAAACAAAAGAACTAGTTAAACGATACGAAATGAATCCTGGCTTATTGGTGGGGGTCACAC
GTGGCAGTCAAGAAAGTCTAATACAGGAATTGGGAGCTAGAAGCAAAATAGTATTCTCGGCAGAACGCGCCATG
ACTAACAGGGCATCTAGCAAAACCGTGTACATAGACGAAGCTTCATTAATAACGTTGCCGCAATTGTGTTGTAT
GCTAACCCCTTTAGTTGAAAATTTAGTAATATCAGGAGACTTGGCACAAATTCCAGCAAAGGAATTCTCCAAGG
TTTGCGGATATCAACCCACCAACATTCTAGAATTTAGCAAAGATACAGGAGCCACGAAAGTAGAGCTTAAAAAA
ACCTGGAGATTTGGAAAGCGAATATGTTCAATTCTTAATGAAGCTTTTGGATTGGATATGCAGTCCGCCACAGA
AAAAGAAACTAATGTTAATCTAGAACACTCCTGTGGAATAGATAAAAACAGCCTGAACAGAATAGTGAAAGAGA
GAGACATAGACACCATCATGGTTTTTACTACTCAGGTGGAGAGACAAGTTAATTCATTGCTAGAACCTGAAAGC
CGTGTCAGGGTAGCAAGAGTTCATAGTTCCCAAGGAAGTTCATTTGATAGAGTTCTAATAGTCCAAGACTATCG
GAAAGGTCCTGCCCAGGGGTCAGAGGAGGTTCAATTCAAGAAAGAGTATGTCATAGTTGCTATGACCAGGTGTA
GAAAAGACGTCACCATACTTTGCACCTACGAGTCATGTAAATGTCGGGAAACATCAAATGAAAGCATAGCCAGG
CATTTAAATCGAGACCTAGGACTACAGTACTTGTACAGAGGTGGCAAGAGTAACAATGTGGATGTGATAGCAAT
CCTAAATGCCGTACAGGATAGGATCTCGGGACTGTTCGAGAGCGTAGAACCCAACCTAGTGAAGGATTGGCTAC
AAGAGCTATGGAAAATAGCAACCGGAGGTATTAAGAAGAAAATGTATATGAGGATGGCGACAATGATCAACGAA
TGCAATGACCGTGAAGTCCTTAAAACATTACTTGACTCTGGAATGCCCATGGTTTCAGACGTAATTACAGAGAA
TGGCAAACTCTATGCTATAATGAATTATGACAACACAAATTGGAACTATGTAAAGAAAAAGACCATGACTCTTT
TCAACAGTAAACAACTCATAGAGTGTTGCAACAACAAGATCCTGATCGGAGGAGTGGAAGTAATTAGCAACCAG
GGATCTGAGGAGATAATAGTGGATGCTAATAAATTCGATTCATGGGAAAGCACTCAAAAACACACCAAACCCAT
AACTATACGCGGGTACCAGCGGAGAATCAATTCGGAAAATTCAGGCAACATTAACATGAACGTGAGACTAATAA
ACCACTCTAGCCAACTTTGCTGGAACGCTGCTAAATTAACTCTTGACATAGAGTATTCCGGGACGAAGTACAGC
ATTAAGCCTACAACCGGCTGCTCCTTGTGTGGAGGAATACAAATCACGAAACAAAACGGCGAGCTCATGGTATT
CATCAACAACATGTACGAAAATTATAGTAGTAGAGATATACAGTTTAAATCCGGAACTGATCCGATAACCAAAT
ACCTATTAGGAAAATGGGATTTAGACCCAAGAGATGATCACCTATGGGAGCTAACAGCTGGGCTCTTACCTAAT
GTCAATCACAATGCACTACACTACACTCTATGGATAGAAAGGATAATGGCTGCAGTAAAAGGGATAAAGAACGG
CAAGACATTTAACACACAGGAAGGACACTTCTTCAGAAACGAACTTGAACTCAACGAAAGGATACTAAGTAGAT
ATAAAAATATAGCGAATGAAGCAGGGATCAAGATAACTTGTGAGAGTGATAGAAATTATTCTTACTTCAAGAAT
CTATCTTTCTTATTTCCAGCCAAATTCAAAGGCAATAGTTGTTACGTGTATTTCCATAAACATAAGAAGTGCGT
ACTAGTTAACAAGACAAAGAAATTTGGTCTGTCTAGGGAATTATCTCCAAAATTGTGGAATGAGGAACTATACA
GGCTACCATACAATTTATCCCTTATGTTCGGAGGATCTGACCTGCAAATTAGGGGACACAAGGAAGTAATTAAC
ATGCTCGACCTAGACCTGGAAAAACACGACCACAACAGAAGTAAACTAATGCTCATGATAGATGAAGATTTAAC
TAAGTTGGCTAAGGACAAGAACTTCGCCAACCCGGAAATTGCCATCCCAAGCAGTCTCATAACAGATGGCAAAG
AACTAGGATTGGCAAGTAACTTTAACATAATACCTGATTTAAATTTAACTGGGATAGGAGCTTCGTATTTGTTA
GATTCGGTAGCTGCAAAATTATTTAGTCTTAACCTGATGGAATCTGGAAAGACTTTCATTACGAGATATTGCAA
CTTATCATTGAGACAAGACATAGAAAGACACATCATGATCAAACCCGTTGATACCAAAATAACCTCTTACAAGG
ATAGTGAAACTTACCAAGATTGCTTTGCTAGAATACTTAGCAAAAAAAATACAGTATCTGAGATGGCCAAGGCA
ACAGACAACTCTGAAACGAAACTCAAACACCTAAGAATTGAACAAATACTACAGAGCATGTTAGAAGGACGTGA
TTTATCAAGCATAATTGTAGAAAATGCTGACACTTGTAAGACTAATATAGGATGTCTTGGAGTGAGTTGCTTGG
AATTTGGGGACATGGAAATTAACAAAATAATGAGAGAATTAAATTGCAAGAGGCTAATTGGCTTCATACCTGAC
CTAAGAAACAAGTCAATTAGGGAAATAGTGAGTTTAAATAAAAACCATTTACTGTTCAAAGGAGACTCCAGAGA
CTATCTAATAAACCCAAAGTGGGTACAGTTGGTTAACATCCTATATAGTAGTGAACGATGGGAATCTAGCATGT
TGCTTGAAAATCTTAAAATAGTAGGTCAAACCGACTTATTCTTAATAGTCGATATAACCGATACTAGTGAAATT
ACAGTAGGCAAAATTCCGGTTTCAATGCATAATGATGAAACCGTAGTAACCGTACCTCAACTAAATCCCATAAA
CGAAATACGAAGAACCGGAAATTTGTTCAATGCCGTGGAATTCGTCATAGATAACGAAACATTGAGCAGGTTAG
TTAGAAGAGCCATGACTCCGGGATGCACGTTGCCGATGTTACAGACAATTGCTAGGAATAGAATGCAAAGTACA

GTAATAACAAAATCCGGAAGAAAAGCAGGAGGCAGAAATGTAATAAGGGACGTGTCCTTGTGTGCTCTCGTTGC
CCAATACATCGCTAACCATAACGATAACCAAATAACTAGGTATCTGGAAAAGATAGATGATTATTTGATGGATA
ACAAAGACTGGAAGAGGTCGTCTGGGGCATTAATACACATGCTTAAGCTAGAAGCGAATACCATAATAGGTAAC
GCACTGAATATTAAGGTGTCGCTTAGCGAATTATCAAAGGTGCTGGGTGAGGTTACATACAATGTCATGACCGA
TAGATCCAGTAAAAAGGGCATCCCCAGCATAAGAGTAATACATCAACCCCATAGAATAATGTATGGGGAATATT
ACTATCATGGACTCGATTCGAAAAATCCGAACTCATTGAACCTTGATCCAAATGTGATAGGCAATATACTCAGG
AATTCCTTTAGATCATGGACTAAAAATGTGGAATCGAAATTAGCCGGCTACTGGGCTGAACCTTCTGTCATAAA
TTCACGATTATCAGACCACTTGACTCCCGATAAAGAGTATAGCGTTAAGTTACTGATCTATGACATGTCCGGCG
GAATAGGAAGAATTCTTCAGAATAGGCTTAAGTTAACTATGAAAACAGACAGCGTGATACATAGCTCAGTATGT
GTCGGCGATAACGAAATTTCTTATGGAAACGGAGTAAAGATAAGCCCGTTGGGATCGCAAATGGTTGGTAAGAC
ATCGAACCCTGTTACATTAGGCAAGATCAAGTTGACTGCCAAAAACATGGAGGAAATAGATAAAATTACTAGTG
AAATATTTATTCCGCGTAAATATAGTCCAATTGGATTGAATTGCAACTTATTCTCGCTTTGGCTCCTGATACAG
TTTGGCTACATGACTAAAATTAAAATAAGTGACAGTAAACTAGAACACTTGGAAAATTTGGCAAGCTTAGTTCC
AGAATTCGGGTCTAAGGTTCCAGAAAACGTGAGGAAATACATAATAGCCTTAAATTCCAAGGTCATGGGAGACG
AGGAGCTAACGGTTAAAATAATGAGGTGTTTTGAGTACACTGTTAAAAGTGAAAGACCCGCTGGAAATAGCAAA
TTAAATAACTTCATAGCACATCAACTCATAATACTAACAGGAAGGAAATTGACAGACCTTAGAATGCAGATACT
TGAGGATGAAAAATCGGAAAGCGAAAGCATGGATGATGATAGCAGACACGACAGTGACTCTGACAGTGATGATG
ACCAAGACAATGACGGGAATAACGGAACAGAACAACAACTCAATGAAGCTGAGGATTTAGAACCGGAAATTGAG
GAGGGAGATAGAGACGCCCGGTCATCTAGTGAAATGAATGAGGACAATGACGAAGAAAATCAAGAAAGCGATTT
CGAGTCCGCGCACTCGGAAAACATGGACGAGGAACAGGCCTCTTGCAACGAAGCAACACTCCCTAGCAATGAAG
AGAACATGCCTTCTCGTATTGATAACACCCCACAAATCGAACGAATTAATACAGGAAACGACACAGAGTCCCAA
TCTAAAACTATTGATAACATTGCTCAAGAAATAACTACAGAAGAAAAGAATACTGACCAAGAAGAAAACGATAC
TGATCGAATAATGGCACCAGATTCTAGGACAGTTTTGGACTTCATCAAGGAAGAAATGGAAAAATTGGGCATAG
ACAAAGTGCCGACCGCAATGTCCACGGCATCCAACGCTATAGGCAAGTTATTTTCTGATGCCGAAATAAATCCA
AGAATAGACCTATATCAAAACGTAAATAAAGAGTTAATAGACAATGCACTATGCATCTGCCAACAGGCCAAAAA
AGACTTCACTGTAACCAAAGATGATCAAAACGGATTAATGAGCAAAAAGGAAAAAGTCGGCACTTACAACGAGA
GACTTAGGTTTAAAATATTGAATAAATTCCAAAAAACGTCAATTTGCGTGGTCGCACTAGGAAGCACTGGTGAC
ACTCTTCCGGTACTAAGTGCCTGCAAGATGCTCAAGCTAGGAGGCGCGTGGATATGCCTGCTGTCACACCCAGA
CATATACGGGTTAGACAACAGCAACCACGATAAATTCATTAAAATAAATAAATCACAGAAAAGAACTACTGGAA
ATATCAAGAGTGACAGTGCCATTGATATCGCGAAACACGCCCAGAGCCACAACATTGAAGCATTACGAACTTTC
AAAGAAGCCACGCAAAACCATGACTTTGATTTAGTCCTTAGTACCCCTCTCGCCCCAGCTGTGACTGGATATTC
TATGTTTCTTGGCTTGAGAACTGCCGAAGCATTCTGCACCTATTGCTGGAACACTGGAGTCGAACAAGGAAATA
GTGAAAGCGATAGTTTTCTATTGCGTTGGTTTGGATTCACGCTAAAATATGCGGCAGTAGACGGGACTCTGCAG
ATACTAAGACAGAGTTTAGCAACTTCAATGCTGAGAGAAATGAACATTGAAAGTCCCGATATTTCAACAGTACC
TAGAGTGGTGCTCAGTTGGGAAGACGTTCATTCTGAAAAGAACATCAAGGATCCTAAGTCTGTATTTATAGGAT
ACACTAGTCCTGGAATGAAAGCAAATCTGCTGAGTAACGACCGGAGTTTCAGGCTTCTCGTGGGATTCGGTAGT
ATGCAGGTCAGGGAGGAACAAATAAACGAAATCAGAAAAGTTGCTAAGTTAATGTCTGGAATAGAAATGATTGT
CACAGTGCACATACAGGATGAGTCACTGAATAAATTGTTGCTTATCACAGTTAAAAACTTATTTCCGAAATGCA
AGGTGCTACTGGGCAATGTAAACCTTGGGGAAATAGTAGCGAATAATGATGCCATGGTGTGCCACGGCGGTATA
GGGACAGTCCAAGAATGCTTAATGGCTTGTTGTGTTCCAATCATAATTCCTTGCTTTGCCGATCAGCCATACGT
GGGCTCCAACTTAGAGAAAAACCAGATTGGAATAATGGTCAGTGACAGCGAAGCCGAACTGACTGCCAAACTGA
AGAAAATACCAATGATGCAGCAGAAATTAAAAAGGAAAAATTATTCCATGACTGATAGCGTTAGAAACTTGACA
GATGCGGTGCTTGATCTAATCGGATCCCAGATTCCGATTTCACAATTAAGAGAAACAGGGCAAAAAGTAGATCG
TAGGTTTGAAGTGCCAACCGGAGTGGTAATACCATATCTAACCATCCCCAGTGAGTCAATTGCACTATCATTAA
CACCAGGAGAGTACCAGGTAAACGGGGTTATAAACGAAGAAACAGTATACAAGATAGGGGAATCCTACTACGGA
GGAGAGTGCTTTAAAGAGGCATTCAATAACGGCATACTCAGACTAAGAGGTCAGGAATACCACATCAGGGCAGT
TCATAGTACGGCATTAATGACCATTGAAACCACTACCGATATCCCAAAATTGAATATATTAGGGTTCTACAATC
ATGTTAACATCCAAGTCTTGGGGCATACAAATAAAACAGTGATCTTTAATGATAGCTGGCCTTTGCTGTCGATA
TACGTGACTAAAATCAGTGAAAGGAGGAAACACAACAGTCTCCATGCATTCATAGTAGCAAATAGAACCGACAT
AATCCGGATTGAACATTTAGCGAAAACTGGTGACAAAATTGCGTCAATCATTGACAACAAGTCTTTGTGTAACC
GTCTGTCATTACCAATAGGTGTCGATCCCAAGATGGCAATTGATTCAGTTATAGGCTGTGATAATAAATTAAAT
GAATGGGACTGGAAATGCTATGGCAGTTTTGAGAACCTAAGAAATAGACTTAAAGGAGAAAACATAGATGCTCT
ATTACAGTCTGATTCTTCTATAGCCATCCCTGTTTTAAGAACAACACAGTATGTGGACGGAGCTTGTTTTGAAA
GCAATGTCAACACCCGGTCCCTGATTGGAGAAGTTGTTTATTGTTTCACAACGATGGGAGTAGTGCCGGGAATT

GTAGTTCGGCAAAGCACTAGTAAAATATACGTGATCACATCACAATCCGTAACTCACCTAAGTGGGTTGATCAG
GTGTAGGTTGCTCCGGAAAAAAACTGGTATGATTGCAGGAACAGAAATCGACAAATTAATGAAAACTAGCTACA
GTTACGCCCTAAACATAGAAACGATCAATAAAATCAAGGAGGTGATACCCGGATCCAACCCTAAACTAATGACT
GCCGCAATACAGGAAATGATTTCCGAATATGTAGTGATATCAACAAATTTTGATACAAGGTTTCATCATTCAGA
TAGAAGTAACTATGAGCGGAAAATCCTGATGGACTTGATAGGAACCGACAAACTGAAACTATTACCAAATGAAC
TAAATGGAAATCTGGCAGCGATAGTTAAACTAACAAACGTAGGATATTATAACTATGCAGCAGGCTGTTTATGG
TACAATATTGAATGCCGTGAGAAAGACTTGTGTCTTTCATTATCTAGAATAGCACAAGAATGCGTGAATAAAAG
AGGTAGTTTCTTCAACAGCATGCTCAACTTTAGAATTGAAGCAGAGGACTCAAATGAATTTAGAGATTTCATCA
AAATAATGAATAACAATTTCGATTTAGTCGACAACGACATTCTAGCGAGGAGCAATGAAAACATACAGGATGAA
AAGATAATCAAGATAAATCTTGAGTGGCTTAATAACAACCTGATTGTAACAGAAAATAAGATAGCCAAGCTTGT
GTTTGCTGAATTCCTTGGAAAGGAGAAGGAGGTAGACGCCATTTGCCAGAAAGCCGTTGAAAACTCAGACATCA
ATTCTTTTCAGTTAGTGAGTGGAGTTTGTTTATCAAGGTCAGGATTAACCATACTAAGCTTGAATCCAATGGTT
AGAATAACTAAAATTAAGATTAAAGGAGGTGCAGAGAACAAGGACGACAAGACCGACTCCGACATGGGAAGAAT
ATTCAACATACAACCAAACGAGACTTTAATGGATCAGCCCTTGGAAATGGTTAGGGGCCCAAAGTTAGGAAAAG
AAAATTTTGGCCGCTATGACAAGTGGCTAAAGCAAAATCCTTATTTGGAGGAACCAAAAAATGCTGATAAAGGC
AAATCTATCATGTCCACCAAGAAATTCCCTTGGGAGGAGATTTACCACACGATACAATCAGACGGCCAAGTGGA
CTGCGAAGTCCAATCAGTATCCAGGGACTATCTACAGATAGATCCGATAGAAAACAATTCCAGGTTTGGAAATT
TGGAAACTAAAGTGTCTGATCCAGCCAGAGGGATCTTAGAGAACCATAAGATAATAGACCTATGGGAAGGAGGC
AGAGACTTGATTGATCACGTTGTCATGCACGGCCCAACCAATGCTCAAAGATACACTGTGAAAGAAGGCTATTA
TTCCGTAATGGAAGTAACAAAGACTATATTCTCCAAATATCCGGTACAGTGTAGACCAATATTTCAAGACGAAG
CCTATGCAAGCCTTAATTCATTGACCGGCAGATTGGGAAGGAGCCTAGAAATAAGGAACATGAAAATAGTGCCT
TCAACAGACGAAGTAATTAAGAAGATGGCTAACTTGTTTTTCCACAAAAATTGGGAAGGAATGACTGATAAATA
TAGAATGGACCCAATAGTCTTCAACGATAAAGACTTCAGGGACTGGGTCATGGGCCACAAGAATGCTGCCAAAG
TAATCAAGGAGTTGGACAGTCTCTGTGCAGAAGGAATAAATACAAATCCTTTCAACAAATTCAGGAGCCATGTC
AAATTGGAAAGCATTAACAAACCTAACGCGATAGAGGACTTTCGCCAGTCAACTCCGAGGGCCATAGTTTGGTT
GCCGTATTGTATGCCAGCACTGTTTAGTTATATTTTTAAATTAGCAAGCAATAGGTTTAAGCTAATCCTCAGGG
ATAATGTTCATTATGCGTCGGGAATAGATGTAAATGACCTACAGAATTACGTAAATTTAGTGGAAGAAGATTGC
TACATCTTTGACAACGATATTAGTAAAATGGATTCCCAGGTTGATAGACACATGATAGAGATAGAGTGGGAAAT
GCTAAAATTAATGGGAGTTGACCCCGAAGTGTTAGAAAGCTACAAGGAACTCAAAAGAAATTGGACTATTTCAA
ACAAGTTCGTCAGGGTTAGTGATAGCTGGTTAAGGCACAGTGGTGAACCAACTACCGCCCTAGGAAACGGAATA
ATAAATCTTGCAATAACAAGTCTTTCATTGTCCAGAACTAAAAGATCTGACATGAAACTATGCTTGTTTGTGGG
AGATGACATGCTAATGGTTACCAAGGAAAAGGAAGACATTGACCTGGTTAAATTAAGGGGAAAGAAATTAGCCA
ACTCATTACTTAAACCCAGTATTAATAAAAGGTGCGGTCCGTTTTGTAGCTTTATAATAGGATACAGCGACATT
TGTACTGGAATGGCTGTTGTTCCGAATGTTAGTAGACTGGCATTTAAATGGGAGGTACCAAATGGACAGCACGA
AACTACTGACGAATCAGTCTTTACTAGACAATTGAGCTATGCCTGCCTACTTGGAAATAACAGCTTCAGTTCAA
TATTGCAGCCCTTAATCAGCAAACAAACCAAATGTGAGCTAGAAATACCAAACTATTACCGTGAATCAGATTTA
ATAAGATTAAATTGCGAGTACAGTAAACTACAAGAAATAGAGTTCATGGATTTACTTAATTTGCTGTACAATAG
AATTCTAAAACCTGAAACTATTCAGGTGAAATTCTTGATAACCTCAGAAAACATAAGGAAAGGAATCAAAAAGA
TGAGCCAGTTGAAAAGTAGTGAACATGAACTCGAAAGCAAGTGCCACGTTAGACTAACTGAAGAAACTGACTGA
TCAGTTATAAGGTGTTATAATACAAAACACATCACAGTAATTTTCTCTCGCCCCCCCCCC

Appendix 5 Polyprotein sequence of Geranium carolinianum endornavirus 1

SECOND MET chosen to begin ORF of polyprotein

DE   Translation of nucleotide sequence generated on ExPASy
DR   SWISS-2DPAGE; VIRT4910; VIRTUAL.
SQ   SEQUENCE   4815 AA.

Conserved domains (potential genes):
Helicase 1= green
Peptidase= blue
UDPGT = red

RdRp = yellow

MTTKCLARTP VTKNYVSLIS KNKKESRKGC LTAKLPIFKP DNIKEILKCV SNTYMISRHK
TKIIPVTPSI GMGEFFSIMM DGLAWDCANN FGLTTFLLGV IQGGKVLSNQ VIGNHNFKSS
YRSFNEKVTS LTADNYKGNH SAERLSHQLT RNINLRLDYT GIRELAVYIG FDYLLFPEED
YNASLGDNFP LDQTENFLKK SLEWCRPSRY AELLGEGYYT EFVTVGKGSA IHNLNNNVFI
CVACGCLNSK CDDEVHTCGK CLSINATVDI DVTELKNYKC LVLVPSCYEF DRSINGPLAA
LSLKNYLDGG LLNLPKLNSE LIAQDVLRYV KTRSRLDVKN QYVCANLSRD ELNYLREEFK
NLEIIVRNTW LDMQGMLMTE SHCHLTTLLN SEKKITEVQG FNRSYTNVGT SDTKFSAANH
LTNWQESNGE MAKNPQVMNY NSIPKMEKLL TMACARKIYV VVPNVFDQYN GFSNGFSFEI
SNERGLVKIL INGTTTVLEY TLEQMKLLLE YDYISCDDKL FEVKLIKKTS NCSLISISKL
RKNSLNVDEL GYKSVQPNKF KLFTLEIPDW QQNVMGLQLG PMIKRRIKFN MRFLKHLITR
CEAWPVSFHG LREYAIVSSF SRSESNDIVK NVFELNFEDI PDHVFCAYNI YLRQQLSTQF
THWLTTERSL GLDKLIQGFA GGIVGLITNL MNDHQNDKSF ISQLLDKCSD LSWLVSAQWD
KIEESIKTWE TDTVKLADFK GKIVESFVDK LETCQHGMEK DLISNGCNCC GRKTNELSGY
CSTCNLEGSC SHPCLHRCNS KIEHFCEGSV TRPDVAVGDN LICGHMVVTC KCCRKPSCQE
LCYKCFEWGQ YEDNLTRMAV VDQTIYKGEG AESAVRRVIK QAKGKQSYPK SIERKRVLYK
KNTKQANKQP QKGTTTSVIT DKPGDSSQKI QVQITKRQER QYFGLTSDSS EEDTKSDPFK
EREVNLNLED EDHDKCSEPE EPVDDTNVEK LKQHIMMNEL RTITNDEYCN IICQRSKPMD
ELLGESIKFK FLYQGEVFVD PTSINSIRPI IVPYTIGFCL RDTMAYYNPI INDITWADAC
SDTGLDESWC ILNDVWKYAE YFEMNVLIIH EMEGEVAGVY CYIHDKYEQV NMIRYSSEKP
GKEDYEELRG HYEPCEVGFN KEPSLPPVYA PDITWEDVNQ VYFNVTGGGD LGEFYDKNIE
DRLTIALALN ENKGLQYASS GFPKIELSNR KLGYMIFNNG NNLHEPRKGK LATFIKESTI
GSQLNIPSLL TRAELQEWYL DNLVKEEDLS NDKDCVRNAI TMIISQYLDL KMSCEETFNN
QNADNSNIAL LKKGIKIIDK GNYSVILAVD GLERLKTGDV IMIRRGPNRF CCQVERNIKR
IMIPKMPGHG TKLLIDIALF KISYTSLIIQ LASVSRPGIS LERAKELLGK ATCILGYPGT
GKTKELVKRY EMNPGLLVGV TRGSQESLIQ ELGARSKIVF SAERAMTNRA SSKTVYIDEA
SLITLPQLCC MLTPLVENLV ISGDLAQIPA KEFSKVCGYQ PTNILEFSKD TGATKVELKK
TWRFGKRICS ILNEAFGLDM QSATEKETNV NLEHSCGIDK NSLNRIVKER DIDTIMVFTT
QVERQVNSLL EPESRVRVAR VHSSQGSSFD RVLIVQDYRK GPAQGSEEVQ FKKEYVIVAM
TRCRKDVTIL CTYESCKCRE TSNESIARHL NRDLGLQYLY RGGKSNNVDV IAILNAVQDR
ISGLFESVEP NLVKDWLQEL WKIATGGIKK KMYMRMATMI NECNDREVLK TLLDSGMPMV
SDVITENGKL YAIMNYDNTN WNYVKKKTMT LFNSKQLIEC CNNKILIGGV EVISNQGSEE
IIVDANKFDS WESTQKHTKP ITIRGYQRRI NSENSGNINM NVRLINHSSQ LCWNAAKLTL
DIEYSGTKYS IKPTTGCSLC GGIQITKQNG ELMVFINNMY ENYSSRDIQF KSGTDPITKY
LLGKWDLDPR DDHLWELTAG LLPNVNHNAL HYTLWIERIM AAVKGIKNGK TFNTQEGHFF
RNELELNERI LSRYKNIANE AGIKITCESD RNYSYFKNLS FLFPAKFKGN SCYVYFHKHK
KCVLVNKTKK FGLSRELSPK LWNEELYRLP YNLSLMFGGS DLQIRGHKEV INMLDLDLEK
HDHNRSKLML MIDEDLTKLA KDKNFANPEI AIPSSLITDG KELGLASNFN IIPDLNLTGI
GASYLLDSVA AKLFSLNLME SGKTFITRYC NLSLRQDIER HIMIKPVDTK ITSYKDSETY
QDCFARILSK KNTVSEMAKA TDNSETKLKH LRIEQILQSM LEGRDLSSII VENADTCKTN
IGCLGVSCLE FGDMEINKIM RELNCKRLIG FIPDLRNKSI REIVSLNKNH LLFKGDSRDY
LINPKWVQLV NILYSSERWE SSMLLENLKI VGQTDLFLIV DITDTSEITV GKIPVSMHND
ETVVTVPQLN PINEIRRTGN LFNAVEFVID NETLSRLVRR AMTPGCTLPM LQTIARNRMQ
STVITKSGRK AGGRNVIRDV SLCALVAQYI ANHNDNQITR YLEKIDDYLM DNKDWKRSSG
ALIHMLKLEA NTIIGNALNI KVSLSELSKV LGEVTYNVMT DRSSKKGIPS IRVIHQPHRI
MYGEYYYHGL DSKNPNSLNL DPNVIGNILR NSFRSWTKNV ESKLAGYWAE PSVINSRLSD
HLTPDKEYSV KLLIYDMSGG IGRILQNRLK LTMKTDSVIH SSVCVGDNEI SYGNGVKISP
LGSQMVGKTS NPVTLGKIKL TAKNMEEIDK ITSEIFIPRK YSPIGLNCNL FSLWLLIQFG
YMTKIKISDS KLEHLENLAS LVPEFGSKVP ENVRKYIIAL NSKVMGDEEL TVKIMRCFEY
TVKSERPAGN SKLNNFIAHQ LIILTGRKLT DLRMQILEDE KSESESMDDD SRHDSDSDSD
DDQDNDGNNG TEQQLNEAED LEPEIEEGDR DARSSSEMNE DNDEENQESD FESAHSENMD
EEQASCNEAT LPSNEENMPS RIDNTPQIER INTGNDTESQ SKTIDNIAQE ITTEEKNTDQ
EENDTDRIMA PDSRTVLDFI KEEMEKLGID KVPTAMSTAS NAIGKLFSDA EINPRIDLYQ
NVNKELIDNA LCICQQAKKD FTVTKDDQNG LMSKKEKVGT YNERLRFKIL NKFQKTSICV
VALGSTGDTL PVLSACKMLK LGGAWICLLS HPDIYGLDNS NHDKFIKINK SQKRTTGNIK
SDSAIDIAKH AQSHNIEALR TFKEATQNHD FDLVLSTPLA PAVTGYSMFL GLRTAEAFCT

YCWNTGVEQG NSESDSFLLR WFGFTLKYAA VDGTLQILRQ SLATSMLREM NIESPDISTV
PRVVLSWEDV HSEKNIKDPK SVFIGYTSPG MKANLLSNDR SFRLLVGFGS MQVREEQINE
IRKVAKLMSG IEMIVTVHIQ DESLNKLLLI TVKNLFPKCK VLLGNVNLGE IVANNDAMVC
HGGIGTVQEC LMACCVPIII PCFADQPYVG SNLEKNQIGI MVSDSEAELT AKLKKIPMMQ
QKLKRKNYSM TDSVRNLTDA VLDLIGSQIP ISQLRETGQK VDRRFEVPTG VVIPYLTIPS
ESIALSLTPG EYQVNGVINE ETVYKIGESY YGGECFKEAF NNGILRLRGQ EYHIRAVHST
ALMTIETTTD IPKLNILGFY NHVNIQVLGH TNKTVIFNDS WPLLSIYVTK ISERRKHNSL
HAFIVANRTD IIRIEHLAKT GDKIASIIDN KSLCNRLSLP IGVDPKMAID SVIGCDNKLN
EWDWKCYGSF ENLRNRLKGE NIDALLQSDS SIAIPVLRTT QYVDGACFES NVNTRSLIGE
VVYCFTTMGV VPGIVVRQST SKIYVITSQS VTHLSGLIRC RLLRKKTGMI AGTEIDKLMK
TSYSYALNIE TINKIKEVIP GSNPKLMTAA IQEMISEYVV ISTNFDTRFH HSDRSNYERK
ILMDLIGTDK LKLLPNELNG NLAAIVKLTN VGYYNYAAGC LWYNIECREK DLCLSLSRIA
QECVNKRGSF FNSMLNFRIE AEDSNEFRDF IKIMNNNFDL VDNDILARSN ENIQDEKIIK
INLEWLNNNL IVTENKIAKL VFAEFLGKEK EVDAICQKAV ENSDINSFQL VSGVCLSRSG
LTILSLNPMV RITKIKIKGG AENKDDKTDS DMGRIFNIQP NETLMDQPLE MVRGPKLGKE
NFGRYDKWLK QNPYLEEPKN ADKGKSIMST KKFPWEEIYH TIQSDGQVDC EVQSVSRDYL
QIDPIENNSR FGNLETKVSD PARGILENHK IIDLWEGGRD LIDHVVMHGP TNAQRYTVKE
GYYSVMEVTK TIFSKYPVQC RPIFQDEAYA SLNSLTGRLG RSLEIRNMKI VPSTDEVIKK
MANLFFHKNW EGMTDKYRMD PIVFNDKDFR DWVMGHKNAA KVIKELDSLC AEGINTNPFN
KFRSHVKLES INKPNAIEDF RQSTPRAIVW LPYCMPALFS YIFKLASNRF KLILRDNVHY
ASGIDVNDLQ NYVNLVEEDC YIFDNDISKM DSQVDRHMIE IEWEMLKLMG VDPEVLESYK
ELKRNWTISN KFVRVSDSWL RHSGEPTTAL GNGIINLAIT SLSLSRTKRS DMKLCLFVGD
DMLMVTKEKE DIDLVKLRGK KLANSLLKPS INKRCGPFCS FIIGYSDICT GMAVVPNVSR
LAFKWEVPNG QHETTDESVF TRQLSYACLL GNNSFSSILQ PLISKQTKCE LEIPNYYRES
DLIRLNCEYS KLQEIEFMDL LNLLYNRILK PETIQVKFLI TSENIRKGIK KMSQLKSSEH
ELESKCHVRL TEETD

Appendix 6 GPS Coordinates of Locations where *G. carolinianum* samples collected

Location	**Coordinates**
AgCenter Botanic Gardens #1	30.409350, -91.109286
AgCenter Botanic Gardens #1	30.410442, -91.111341
AgCenter Botanic Gardens #1	30.408259, -91.112226
AgCenter Botanic Gardens #1	30.412028, -91.112932
AgCenter Botanic Gardens #1	30.415577, -91.118631
AgCenter Botanic Gardens #1	30.410190, -91.113195
AgCenter Botanic Gardens #1	30.408478, -91.104593
LSU Central Research Station #1	30.359984, -91.172370
LSU Central Research Station #1	30.359989, -91.172717
LSU Central Research Station #1	30.357304, -91.171994
LSU Central Research Station #1	30.357216, -91.172105
LSU Central Research Station #1	30.357387, -91.172057
LSU Central Research Station #1	30.357391, -91.172026
LSU Central Research Station #1	30.369064, -91.169932
LSU Central Research Station #1	30.369091, -91.169956
LSU Central Research Station #1	30.368996, -91.169668
LSU Central Research Station #1	30.369004, -91.169495
LSU Central Research Station #2	30.375322, -91.169994
LSU Central Research Station #2	30.375129, -91.169819
LSU Central Research Station #2	30.375013, -91.169328
LSU Central Research Station #2	30.366321, -91.170524

LSU Central Research Station #2	30.366321, -91.170492
LSU Central Research Station #2	30.360055, -91.172881
LSU Central Research Station #2	30.360064, -91.172982
LSU Central Research Station #2	30.357282, -91.172091
LSU Central Research Station #2	30.357255, -91.172144
LSU Central Research Station #2	30.357198, -91.172198
AgCenter Botanic Gardens #2	30.409142, -91.109508
AgCenter Botanic Gardens #2	30.410631, -91.110969
AgCenter Botanic Gardens #2	30.410582, -91.111311
AgCenter Botanic Gardens #2	30.410540, -91.111243
AgCenter Botanic Gardens #2	30.412033, -91.111832
AgCenter Botanic Gardens #2	30.414554, -91.117948
AgCenter Botanic Gardens #2	30.410148, -91.113234
AgCenter Botanic Gardens #2	30.408409, -91.111806
AgCenter Botanic Gardens #2	30.408149, -91.104191
AgCenter Botanic Gardens #2	30.407988, -91.104581
Tulane University Biodiversity Research Institute #1	29.891374, -89.953971
Tulane University Biodiversity Research Institute #1	29.891562, -89.955144
Tulane University Biodiversity Research Institute #1	29.891498, -89.955776
Tulane University Biodiversity Research Institute #1	29.890978, -89.955152
Tulane University Biodiversity Research Institute #1	29.889817, -89.955712
Tulane University Biodiversity Research Institute #1	29.888761, -89.956419
Tulane University Biodiversity Research Institute #1	29.889089, -89.956489
Tulane University Biodiversity Research Institute #1	29.888989, -89.956273
Tulane University Biodiversity Research Institute #1	29.886897, -89.954567
Tulane University Biodiversity Research Institute #1	29.887099, -89.959725
Tulane University Biodiversity Research Institute #2	29.891368, -89.953947
Tulane University Biodiversity Research Institute #2	29.891347, -89.954156
Tulane University Biodiversity Research Institute #2	29.891515, -89.954254
Tulane University Biodiversity Research Institute #2	29.891617, -89.954801
Tulane University Biodiversity Research Institute #2	29.891429, -89.954801
Tulane University Biodiversity Research Institute #2	29.891325, -89.954925
Tulane University Biodiversity Research Institute #2	29.890200, -89.955511

Tulane University Biodiversity Research Institute #2	29.890256, -89.955491
Tulane University Biodiversity Research Institute #2	29.889013, -89.956175
Tulane University Biodiversity Research Institute #2	29.887220, -89.957382

VITA

Rachel Herschlag is from Easton, Pennsylvania. She graduated from Easton Area High School in 2011, and in the fall of that same year began her undergraduate studies at Tulane University in New Orleans, LA. During her time at Tulane University, she conducted research on Podostemaceae, also known as the riverweed family. She studied the evolution of floral morphology within the family, and co-authored the description of a new species within Podostemaceae, *Ledermanniella achoundongii*. She graduated in the spring of 2015 with a B.S. in Ecology and Evolutionary Biology, and a minor in Public Health. In the fall of 2015, she joined Dr. Valverde's lab in the Louisiana State University Department of Plant Pathology and Crop Physiology as a graduate student. She anticipates graduating with her master's degree in December 2017.