EDNA-HOST: DETECTION OF GLOBAL

PLANT VIROMES USING

HIGH THROUGHPUT SEQUENCING

By

LIZBETH DANIELA PENA-ZUNIGA

Bachelor of Science in Biotechnology
Escuela Politecnica de las Fuerzas Armadas (ESPE)
Sangolqui, Ecuador
2014

Submitted to the Faculty of the

Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY

May 2020

EDNA-HOST: DETECTION OF GLOBAL

PLANT VIROMES USING

HIGH THROUGHPUT SEQUENCING

Dissertation Approved:

Francisco Ochoa-Corona, Ph.D.

Dissertation Adviser

Committee member Akhtar, Ali, Ph.D.

Committee member Hassan Melouk, Ph.D.

Committee member Andres Espindola, Ph.D.

Outside Committee Member Daren Hagen, Ph.D.

ACKNOWLEDGEMENTS

I would like to express sincere thanks to my major adviser Dr. Francisco Ochoa –Corona for his guidance from the beginning of my journey believing and trust that I am capable of developing a career as a scientist. I am thankful for his support and encouragement during hard times in research as well as in personal life. I truly appreciate the helpfulness of my advisory committee for their constructive input and guidance, thanks to: Dr. Akhtar Ali for his support in this research project and his kindness all the time, Dr. Hassan Melouk for his assistance, encouragement and his helpfulness in this study, Dr. Andres Espindola, developer of EDNA MiFi™, he was extremely helpful in every step of EDNA research, and for his willingness to give his time and advise; to Dr. Darren Hagen for his support and advise with bioinformatics and for his encouragement to develop a new set of research skills. I deeply appreciate Dr. Astri Wayadande thoughtful advice through my writing process, her invaluable support inspired me in unique way.

I also want to thank the head of the Department of Entomology and Plant Pathology (EPP), Dr. Phil Mulderfor his support. I like to extend my gratitude to Dr. Kitty Cardwell, the director of the Institute of Biosecurity and Microbial Forensics (IBMF) at Oklahoma State University for her assistance and encouragement. I am sincerely thankful to the founding sources that supported my research, Oklahoma Department of Agriculture, Food and Forestry (ODAFF), Combating Rose Rosette Disease USDA project. I also want to thank the students, staff and professors of EPP who were incredible cooperative and friendly during these years. Finally, I would like to express my sincere love and gratitude to my family who are my biggest fans. My parents believed, encouraged and supported me all the way to accomplish each of the dreams in my life. Also, to my sister whose presence in my life gave me support and friendship. I have been extremely blessed to have them as my family and moving along with my dreams helping to make them possible. I could not accomplished this without each of you.

Name: LIZBETH DANIELA PENA-ZUNIGA

Date of Degree: MAY, 2020

Title of Study: EDNA-HOST: DETECTION OF GLOBAL PLANT VIROMES USING HIGH THROUGHPUT SEQUENCING

Major Field: PLANT PATHOLOGY

Abstract:

The aim of this research was to develop the concept of host-specific virome detection methods. Three model systems were selected roses, cucurbits, and water. EDNA MiFi™ was used to generate pathogen electronic probes (E-probes) for detection of viruses *in-silico* and *in-vitro* using field samples for validation. The selected three host model allows a comprehensive detection system for specifically targeting the three pre-determined host viromes (117 plant viruses). This research also presents a validation of the analytical sensitivity determining the *in-vitro* limit of detection (10pg) quantified for a specific set of E-probes (ArMV-RNA2 60nt). The database of EDNA-Rose MiFi™ includes E-probes (523) specifically designed to detect 22 reported viruses infecting roses worlwide. The EDNA-Cucurbits MiFi™ includes E-probes (412) for 15 reported viruses. Finally, the MiFi™ database EDNA-Water is composed by a selection of highly specific E-probes (1730) for 80 reported viruses within the genera *Potexvirus, Tobamovirus* and *Tombusvirus*. Detection of the presence of multiple viruses in a host is challenging. Hence, the relevance of developing and applying a broad detection system to detect viruses out of a virome in a sample. The application of EDNA-Host extends to plant virus detection to all staple crops, virus-free certification programs, breeding, veterinary and life sciences.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER III

CHAPTER IV

CHAPTER V

CHAPTER VI

LIST OF FIGURES

CHAPTER III

CHAPTER IV

CHAPTER I

INTRODUCTION

Plant virus diagnostics methods rely upon key viral features comprising biological activity within the host and vector, morphological, physical, and chemical characteristics of the viral particle, its proteins, and viral nucleic acid attributes. Multiple viral infections play a key role in symptom expression during active infection (Moreno & López-Moya, 2020). Simultaneous mixed infections can be challenging to discern during virus detection and diagnosis because of the multiple viral gene expressions occurring at the same time with the plant response. The plant virome is defined as the virus population inhabiting a host (Mascia & Gallitelli, 2016). Plant viromes are specific to a host and its rage is determined by the host susceptibility. The overall biological interaction with the hosts, and allowance of main molecular pathway for viral genomic replication (Wolf at al., 2018).

Plant viruses are continuously evolving in a changing environment. Hence, it is highly relevant to study plant viruses in the context of their relationship with surrounding organisms.

The increasing interest in studying plant virus-host interactions has heightened research for establishing biological interactions. These interactions can be studied in perspective considering the virus inside the host, or interacting with the vector. The study of plant virus-host-vector interactions have been contributing current knowledge about synergy development and the progression of disease in a host (Mascia & Gallitelli, 2016).

A number of useful technologies have been conceived and validated for identification of viral causal agents when the disease etiology is both known or unknown. Most plant virus epidemiological studies are based on known viruses that affect crops (Jeger et al., 2006). It is important understanding many viral outbreaks and symptoms in the field may be caused by more than one virus. Hence, the diagnosis of early viral infection is cornerstone to design and to deploy integrated pest management (IPM). Sampling methods for plant virus detection derive from a small-scale epidemiological observation and are developed when characteristic symptoms are identified and associated with a disease while scouting (Agrios, 2005). Subsequently, symptomatic tissue is to be used as a reference specimen and has to be equally sampled with non-symptomatic sample tissue to achieve understanding of the complex interplay between the plant virion, the plant metabolism, and abiotic factors that cause disease. Decision making for management and control after plant virus detection needs further consideration regarding the sensitivity of the technique, the number of samples collected, the type of plant material sampled, the reliability of the selected method, equipment, and expertise available (Jeger et al., 2006).

Although detection and diagnostic technologies have been reported to identify single and also multiple plant viruses in a single test, these techniques are limited to identify a few viruses only in a single sample. Moreover, the partial amplification of targeted nucleic acid (PCR based tools) partially detects a segment of the sequence of the targeted pathogen amplifying a diagnostic segment in a range from ~50-3000 bp (Lévesque, 2001). Therefore, because most plant viruses have small RNAs, the nucleic-based detection method used may not cover the complete genome of the target. High

Throughput Sequencing (HTS) has emerged as a useful tool for analyses of microbial community studies to novel virus discovery. The impressive coverage and depth offered by available HTS platforms are useful and can be extended to detection and diagnosis methods (Maree et al., 2018).

Stobbe et al., (2013) pioneered the development of a bioinformatics pipeline capable of identifying DNA specific fingerprints in massive-parallel sequencing. Electronic Diagnosis Nucleic acid Analysis (EDNA) function as a reference Electronic probe (E-probe) guided database. E-probes are developed in specific non-conserved regions of the targeted genome and curated against taxonomically, biologically and phylogenetically related near neighbors. The original EDNA reported in 2013 has evolved to be MiFi™, a graphic user interface. The users of this new platform can build up databases for self-designed pathogen-specific E-probes, then uploading unassembled metagenomic reads from any HTS platform to identify plant pathogens in raw reads.

Stobbe et al., (2013), Blagden et al., (2016), and Espindola et.al., (2018) have all conducted pathogen-specific detection research, targeting fungal, oomycete, bacterial, and viral plant pathogens. Advances in sequencing platforms allowed depth and high coverage of the sequenced sample, enabling a broad detection of microbial communities related to a plant system. Nonetheless, these reports did not consider multiple pathogen infections in a plant host. In order to fully study the complexity of plant-host interactions, it is important to expand the capacity of the detection system considering the presence of a multiple-populated virome or pathosystem in a host. This research is the first studies that pursues complete virome detection. Rydzak et al.,(unpublished-2019); pursue to detect the cereal virome, as well as establishing a sensitivity limit of detection of EDNA-Cereals.

In this project, three model pathosystems were selected (rose, cucurbits, and water). The main focus is to validate a comprehensive detection system targeting a host pathosystem. The selection of the model hosts was based on the importance of garden rose and cucurbit crops to Oklahoma growers and its economy. Rose is an economically important specialty crop and its susceptibility to the *Rose*

rosette virus (RRV), which is part of a Nationwide project to control RRV, the Combating Rose Rosette Disease project sponsored by the USDA. The EDNA-Rose database is located at the MiFi™ server and includes E-probes designed for 22 reported viruses. Cucurbits are another important group of economically important specialty crops in Oklahoma. Ali, (2012) reported the presence of cucurbit infecting viruses in Oklahoma. The occurrence of plant viruses infecting cucurbits in Oklahoma has been monitored and recorded in the field since 2012. Therefore, the importance of developing a broad detection system able to find cucurbit viral-related presence in cucurbit field production areas, as well as, analyzing the presence of cucurbit viruses in genetically different hosts belonging to Cucurbitaceae family. The EDNA-Cucurbits database at the MiFi™ server includes E-probes for 15 reported viruses. Finally, three water-borne genera of plant viruses: *Potexvirus*, *Tombusvirus*, and *Tobamovirus*, previously reported to be waterborne in agricultural irrigation systems and nutritive-based solutions in recirculating irrigation systems at greenhouse production areas were addressed developing EDNA-Water MiFi™ and includes E-probes for 80 reported plant waterborne viruses.

Reverse Transcriptase quantitative PCR (RT-qPCR) is a powerful tool used for reliable, accurate, and sensitive detection. RT-qPCR coupled to High-Resolution Melting (HRM) allows verifying virus presence before sequencing (Olmos et al., 2018). A previous amplification of a partial plant viral sequence in a multiplex design allows recognition and quantification of the plant viral pathogen in a sample. However, although the multiplex RT-qPCR-HRM developed during this study is accurate, reliable, and sensitive it does not allow detection of more than five viruses in a single sample (Lévesque, 2001). All of these highlights the need for detection systems able to detect and identify multiple (more than five) pathogens in a single sample. This study uses RT-qPCR as a confirmation and validation tool for HTS testing.

This study provides valuable insights into HTS based diagnosis and introduces a new perspective for the application of EDNA MiFi™ in plant-pathogen diagnostics. EDNA-MiFi™ is a bioinformatics pipeline able to detect viruses as components of a complete reported virome of a host with an

acceptable limit of detection for viral pathogens. This study proposes a newly multiple detection strategy is applicable in plant diagnostic clinics, quarantine stations, border control, irrigation water testing, and plant mother stock plant viromes rapid screening.

This research contributes a multiple virus detection tool based in HTS. Moreover, research in plant virome diagnostics can assist identifying the epidemiological changes in the host during infection that leads to resistance. Other applications may also take the use of EDNA as a comprehensive breeding method able to target multiple viral infections and their synergy.

The overall goal of this project is to detect known virus species present in model host pathosystems (rose, cucurbits, and water) for:

1. Developing and validating E-probes for predetermine Host-Virus universes.
2. To analyze the presence of the global rose virome in garden rose varieties in Oklahoma field trials.
3. To detect and determine the distribution of cucurbit viruses in Eastern Oklahoma.
4. To assess the filtration of waterborne plant viruses for detection.

References:

Agrios, G. N. (2005). Plant diseases caused by nematodes. *Plant pathology*, *4*, 565-597.

Blagden, T., Schneider, W., Melcher, U., Daniels, J., & Fletcher, J. J. J. o. f. p. (2016). Adaptation and validation of E-probe diagnostic nucleic acid analysis for detection of Escherichia coli O157: H7 in metagenomic data from complex food matrices. *79*(4), 574-581.

Espindola, A. S., Schneider, W., Cardwell, K. F., Carrillo, Y., Hoyt, P. R., Marek, S. M., . . . Garzon, C. D. J. P. o. (2018). Inferring the presence of aflatoxin-producing Aspergillus flavus strains using RNA sequencing and electronic probes as a transcriptomic screening tool. *13*(10), e0198575.

Jeger, M. J., Seal, S. E., & Van den Bosch, F. (2006). Evolutionary epidemiology of plant virus disease. *Advances in Virus Research*, *67*, 163-203.

Lévesque, C. A. (2001). Molecular methods for detection of plant pathogens—What is the future?. *Canadian journal of plant pathology*, *23*(4), 333-336.

Maree, H. J., Fox, A., Al Rwahnih, M., Boonham, N., & Candresse, T. (2018). Application of HTS for routine plant virus diagnostics: state of the art and challenges. *Frontiers in plant science*, *9*, 1082.

Moreno, A. B., & López-Moya, J. J. (2020). When viruses play team sports: Mixed infections in plants. *Phytopathology*, *110*(1), 29-48.

Olmos, A., Boonham, N., Candresse, T., Gentit, P., Giovani, B., Kutnjak, D., ... & Nakhla, M. K. (2018). High-throughput sequencing technologies for plant pest diagnosis: challenges and opportunities. *EPPO Bulletin*, *48*(2), 219-224.

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., . . . Schneider,

W. J. J. o. m. m. (2013). E-probe Diagnostic Nucleic acid Analysis (EDNA): a theoretical approach for

handling of next-generation sequencing data for diagnostics. *94*(3), 356-366.

CHAPTER II

REVIEW OF LITERATURE

Agricultural biosecurity

Agricultural crops are increasingly vulnerable to plant pathogens due to globalized, trading, importing pathogen-infected plant goods of germplasm. Biosecurity policies and regulations prepare and protect a country or selected environments from transmission of infectious pathogens in agriculture. Protective protocols and policies has been developed to avoid or minimize the introduction of plant pathogens that may threaten economically important crops. Plant biosecurity faces continuously challenges based on safety, security, risk assessment, and global health sustainability. Plant pathogen introduction (intentional or unintentional) may interfere not only with crop industries but also can alter negatively the surrounding environmental ecology (Miley, 2020).

The global movement of plant germplasm increases the risk of introducing and establishing of invasive plant pathogens. Agricultural Biosecurity protocols, efficient biological characterization of pathogens, quarantine regulations, and risk assessment are mandatory efforts to minimize pathogen introduction and establishment (Ochoa-Corona, 2011).

Agricultural stakeholders need comprehensive programs that integrates plant protection against invasive species applying integrated methods to reduce pest impact in agricultural ecosystems before and after the trade takes place (Ireland et al., 2020). Therefore, there is a need to reduce the risk of introduction assisted by accurate pathogen identification, and quarantine and containment measures. Plant pathogens can cause disease in susceptible crops. They cause disease and weaken the host by deriving nutrients, disturbing host metabolism by own enzymes, toxins or by secreting growth-regulating substances. All of which block nutrient and water uptake, and uses the host cell machinery for reproduction and spreading (Agrios, 2005). It is important to develop a strategic screening plan that identifies all potential agricultural threats of crops. At present, diagnosis research is able to detect all the pathogens present in a host using a single metagenomics based-test, including fungi, bacterial, nematode, or virus groups, which are the most common agricultural threats (Agrios, 2005).

Plant virus origin and evolution

Viruses are infectious nucleic acids and protein encapsidated particles and the most diverse group of plant pathogens. About 30 families and over 145 reported genera of plant viruses are responsible to cause significant annual yield losses in Agriculture (Hull, 2013). Plant viruses widespread in diverse environments and virus evolution theories relay on about the origin of RNA molecules. Since the early description of the first plant virus, *Tobacco mosaic virus* (TMV) (Stanley, 1935), scientists have been studying the structure, organization, and genetic diversity of plant viruses (Hull, 2013). Plant virus classification uses the nature of the genome organization, w genome composition (dsDNA, ssDNA, ssRNA +/-, and dsRNA) among other features (Gergerich & Dolja, 2006). The plant global virome is highly comprised of RNA viruses, this genetic polarity derives in a replication advantage since the ss (+) RNA virus are similar to messenger RNAs (Wolf et al., 2018). Plant viruses are biotrophic obligate entities. Although the life cycle of each plant virus family differs from each other, the general life cycle is described as follows:

1). entering of the virion particle into the host cell; once in the cell, plant viruses use host cell machinery to form proteins needed for own replication of its genome, structure, and movement. 2). During replication, plant viruses form a complex of proteins to multiply in number their genome before moving to the next cell. Structural proteins encapsidate the newly formed genomes to produce a new viral particle, and 3). Movement proteins are formed by the virus to leave the cell and advance from a local infection through a systemic infection (Hull, 2013).

The International Committee on Taxonomy of Viruses (ICTV) is the official authority for virus classification. Virus taxonomy relays on phenotypic properties, genomic organization, and sequence-based framework for comprehensive virus characterization. Deep sequencing technologies have been helpful to uncover microbial communities related to specific host. Viral metagenomics data are changing our understanding of plant virus diversity by challenging the current plant-virus-vector interaction (Simmonds et al., 2017). Studying microbial diversity and interactions within a population has benefitted from the metagenomics approach allowing the analysis of the complete genomic content in a sample.

Disease symptoms and host range

Plant virus infection under the appropriate conditions may cause disease. Plant virus infection causes microscopic changes and macroscopic symptoms. Local symptoms are lesions delimited to the area where the virus penetrates the cell producing limited necrosis. Host infected cells present a loss of chlorophyll or pigments as a consequence of reduced area for photosynthesis (Agrios, 2005). Systemic symptoms derive from active disease development, the disease involves a sequential symptomatology development as stunting, mosaic patterns, yellowing leaves, leaf rolling, ringspot, necrosis, developmental abnormalities, wilting, reduced nodulation, and genetic effect in the host. Also, it is important to mention that some viral infection does not result in symptom production. The absence of symptoms in the plant is can be due to a mild strain of the virus, a tolerant host, age of the leaves, a mosaic pattern of dark green islands, or infection with a

cryptic virus (Hull, 2009). Other factors can produce virus-like symptomatology in plants such as small cellular parasites, bacterial infection, toxins produced by arthropods, genetic abnormalities by inherited plastid effects on the plant, transposons, nutritional deficiencies, high temperatures, hormonal damage, insecticides, and air pollutants (Hull, 2013).

Plant virus infection is often limited to a higher proportion of related plant species in which the virus can replicate, and this determines the host range. Host range evolution is shaped by environmental factors and spatio-temporal discontinuities among plant-virus interactions. Competitive evolutionary effects are often influenced by natural enemies, anthropogenic disturbance, and biotic-abiotic interactions (McLeish et al., 2018). Determinants in the host range are based on small changes in the genomic sequence of the virus. Biological determinants include recognition of a suitable host cell or organelle, lack of specificity in the uncoating process, presence of an appropriate suppressor tRNA in the host during virus replication, disruption in the cell to cell movement, the specific response of host resistance genes during stimulation of host cell defenses in virus infection, or an incompatible interaction between host genes affecting host range (Hull, 2013).

Economic impact

Plant virus infections in Agriculture cause a reduction in yield and quality of the product. Furthermore, the economic impact of plant viruses has a significant effect on the growth and nutrition uptake of a plant. Crop loss productivity has a burden on food security. Damage associated with a crop may cause direct and indirect effects. Reduction in growth may include yield reduction even with symptomless infections. Reduction in vigor by increasing sensitivity to frost and drought, as well as, increase in predisposition to plant pathogens. Reduction in the quality of the marketable product like visual defects in the fruit size, shape, taste, color, and texture. Agricultural stakeholders cover the cost associated to maintain crop health (Hull, 2013).

Persistence and dissemination

Generally, plant viruses, are actively transmitted by an insect vector or passively transmitted by seed, soil, mechanically inoculated, and water (Ertunc, 2020). Long-term persistence is rare for plant viruses (*Tobacco mosaic virus*, TMV). Viruses need a living organism to function as a vector. Virus transmission may require a specific kind of vector. Plant viruses have developed biological complex interaction with their vectors (arthropod, nematode, and fungi). The specific association between the plant virus and the vector reflects the complexity of the pathosystem (Gray & Banerjee, 1999). Interactions between plant viruses and an insect vector's body are characterized in four types. Non-persistent transmission, semi-persistent, circulative, and propagative transmission have been described (Watson and Roberts, 1939; Kennedy et al., 1962; Sylvester, 1958; Nault, et al., 1997).

Virus transmission by insect vectors in crops is the most economically important type of transmission. In non-persistent transmission, viruses are acquired within seconds after feeding on an infected plant and are carried at the tip of their mouthparts for 24 hr or less. In semi-persistent transmission, virions from the mouthparts and foregut are inoculated in the plant as the insect salivates during feeding. In circulative transmission, plant viruses accumulate inside the body of the vector before passing to plant tissue via ejection from the salivary glands via salivation at feeding. Lastly, in propagative transmission plant viruses replicate inside the insect vector, plant virus titer increases before inoculation in plant tissue (Gray & Banerjee, 1999).

Vegetative propagation (budding, grafting, cutting, and infected tubers) is another pathway for plant virus transmission. Once the virus infects the propagative material it will reduce the quality of the progeny. Mechanical transmission through sap takes place between closely planted crops. Virus-infected sap can be acquired by an insect and lead to the infection of more plants. In the case of seed transmission, plant viruses are present in the ovule, pollen of an infected plant.

Seedling infected with plant viruses may induce disease in the plant producing symptoms (Agrios, 2005).

Ecology and epidemiology

There are complex interactive factors that drive the ecology and epidemiology of plant viruses. Plant virus epidemiology studies the determinant causes, dynamics, and distribution of disease in a susceptible crop. Epidemiology of viruses focuses on vector-host-pathogen interactions in cultivation systems. Virus vector presence in the field will determine virus dynamic movement and spread. Moreover, environmental conditions and availability of a suitable host will regulate virus entry, replication, movement, and disease establishment in the host. Environmental conditions may be favorable for continual replication and movement of the virus in the plant as well as the survival and persistence of the vector in the field. Plant virus ecology considers factors that may influence virus behavior like host response, tissue tropism, host range, and virus pathogenesis. In addition, studied ecological factors allow to conclude the infected field interaction within the surrounding environment (Hull, 2013).

Management

Plant virus management strategies are challenging and focus mainly on controlling the vector. Application of broad-spectrum chemical control, cultural control, the input of organic amendments in the soil may reduce virus transmission by arthropod vectors (Perring et al., 1999). Cultural control relies on removal or avoidance of the source of infection, a plant of virus-free rootstocks, virus-free seeds, heat therapy, low temperature – cryotherapy, meristem tip culture, tissue culture, and modifying plant and harvest procedures (Hull, 2013). Biological control agents have been studied using predatory natural enemies present in the rhizosphere against plant viruses. Few resistant cultivars are available. Screening for resistant cultivars showed that the most resistant varieties, showing high yield and have become a target for breeding programs. Quantitative trait loci (QTLs) breeding programs have not found consistency in resistant markers so far. However,

breeding research in the future may include successful markers from QTL analysis (Maule, 2007). Recently, CRISPR-Cas13a technology has been applied to provide genetic resistance to susceptible cultivars (Khan et al., 2018). Integrated pest management minimizes future outbreaks (Wisler & Norris, 2005).

Host-pathogen-vector interaction

The host range of plant viruses varies from narrow to broad among plant species. The virus interaction with the host is a product of co-evolution. Amazingly, viruses are diverse in nature, their genomic features, and interaction with the cells of the host. From a genomic compatibility perspective plant virus once in the host induce disease as a complex process relaying mostly on two or more viral genes expressed by their genome. Mutations and gene products play an important role during the virus infection cycle (Hull, 2013).

The viral genome form networks during its interaction with the host genome. Similarly, a successful plant viral transmission and interaction depends on the interplay between the genomes of the virus, host, and vector. These type of interactions have two main pathways, the first that initiates the infection cycle, and the second that interferes the metabolic pathways of the host hijacking plant immunity and leading to disease. Structural protein-protein and nucleic acid interactions can also occur between and within viral species (Hull, 2013).

A plant virome is defined as the assortment of all virus nucleic acid components that infect plants (Coetzee et al, 2010). The plant virome distribution and selection among a host had a recent breakthrough from research in ecological metagenomics. It was reported genes of the coat protein interact with house-keeping genes or interactive modules (gene activation) in a way to produce disease in their host (Dolja & Koonin, 2011).

Virus-virus interactions and mixed plant virus infections

Multiple infections in a single host had revealed different types of interaction among viral particles. Four types of virus-virus interactions among conserved sequence have been described: susceptibility, competition, synergy, and interference (Melcher & Ali, 2018). These interactions involve the structure of the viral particle, either proteins and/or the nucleic acids. The virus-virus interactions are also influenced by spatial-tempo factors of the cellular pathway and its interplay with the vector. During the competition, non-viral components lead to a rivalry for host and vector components for viral survival and evolution. The synergistic interactions can produce severe symptoms (higher virus titer and movement in the host), but synergistic effects may also be unequal due to variation in virus strains. Interference and cross-protection are negative interactions among the viruses involved and may restrict symptom production (Melcher & Ali, 2018).

Plant virus accumulation levels, host responses, and interactions with the vector are dynamic factors that impact the virus dissemination during a mixed infection. Pathogenic viruses can spread quickly in the environment posing complex challenges to stop virus dissemination. Hence, evolutionary rates during mixed infection may change the host and virus fitness because competition for nutritional resources and plant cell machinery (Malpica et al., 2006). If comparing mixed and single infections, mixed infections are the result of successfully synergistic interactions of viral particles benefiting the viruses by an increase of virus titer, pathogenicity level, and the capacity to interact with the vector (Moreno & López-Moya, 2020).

Plant virome

The virome is a relatively new concept based on intensive research to describe all viruses associated with a specific host environment. The concept of the plant virome arose from new advances in research of metatranscriptomic and metagenomic analyses of viral metagenomes (Pooggin, 2018). The term virome also can be associated with all the viruses related to a host. In

general, the plant virome is heavily composed of positive-strand RNA viruses in relation to DNA viruses (Dolja & Koonin, 2011), with emphasis on the viromes associated with specific marketable plant species. Al Rwahnih et al., (2011), studied the virus community infecting grapevine plants, and reported 60 viruses and mycoviruses present in the pathosystem. Whole-community sequencing for the metagenomics study of virus-associated to grapevine helped to reveal the characterization of those viruses. There are few complete reports of the rose (Milleza et al., 2013) and the cucurbit (Lecoq et al., 1998) viromes. In this research, viromes associated with garden rose, cucurbits, and water are detected to determine the biological interaction with their hosts.

Plant virus Viromics

Agronomic, horticultural, and vegetable crops are often affected by plant virus infection. The deep study of molecular patterns and interplay between the virus, the vector and the host during the infection cycle is called Plant Viromics. The revelation of such complex interactions is a direct result of advances in genomics, transcriptomics, and proteomics combined by molecular techniques, and supporting evidence of coevolution of virus, host, and vector (Hull, 2013). High throughput sequencing (HTS) and advances in bioinformatics are a revolutionary tools for virus discovery in all types of ecosystems. Moreover, understanding the virome of a host provides new insights for studying virus in diversity, population dynamics, virus interactions, and presence of a virus community in a specific host (Tatineni et al., 2020).

Genomic interactions have lead and contributed knowledge to virus evolution and interaction within a life kingdom (Dolja & Koonin, 2011). Synergistic and antagonistic interactions can be a result of a virus complex infection by two or more viruses in a single host, those viruses are able to modify the pathogenic potential to cause disease within that host. Multiple viral infections can interfere with host cell machinery to produce unexpected symptoms, infection, accumulation and transmission pathways (Syller & Grupa, 2016).

Diagnosis

Increased crop vulnerability in agriculture is impact due to delays occurring since the pathogen introduction take place and an accurate pathogen detection, diagnosis, and response is completed (Savary et al., 2019). The detection and identification of plant viruses rely upon traditional and modern methods. The screening process is often based on symptom expression and identification during field scouting. Suspected viruses are identified by either serological mainly Enzyme-Linked Immunosorbent Assay- ELISA or molecular-based methods mainly Polymerase Chain Reaction - PCR. Despite the accuracy and reliability of both methods, each lacks a broad multiplexing capacity, and are not suited for detecting variants, races, and/or new strains (Adams et al., 2009).

Metagenomics has emerged (Massart et al., 2017) as a powerful tool able to support plant virus diagnosis increasing multiplexing capacity. The metagenomics outputs sequence contain all pathogens present in the sample (Simmonds et al., 2017). Plant virus identification starting from a partial section of a targeted genome generated the need for diagnosis methods or pipelines able that comprises the complete genomes of plant pathogens. Therefore, an optimized diagnostic method able to process the genetic diversity of strains and races infecting a host suitable for plant diagnosis clinics is needed (Massart et al., 2017).

New plant virus diseases management systems can developed upon the basis of accurate, reliable, and sensitive diagnostic, if Metagenomics is an alternative for rapid identification of disease causal agents that may support epidemiological studies (Adams et al., 2009).

Polymerase chain reaction

Polymerase chain reaction (PCR) at present is a fundamental pathogen detection tool. PCR is a DNA amplification technology that uses short DNA flanking oligonucleotide sequences (primers) complementary to the target DNA. Along with the anchoring of these short single-stranded DNA sequences, the DNA amplification process occurs through cycles (~30-40 cycles) of controlled temperatures required for denaturation, annealing, and extension (Mullis et al., 1986). *Thermus*

aquaticus (Saiki, 1988) is the selected bacterial species from where the polymerase enzyme was isolated for further use and DNA amplification. After amplification, DNA may be loaded onto agarose gel to perform electrophoresis, also fluorescent labeled products can be used to visualized amplified DNA. Electrophoresis uses electronic charges to allow DNA migration which varies based on the molecular weight of the target DNA. Finally, UV light is used to visualize the target amplicon in the gel (Benett et al., 2004). Advances in PCR technology may include an additional step of reverse transcription. Modifications of PCR include real-time detection, quantification, droplet PCR, and isothermal amplification (Lauerman, 2004).

High-Resolution Melting (HRM)

High-Resolution Melting (HRM), is a method commonly applied to genotype differentiation. HRM in plant pathology is used to discriminate highly polymorphic regions in plant genomes (Mackay et al., 2008). The advantages offered by discrimination of single-nucleotide of samples, makes HRM extensively used as a detection and discrimination method (Bester et al., 2012; Komorowska et al., 2014; Farrar & Wittwer, 2017). HRM is a post-PCR method that dissociates previously PCR amplified DNA sequences. The sample containing the amplicon is heated from 65 to 99 ^{o}C. Once the double-stranded DNA dissociates, the attached fluorescence marker or chromophore is released and the lack of fluorescence is registered. The temperature of melting (TM) is calculated by software resulting in a unique Tm per specimen (Massart et al., 2017)

Sequencing

Sequence-based genome mapping was first developed by Sanger et al. (1977). Sanger sequencing is based on the chain-termination method using the Maxam-Gilbert chemical method and chain-termination biochemistry for small fragments (~500-600 nucleotides). Later, the Human Genome Project (Venter et al., 2001), led to the development of the shotgun sequencing. The shotgun sequencing platforms are based on random fragmentation of DNA strands that undergo further sequencing by synthesis. Afterwards, sequencing in parallel was developed from the basis of

shotgun sequencing. Sequencing by synthesis in parallel allowed the DNA sequence determination of DNA (Kreuze et al., 2009). Current commercial platforms sequence short and long reads. Platforms available for High Throughput Sequencing are Illumina® (short) (Illumina, San Diego, CA), Pacific Biosciences® (long) (PacBio, Menlo Park, CA), and Oxford Nanopore (long) (Oxford Nanopore, Oxford, UK).

High throughput sequencing (HTS) in plant diagnostics

The development of a robust framework for plant virus sequence-based diagnosis is essential for a comprehensive characterization of viromes associated to a host. The advance of HTS studies had revealed the existance of large viromes within a single ecosystem (Simmonds et al., 2017). Furthermore, HTS had uncovered host phytobiomes associated to a particular studied ecosystem. Also confirming the presence of pathosystems related to the host being studied. Moreover, HTS contributed to discover the interaction among plant pathogens alternative host reservoir (Hily et al., 2018). Alo, asymptomatic infected plants timely diagnosed, and novel cryptic viruses can be discovered with deep sequencing analysis (Massart et al., 2017).

Oxford Nanopore ™ developed a portable sequencer, the MinION™, which has the potential to sequence a single molecule. Synthetic biological proteins forming nanopores are fixed in a synthetic electric-resistance membrane, holding around ~1600 pores per flow cell. Once the DNA has been previously enriched and specially required adapters attached, the DNA to be sequenced it will pass through nanopores assisted by an electrical current flow. While the DNA molecules passes through the nanopore, their electrical conductivity is measured with the assistance of the MinKNOW™ software, which is recording the DNA sequence in long reads. Base pair identification is done by detecting and discriminating differences in electronic resistance putative to each nucleotide. Reads are commonly stored in the computer hard drive as .fast5 files. Subsequently, fast5 files are basecalled by GUPPY™ (software)to generate fastq files (Bayley, 2015; Laver et al., 2015; Stoddart, Heron, Mikhailova, Maglia, & Bayley, 2009).

Traditional bioinformatics approach for virus detection

The HTS approach was used for inferring ecological relationships, microbial discovering, and biological characterization of a pathogen (Massart et al., 2017). The HTS pipeline provides invaluable data that helps to understand the current phytosanitary status of a crop (Hily et al., 2018). *De novo* assembly and characterization strategy is for identification of plant viruses from HTS metagenomic data are consist of: *de novo* assembly, which generates contigs and scaffolds; mapping contigs against the reference genome, searching homology in genomic databases, and finally identifying taxonomically the pathogen. Often, the whole process is time-consuming, and algorithms must be run by a trained bioinformatician scripting to reduce time. A standard process can take as much as three weeks before a comprehensive diagnosis report is completed. Trade and quarantine stations at the border are delayed by held up decision making. As a consequence the detection time may takes long and approximately more than a week. Hence, a rapid, reliable, accurate, sensitive, and specific HTS diagnosis tool is needed (Massart et al., 2017; Pirovano, Miozzi, Boetzer, & Pantaleo, 2015).

Electronic Diagnosis Nucleic Acid Analysis (EDNA)- MiFi ™

Electronic Diagnosis Nucleic acid Analysis (EDNA) is a reference-guided approach, able to report components of a given plant virome in a host examining complete genome by electronic probes (E-probes) coverage. E-probes are unique specific sequences of the targeted pathogen organized in the database, MiFi™, ready to match raw HTS reads from a suspect sample (Blagden et al., 2016; Espindola et al., 2018; Stobbe et al., 2013).

EDNA has been validated for detection of plant pathogenic fungal, bacterial, oomycetes, and viruses (Stobbe et al., 2013) and has been applied to gene expression and transcriptomic quantification (Espindola et al., 2018). The EDNA pipeline is designed to quickly examine and detect plant pathogens based on target-specific E-probes. The uniqueness of the E-probes depends on the comparison of the target genome with biologically and phylogenetically related genomes of

near neighbor organisms (Stobbe et al., 2013). The use of EDNA E-probes proved to be a relatively rapid detection method for specific and discrete targeted viruses if compared with other HTS bioinformatics analyses.

EDNA MiDetect™ calculate the statistical probability of finding a complete match between a unique E-probe in an unassembled metagenomics dataset containing a plant pathogen. The semi-quantitative version of EDNA MiFi™ allows recording the number of hits from E-probes thar target a sequence (Espindola et al., 2018) and the qualitative detection of EDNA MiFi™ provides a presence/absence of the pathogen in a tested sample(Blagden et al., 2016; Stobbe et al., 2013).

EDNA outputs can be valudated in preference comparing results with a validated standard method Polymerase chain reaction (PCR), and it is Retro Transcriptase (RT) quantification version (q). RT-qPCR is a reliable confirmation method for metagenomics-based research or diagnosis.

The model hosts selected for this study

A host is defined as an organism from which nutrients or means for replication can be derived by a parasite (Agrios, 2005). To understand biological interactions between plant viruses and their susceptible hosts, model systems have been selected and defined to facilitate the study of these interactions (Agrios, 2005). Model systems allow to infer biological conclusions based on the plant and how the plant viruses establish a relationship to self-replicate producing infection. The obtained biological conclusions which are the result of data collection and interpretation allows the understanding of the model system and further application (Bolker, 1995).

The selection of the model hosts in this study considered the horticultural importance and value of the production of garden roses and their inbred varieties. For cucurbits, there was also interest for understanding of the occurrence of related viruses present in the Oklahoma fields. Although, water is not a living organism *perse,* it is an essential element for life that allows survival,

persistence of virus, and transportation of microorganisms. Therefore, water was selected as a host because its use in irrigation and pathway for contamination.

Rose

The genus *Rosa* is the result of many years of selective breeding. Active speciation of hybrids throughout the years of breeding has established the horticultural varieties used today. Due to interspecific hybridization, rose breeders developed cultivars with 2x, to 8x chromosomes seeking disease resistance, color, bloom, and fragrance (Horst & Cloyd, 2007). Genetic variability among cultivars generate resistance or susceptibility to disease. Marker-assisted selection in 2-3-year disease trials had helped rose breeders to increase and search for disease resistance against fungi and Rose Rosette Disease respectively. A 2-3-year trial cultivars that showed the best performance to disease resistance in a recurrent phenotypic selection are selected as resistant cultivars (Debener & Byrne, 2014). For example, *R. bracteata*, *R. clinophylla*, and *R. multiflora* have resistance genes against black spot. To present, *R. seitigera* is the rose variety with the highest level of resistance to Rose Rosette Disease (Byrne, 2013). Garden rose is a specialty horticultural crop with economic importance in the US (Horst & Cloyd, 2007). In Oklahoma, garden rose varieties were planted in a disease trial from the USDA founded grant to combat Rose Rosette Disease. The rose varieties used in the combating Rose Rosette disease trial were the same in all fields trials replicated in multiple states in the U.S. seeking for cultivars with Rose Rosette disease resistance (Byrne, 2013). Rose Rosette Disease and its mite vector were found in Oklahoma, which created a need to study the viruses infecting roses and its interaction in susceptible and resistant cultivars (Olson et al., 2017).

Cucurbits

Cucurbits are a source of fresh food, seed oil, and vitamins. The family *Cucurbitaceae* includes a group of species that are genetically and phenotypically diverse, with about 1000 species and 96

genera distributed into 5 clades (Grumet et al., 2017). The genetic difference among species is interesting if seeing from a plant pathological and biological perspective. The over more than 100 years of breeding have improved the cucurbit germplasm worldwide. Breeders also had focused on producing plant pathogen-resistant cultivars (Grumet et al., 2017). In Oklahoma, several studies of cucurbit viral diseases have been made by Ali et al., (2012) in different cucurbit crops. Viral diseases can cause important yield losses to the producers. During summer the Oklahoma cucurbit fields present a complex and changing pathosystem of viral diseases. Field surveys were conducted in mayor cucurbit growing counties in Oklahoma during 2008 and 2010 to study the plant virus distribution of major cucurbit crops (Ali et al., 2012). The result of these studies contributed to new research questions about how the viruses are infecting cucurbits can infect cucurbits despite the genetic diversity of the host. Also, further methods of virus detection with multiple targets able to screen large germplasm.

Water

Water is the most important element needed for plant production and survival. The sanitation of irrigation water sources is critical for agricultural success. Plant health relay importantly on water quality (Sevik, 2011). Waterborne pathogens can survive and disseminate in water causing agricultural yield losses if encountering their susceptible host. Irrigation and recirculated water contaminated with plant pathogens provide a continuous source of inoculum. Therefore, timely plant pathogen detection and monitoring of irrigation water are important for implementing water biosecurity policies and to avoid the establishment of detrimental microorganisms in the crop. Aquatic ecology is also a promising field of study to develop ecologically-based water mitigation plans for plant-pathogen systems (Hong & Moorman, 2005). There are no conclusive research studies that addresses the validation of the structure of a water filtration system for the metagenomic plant virus detection method. The outcome of the aquatic ecological studies can contribute with

important information applicable for Integrated Pest Management strategies aiming to decontaminate water sources.

Viromes associated with the model hosts

Each crop has a unique microbiome population associated to it. The plant virome is mainly composed by RNA single strand positive sense viruses with few exceptions (Dolja & Koonin, 2011). In this study, three viromes infecting specific hosts as model systems were selected: rose, cucurbits, and water.

Rose virome

Roses host twenty-six reported viruses worldwide (Converse & Bartlett, 1979; Di Bello, Ho, & Tzanetakis, 2013; Fulton, 1970; Golino, Sim, Cunningham, & Rowhani, 2011; He et al., 2015; Moury, Cardin, Onesto, Candresse, & Poupet, 2001; Rivera & Engel, 2010; Thurn, Lamb, & Eshenaur, 2019; I. Tzanetakis, Reed, & Martin, 2005; I. E. Tzanetakis, Halgren, Mosier, & Martin, 2007). Rose-associated viruses can be founded in mix infections causing multiple symptoms by a synergistic association. This study focuses on the presence of plant viruses belonging to the rose virome among susceptible and resistant cultivars in Tulsa and Perkins counties, in Oklahoma.

Cucurbit virome

Phylogenetic divergence among clades in the family Cucurbitaceae has become biological interesting in plant pathology. The Begoniaceae and Cucurbitaceae clades comprises the most common cucurbit cultivars. Despite this genetic differences plant viruses can infect and produce similar symptomatology on infected cucurbits. Moreover, cucurbits can be infected from over 35 plant viruses reported worldwide. Cucurbit viromes depend upon the host diversity to establish infection, viral genome replication and symptom development (Lecoq et al., 1998).

Water virome

Plant virus-infested water increases yield loss and virus dissemination in crops. Aqueous environments can hold seven different genera of viruses (Mehle, N., & Ravnikar, 2012). This study selected three genera of water-borne plant viruses: *Potexvirus, Tobamovirus,* and *Tombusvirus* (80 viruses). These genera have been reported to include water-borne virus species, and a number of them occur in the U.S. Viral infections in irrigated fields or hydroponic crops are not detected until symptoms are evident (Mehle & Ravnikar, 2012).

This project aims to detect plant reported viromes from three model systems (Hosts) using HTS combined with bioinformatic pipelines (EDNA MiFi$^{TM)}$. The goal is to timely detect known virus species belonging and known to be present in a specific host (rose, cucurbits, and water), as well as understanding the biological implication of the virus interaction with the host during infection.

References

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., . . . Boonham, N. J. M. p. p. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *10*(4), 537-545.

Agrios, G. N. (2005). Plant diseases caused by fungi. *Plant Pathology*, *4*.

Ali, A., Abdalla, O., Bruton, B., Fish, W., Sikora, E., Zhang, S., & Taylor, M. J. P. h. p. (2012). The occurrence of viruses infecting watermelon, other cucurbits, and weeds in the parts of the southern United States. *13*(1), 9.

Al Rwahnih, M., Daubert, S., Urbez-Torres, J. R., Cordero, F., & Rowhani, A. (2011). Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. *Archives of virology*, *156*(3), 397-403.

Bayley, H. J. C. c. (2015). Nanopore sequencing: from imagination to reality. *61*(1), 25-31.

Benett, W. J., Richards, J. B., Stratton, P. L., Hadley, D. R., Milanovich, F. P., Belgrader, P., & Meyer, P. L. (2004). *U.S. Patent No. 6,699,713*. Washington, DC: U.S. Patent and Trademark Office.

Bester, R., Jooste, A. E., Maree, H. J., & Burger, J. T. (2012). Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated virus 3 variant groups I, II, III and VI. *Virology journal*, *9*(1), 219.

Blagden, T., Schneider, W., Melcher, U., Daniels, J., & Fletcher, J. J. J. o. f. p. (2016). Adaptation and validation of E-probe diagnostic nucleic acid analysis for detection of Escherichia coli O157: H7 in metagenomic data from complex food matrices. *79*(4), 574-581.

Bolker, J. A. (1995). Model systems in developmental biology. *BioEssays*, *17*(5), 451-455.

Byrne, D. H. (2013, August). Advances in rose breeding and genetics in North America. In *VI International Symposium on Rose Research and Cultivation 1064* (pp. 89-98).

Coetzee, B., Freeborough, M. J., Maree, H. J., Celton, J. M., Rees, D. J. G., & Burger, J. T. (2010). Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology*, *400*(2), 157-163.

Converse, R. H., & Bartlett, A. J. P. D. R. (1979). The occurrence of viruses in some wild Rubus and Rosa species in Oregon. *63*(6), 441-444.

Debener, T., & Byrne, D. H. (2014). Disease resistance breeding in rose: current status and potential of biotechnological tools. *Plant Science*, *228*, 107-117.

Di Bello, P., Ho, T., & Tzanetakis, I. (2013). *Identification of the rose rosette disease agent.* Paper presented at the VI International Symposium on Rose Research and Cultivation 1064.

Dolja, V. V., & Koonin, E. V. J. C. o. i. v. (2011). Common origins and host-dependent diversity of plant and animal viromes. *1*(5), 322-331.

Ertunc, F. (2020). Emerging Plant Viruses. In *Emerging and Reemerging Viral Pathogens* (pp. 1041-1062). Academic Press.

Espindola, A. S., Schneider, W., Cardwell, K. F., Carrillo, Y., Hoyt, P. R., Marek, S. M., . . . Garzon, C. D. J. P. o. (2018). Inferring the presence of aflatoxin-producing Aspergillus flavus strains using RNA sequencing and electronic probes as a transcriptomic screening tool. *13*(10), e0198575.

Farrar, J. S., & Wittwer, C. T. (2017). High-resolution melting curve analysis for molecular diagnostics. In *Molecular diagnostics* (pp. 79-102). Academic Press.

Fulton, R. J. P. V. N. (1970). Prunus necrotic ringspot virus. CMI/AAB Descript.

Gergerich, R. C., & Dolja, V. V. (2006). Introduction to plant viruses, the invisible foe. *The plant health instructor*, *1*, 24.

Golino, D., Sim, S., Cunningham, M., & Rowhani, A. (2011). *Evidence of root graft transmission of two rose mosaic viruses, Prunus necrotic ringspot virus and Apple mosaic virus in rose rootstocks.* Paper presented at the Phytopathology.

Gray, S. M., & Banerjee, N. (1999). Mechanisms of arthropod transmission of plant and animal viruses. *Microbiol. Mol. Biol. Rev.*, *63*(1), 128-148.

He, Y., Yang, Z., Hong, N., Wang, G., Ning, G., & Xu, W. J. M. p. p. (2015). Deep sequencing reveals a novel closterovirus associated with wild rose leaf rosette disease. *16*(5), 449-458.

Hily, J.-M., Candresse, T., Garcia, S., Vigne, E., Tannière, M., Komar, V., . . . Hommay, G. J. F. i. m. (2018). High-throughput sequencing and the viromic study of grapevine leaves: From the detection of grapevine-infecting viruses to the description of a new environmental Tymovirales member. *9*, 1782.

Holmes, F. O. (1929). Local lesions in tobacco mosaic. *Botanical Gazette*, *87*(1), 39-55.

Hong, C. X., & Moorman, G. W. (2005). Plant pathogens in irrigation water: challenges and opportunities. *Critical Reviews in Plant Sciences*, *24*(3), 189-208.

Horst, R. K., & Cloyd, R. A. (2007). *Compendium of rose diseases and pests*. American Phytopathological Society (APS Press).

Hull, R. (2009). *Comparative plant virology*. Academic Press.

Hull, R. (2013). *Plant virology*: Academic press.

Ireland, K. B., van Klinken, R., Cook, D. C., Logan, D., Jamieson, L., Tyson, J. L., ... & Rodoni, B. (2020). Plant Pest Impact Metric System (PPIMS): Framework and guidelines for a common set of metrics to classify and prioritize plant pests. *Crop Protection*, *128*, 105003.

Kennedy, J.S., Day, M.F., Eastop, V.F. 1962. A Conspectus of Aphids as Vectors of Plant Viruses. London: Commonwealth Institute of Entomology 114 pp.

Khan, M. Z., Amin, I., Hameed, A., & Mansoor, S. (2018). CRISPR–Cas13a: prospects for plant virus resistance. *Trends in Biotechnology*, *36*(12), 1207-1210.

Komorowska, B., Fiore, N., Zamorano, A., & Li, R. (2014). Simultaneous detection of Cherry necrotic rusty mottle virus and Cherry green ring mottle virus using real-time PCR and high resolution melting analysis. *Molecular and cellular probes*, *28*(4), 186-191.

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

Lauerman, L. H. (2004). Advances in PCR technology. *Animal health research reviews*, *5*(2), 247-248.

Laver, T., Harrison, J., O'neill, P., Moore, K., Farbos, A., Paszkiewicz, K., . . . quantification. (2015). Assessing the performance of the oxford nanopore technologies minion. *3*, 1-8.

Lecoq, H., Wisler, G., & Pitrat, M. (1998). Cucurbit viruses: the classics and the emerging. In *Cucurbitaceae* (Vol. 98, pp. 126-142).

Mackay, J. F., Wright, C. D., & Bonfiglioli, R. G. (2008). A new approach to varietal identification in plants by microsatellite high resolution melting analysis: application to the verification of grapevine and olive cultivars. *Plant Methods*, *4*(1), 8.

McLeish, M. J., Fraile, A., & García-Arenal, F. (2018). Ecological complexity in plant virus host range evolution. In *Advances in virus research* (Vol. 101, pp. 293-339). Academic Press.

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., . . . Škorić, D. J. F. i. m. (2017). A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. *8*, 45.

Maule, A. J., Caranta, C., & Boulton, M. I. (2007). Sources of natural resistance to plant viruses: status and prospects. *Molecular plant pathology*, *8*(2), 223-231.

Malpica, J. M., Sacristán, S., Fraile, A., & García-Arenal, F. (2006). Association and host selectivity in multi-host pathogens. *PloS one*, *1*(1).

Mehle, N., & Ravnikar, M. (2012). Plant viruses in the aqueous environment–survival, water-mediated transmission and detection. *Water Research*, *46*(16), 4902-4917.

Melcher, U., & Ali, A. (2018). 15 Virus–Virus Interactions in Plants. *Plant Viruses: Diversity, Interaction, and Management*.

Miley, K. (2020). Global Health Biosecurity in a Vulnerable World–An Evaluation of Emerging Threats and Current Disaster Preparedness Strategies for the Future. In *Global Health Security* (pp. 79-102). Springer, Cham.

Milleza, E. J. M., Ward, L. I., Delmiglio, C., Tang, J. Z., Veerakone, S., & Perez-Egusquiza, Z. (2013). A survey of viruses infecting Rosa spp. in New Zealand. *Australasian Plant Pathology*, *42*(3), 313-320.

Moury, B., Cardin, L., Onesto, J.-P., Candresse, T., & Poupet, A. J. P. (2001). Survey of Prunus necrotic ringspot virus in Rose and Its Variability in Rose and Prunus spp. *91*(1), 84-91.

Moreno, A. B., & López-Moya, J. J. (2020). When viruses play team sports: Mixed infections in plants. *Phytopathology*, *110*(1), 29-48.

Mullis, K. et al. (1986). Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 51: 263–73.

Nault, L. R., Styer, W. E., Knoke, J. K., and Pitre, H. N. 1973. Semipersistent transmission of leafhopper-borne *maize chlorotic dwarf virus*. Journal of Economic Entomology, 66(6), 1271-1273.

Ochoa-Corona, F. M. J. A. P. P. (2011). Biosecurity, microbial forensics, and plant pathology: education challenges, overlapping disciplines, and research needs. *40*(4), 335.

Olson, J., Rebek, E. J., & Schnelle, M. A. (2017). Rose rosette disease.

Perring, T. M., Gruenhagen, N. M., & Farrar, C. A. (1999). Management of plant viral diseases through chemical control of insect vectors. *Annual review of entomology*, *44*(1), 457-481.

Pirovano, W., Miozzi, L., Boetzer, M., & Pantaleo, V. J. F. i. m. (2015). Bioinformatics approaches for viral metagenomics in plants using short RNAs: a model case of study and application to a Cicer arietinum population. *5*, 790.

Pooggin, M. M. (2018). Small RNA-omics for plant virus identification, virome reconstruction, and antiviral defense characterization. *Frontiers in microbiology*, *9*, 2779.

Rivera, P. A., & Engel, E. A. J. V. g. (2010). Presence of rose spring dwarf-associated virus in Chile: partial genome sequence and detection in roses and their colonizing aphids. *41*(2), 295-297.

Sanger F., Nicklen S., Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Science USA 74: 5463–67.

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., Nelson, A. J. N. e., & evolution. (2019). The global burden of pathogens and pests on major food crops. *3*(3), 430.

Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., . . . Harrach, B. J. N. R. M. (2017). Consensus statement: Virus taxonomy in the age of metagenomics. *15*(3), 161.

Stanley, W. M. (1935). Isolation of a crystalline protein possessing the properties of tobacco-mosaic virus. *Science*, *81*(2113), 644-645.

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., . . . Schneider, W. J. J. o. m. m. (2013). E-probe Diagnostic Nucleic acid Analysis (EDNA): a theoretical approach for handling of next-generation sequencing data for diagnostics. *94*(3), 356-366.

Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G., & Bayley, H. J. P. o. t. N. A. o. S. (2009). Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *106*(19), 7702-7707.

Syller, J., & Grupa, A. J. M. p. p. (2016). Antagonistic within-host interactions between plant viruses: molecular basis and impact on viral and host fitness. *17*(5), 769-782.

Sylvester, E.S. 1958. Aphid transmission of plant viruses. International Congress of Entomology Montreal, Canada. Presentation.

Tatineni, S., Stewart, L. R., Sanfaçon, H., Wang, X., Navas-Castillo, J., & Hajimorad, M. R. (2020). Fundamental Aspects of Plant Viruses− An Overview on Focus Issue Articles. *Phytopathology*, *110*(1), 6-9.

Thurn, M., Lamb, E., & Eshenaur, B. (2019). Disease and Insect Resistant Ornamental Plants: Rosa (Rose).

Tzanetakis, I., Reed, J., & Martin, R. J. A. o. v. (2005). Nucleotide sequence, genome organization and phylogenetic analysis of Strawberry pallidosis associated virus, a new member of the genus Crinivirus. *150*(2), 273-286.

Tzanetakis, I. E., Halgren, A., Mosier, N., & Martin, R. R. J. V. r. (2007). Identification and characterization of Raspberry mottle virus, a novel member of the Closteroviridae. *127*(1), 26-33.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., and Gocayne, J.D. 2001. The sequence of the human genome. Science 291(5507):1304-1351.

Watson, M.A., and Roberts, F.M. 1939. A comparative study of the transmission of Hyocyamus virus 3, potato virus Y and cucumber virus 1 by the vectors *Myzus persicae* (Sulz.), *M. circumflexus* (Buckton), and *Macrosiphum gu* (Koch). Proceedings of the Royal Society of London 127:543–76.

Wisler, G. C., & Norris, R. F. (2005). Interactions between weeds and cultivated plants as related to the management of plant pathogens. *Weed Science*, *53*(6), 914-917.

CHAPTER III

DEVELOPMENT AND VALIDATION OF E-PROBES FOR PREDETERMINE HOST-VIRUS UNIVERSES

Abstract

Global trade of plants and goods had increased the risk for plant pathogen introduction and establishment, into new regions or countries threatening agriculture and their horticulture production. There is a need for developing a robust, rapid, specific, and sensitive detection method using High Throughput Sequencing (HTS) metagenomes to detect plant viruses avoiding assembling the raw reads. The E-probe diagnostic nucleic acid assay – Microbe Finder (EDNA-MiFi™) is a bioinformatic pipeline for plant pathogen detection and identification that uses raw HTS datasets. EDNA-MiFi™ has passed from a proof-of-concept to beta-testing. This diagnostic tool has been applied for detection of plant pathogens. The objective of this project is to test and to validate EDNA-MiFi™ a bioinformatic pipeline sensitive and specific for pathogen detection of predetermined plant viromes in three host models (rose, cucurbits, and irrigation water). Pathogen-specific electronic probes (E-probes) were designed selecting unique sequences, then curated for target specificity. The new designed specific E-probes were then tested *in silico* generating simulated spiked metagenomes with different abundance levels of the target pathogen and host sequences. The *in-vitro* limit of detection,10 pg (10^{-3}), of EDNA-MiFi™ correlates with the results obtained *in silico*. The bioinformatics pipeline EDNA-MiFi™ can be rapidly queried for plant viruses without need for sequence assembly.

1. Introduction

The understanding of biological implications of the interaction between the virus and the host changed by metagenomics and metatranscriptomic research. Recent metatranscriptomic and metagenomic studies have elucidated the role of mixed infections in plant hosts (Moreno & López-Moya, 2020). Global virome studies contributes to overcome challenges on classification, isolation, and detection of viruses related to a single plant species (Paez-Espino et al., 2016).

Importation of germplasm is a pathway for introduction exotic pathogens entering new regions threatening agricultural crops. Plant viruses can be disseminated by intermediate vectors, such as insects, fungi, water, and soil into new regions growing economically important crops. During infection, plant virus replication rate increases causing symptom severity and a cascade of metabolic dysfunctions in the plant defense system (Mine et al., 2014). Some detrimental plant virus diseases are reported to date: Plum pox, Citrus Tristeza, Cassava mosaic disease, Maize streak virus disease, Rose rosette disease, among others (Agrios, 2005, Matthews, 2019). Therefore, the ability to quickly detect potential plant virus infection in a susceptible crop can prompt the development and implementation of management strategies that minimizes the likelihood of a new virus spread and establishment between farms, countries, and continents.

High throughput sequencing technologies had evolved to generate millions of sequencing reads from a single sample. Deep sequencing metagenome analysis requires data assembly, annotation, and taxonomical identification before achieving a conclusion (Massart et al., 2017). Hence, a bioinformatic detection pipeline that reduces the computational time and with a graphical user interface was needed. EDNA has the capability of detecting all plant pathogens present in a single sample. EDNA MiFi^{TM} can consistently detect multiple plant pathogens in a single metagenome sample (Stobbe, et.al., 2013). To date, EDNA has been tested in numerous systems including viruses (Stobbe et al., 2014), fungi and oomycetes (Espindola et al., 2015, and Espindola et al.,

2018), plant infecting bacteria like spiroplasmas in leafhoppers (Andearson, 2016 unpublished dissertation), and human pathogens (Blagden et al., 2016). EDNA MiFi™ platform evolved to be a graphical user interface that rapidly analyzes metagenomic data to accurately find plant pathogen fingerprint sequences (Espindola 2015). At present, the EDNA MiFi™ validation focuses on sensitivity, specificity, accuracy and statistical adjustments. This study focuses on studying the sensitivity of the method *in-silico* and *in-vitro* research.

The purpose of this study is to apply and adapt the pipeline EDNA MiFi™ My probe for development pathogen-specific E-probes for accurate detection of plant viruses in three host model systems (rose, cucurbits, and water) from metagenomic datasets. Roses and cucurbits model systems were selected because their economic importance in Oklahoma. Water was chosen because of its environmental importance, and as a potential source of virus transmission in agriculture.

2. Materials and Methods

2.1. E-probe design

The genome sequences from reported plant viruses were obtained from the NCBI Genbank. E-probes of 20-30 nt length were generated using the E-probe pipeline for each model system (rose, cucurbits, and water) EDNA MiFi™ (Stobbe, et.al., 2013). The E-probe pipeline within EDNA uses a modified version of Tools for Oligonucleotide Fingerprint identification (TOFI) (Satya et al., 2008). The modified TOFI pipeline eliminates redundant genome regions in the process of developing E-probes. Also, the modified TOFI pipeline compares the target sequence of the plant virus (full genome) to taxonomically and biologically related sequences to identify unique, highly-specific E-probes. EDNA does not use thermodynamic optimization process of TOFI since is not necessary. After TOFI comparison, designed E-probes undergo a specific BLAST for E-probe length restriction and elimination of homo-oligomers. Afterwards, EDNA MiFi™ performs a Basic Local Alignment Search Tool (BLAST) curation search to achieve the uniqueness of the E-probes.

Pairwise alignments were performed to determine the E-probe specificity using a stringency of 100% identity and query coverage by Basic Local Alignment Search Tool Nucleotide (BLASTn) with the nucleotide database at the National Center of Biotechnology Information (NCBI). The obtained target-specific E-probe datasets are then uploaded to MiFi^{TM} to query raw sequencing reads without read assembly or annotation.

For increasing specificity, EDNA MiFi^{TM} E-probe libraries eliminates E-probes prone to false-positive results. To ensure e-probe specificity, draft E-probes were queried against the NCBI nucleotide database at an E-value of 1×10^{-9}. Therefore, each host E-probe database went through a curation step before being uploaded to MiFi^{TM}. The reverse sequences of the final E-probe set, were used as decoy E-probe set, and used as internal control for statistically random matching or false positives avoidance within the metagenome datasets. Finally, generated E-probes were uploaded to the EDNA MiFi^{TM} database.

in-silico

2.2 Generation of simulated metagenomes and query

To estimate the sensitivity and accuracy of curated E-probe sets, artificially simulated metagenomes were constructed with known percentage of the targeted virus including background host sequences (rose, cucumber, and aquatic irrigation water metagenome). Simulated metagenomes were developed using MetaSim®, a software for deep sequencing run simulation (Richter et al., 2008). The MetaSim® creates metagenomes from target virus genomes combined with the host genome included as background sequences.

Rose genome sequences were kindly provided by Dr. David Byrne from Texas A&M. *Rosa multiflora* assembled and annotated contigs were used as a host background. The selected rose metagenome was spiked with sequences of the *Rose rosette virus* (RRV) genome from NCBI RefSeq NC_015298.1, NC_015299.1, NC_015300.1, NC_015301.1, NC_034979.1, NC_034980.,

NC_034981.1 and uploaded to the metagenome simulator program, MetaSim®. Ten million synthetic reads were generated using an empirical error to mimic the Illumina sequencing platform. The metagenome simulations were tested as a single infection with RRV, and multiple infections using *Prunus necrotic ringspot virus* (PNRSV), and RRV.

The cucumber metagenome was retrieved from the Sequence Read Archive (SRA) of NCBI Bio project: PRJNA33619 BioSample: SAMN02953750; Assembly: GCA_000004075.2; Accession Number: PRJNA33619 submitted in 2011 (Li et.al., 2011). The *Cucumis sativus* cultivar 9930 metagenome from a whole-genome shotgun sequence (Chromosome1-7) served as the host background. The selected cucumber metagenome was spiked with sequences from the *Cucumber mosaic virus* (CMV) genome and uploaded to the metagenome simulator program MetaSim®. Ten million synthetic reads were generated using an empirical error to mimic the Illumina sequencer. The metagenome simulations were tested as a single CMV infection, and multiple infections of *Papaya ringspot virus* (PRSV), *Squash mosaic virus* (SqMV), *Watermelon mosaic virus* (WMV), and *Zucchini yellow mosaic virus* (ZYMV).

The experiment validation for water-borne plant viruses was done *in-silico* using an aquatic metagenome of an environmental sample from the SRA of NCBI Bio project: PRJNA232936 BioSample: SAMN02566839; Assemble name: DSRE_500; Accession Number: JFZN01000001.1 submitted 2014 (Decker & Parker, 2014). The aquatic metagenome from a whole-genome shotgun sequence served as host background. The selected aquatic metagenome was spiked with sequences of *Pepino mosaic virus* (PepMV) genome and uploaded to the metagenome simulator program MetaSim®. Ten million synthetic reads were generated using an empirical error to mimic the Illumina sequencing platform. The metagenome simulations were tested as a single infection of PepMV and multiple infections using one virus of each genera PepMV (*Potexvirus*), *Cucumber necrosis virus* CuNV (*Tombusvirus*), and *Pepper mild mottle virus* PMMoV (*Tobamovirus*).

2.3 Validation of EDNA Water -MiFi™ with published SRA data

Agricultural biosample sequences generated from Illumina HiSeq 2500 run from Sequence Read Archive (SRA) SRR5995660-5995697 of viral metagenomics sequencing of freshwater samples SRX3151361 were retrieved from NCBI from Singapore (Gu et.al., 2018). Raw reads from freshwater metagenomes were directly uploaded to the EDNA-Water MiFi™ server and tested against the database for plant water-borne virus *Potexvirus*, *Tombusvirus*, and *Tobamovirus* genera.

2.4. Blind test of EDNA Rose-MiFi™

A blind test is a process useful for validating a technology, by comparing of results among laboratories. Rose metagenomes kindly provided by Dr. Golino, University of California Davis. A total of twenty-three Illumina HiSeq 2500 sequencing runs from rose samples were sequenced at the Foundation Plant Service. Raw reads from rose metagenomes were directly uploaded to the EDNA-Rose MiFi™ platform and tested against the database for rose related viruses.

2.5. Statistical Analysis

The modified EDNA-MiFi pipeline (Espindola et al., 2018) allows counting high quality hit alignments for each E-probe based on a set of metrics selected by the user (hit counts). These hits are transformed into a score to allow the use of parametric statistics. Each e-probe will have a unique score that will be later compared with negative control e-probes. Pairwise T-test were conducted on parsed alignments of metagenomes to compare specific E-probe hit scores with negative/decoy e-probes (internal negative control). P-values lower than 0.05 are considered positive and those higher than 0.05 suspect or negative.

2.6. *In vitro* Sensitivity

To validate the sensitivity of EDNA-MiFi™ is important to have control of the pathogen concentration in a given sample. Therefore, an artificial sequence of the RNA2 of the *Arabis mosaic virus* (ArMV) was artificaially synthesized by GenScript®. The insert was 3820 nucleotides long was inserted into the plasmid pUC57 (Figure3.1). The lyophilized plasmid

carrying the RNA2 of the virus was inserted into TOPO-TA cloning *E.coli* competent cells (Thermo Fisher®, Waltham, MA). The plasmid in *E.coli* competent cells were transformed by heat-shock (42 ºC), and incubated at 37 ºC for 8h in Ampicillin enriched LB media.

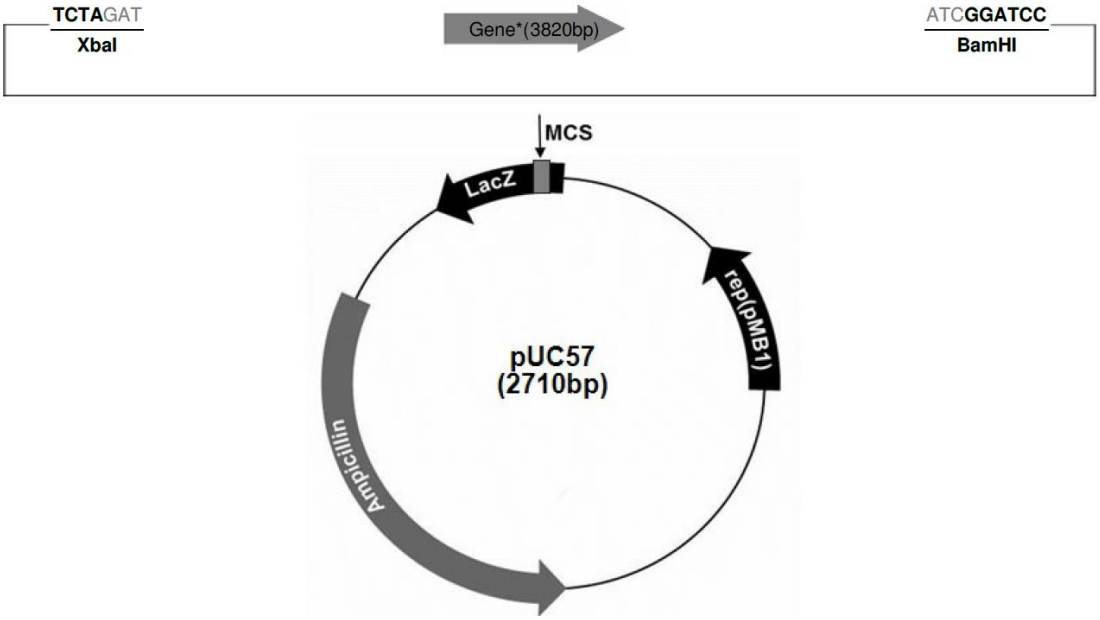


Figure 3. 1 Cloned pUC57 plasmid carrying the *Arabis mosaic virus* (ArMV) RNA2 genome insertion. The RNA 2 insert (3820bp) of the virus was done in the Multiple Cloning Site (MCS) at the LacZ expression promotor of the plasmid.

Isolated colonies were selected and enzyme digestion was performed to verify the presence of the RNA2 genome. Twenty nanograms of plasmid were digested by SpeI enzyme as per the manufacturer's protocol (New England Biolabs, Ipswich, MA). A 1% agarose gel was run with the digested products at 90V for 40 minutes. In the gel, two bands were observed, the inserted RNA2 genome from ArMV is separated showing a band at ~4000 bp along with another band of ~ 3000 bp belonging to the plasmid.

Plasmid DNA purification

Plasmid DNA extraction from transformed *E.coli* cells was performed using QIAprep® Spin miniprep Kit (QIAGEN, Hilden Germany) following the manufacturer's protocol. The resulting DNA was quantified by Quant-iT™ PicoGreen™ (ThermoFisher Scientific, Waltham, MA).

Plasmid DNA– genomic host DNA dilution ratio

Once concentration ratios were determined, the purified DNA of the plasmid containing the RNA2 genome from ArMV was combined with the genomic DNA of the host as described in Table 3.1.

Table 3. 1 Concentration ratios were determined to simulate differences in gene expression between the virus in the host before Illumina Nextseq 500® (San Diego, CA) Sequencing. Grapevine DNA was used as a negative control. Each sample was sequenced by triplicates.

ArMV RNA2 construct (ng/uL)	Grapevine DNA (ng/uL)	Total [] ng
55	0	412.5
27	27	405
0.1	54	405.75
0.01	54	405.075
0.001	54	405.0075
0.0001	54	405.075
0.00001	54	405.0075
0.000001	54	405.00075
0.0000001	54	405.000075

Library preparation was then performed for each of the twenty-seven mixtures using the KAPA HyperPlus library prep kit (Roche, Basel, Switzerland) and KAPA Single Index Adaptors were included (Roche, Basel, Switzerland) as per manufacture's protocol. Each library was then sequenced with an Illumina Nextseq 500 (San Diego, CA) and metagenomic data was generated for each dilution and replicated three times per sample.

2.7. EDNA-MiFi™ *in-vitro* sensitivity detection

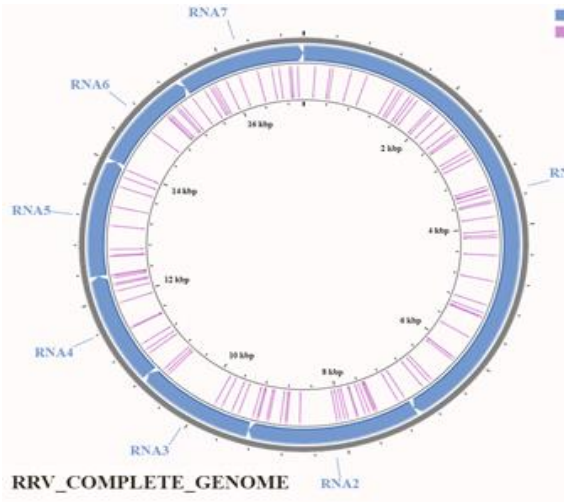
The obtained metagenomes were uploaded to the EDNA MiFi™ platform and queried against the ArMV 60nt E-probes (sourced by Dr. Andres Espindola, Oklahoma State University, Stillwater, OK). The sum of scoring results for each detection determined the total number of hits for the ArMV E-probe set. This data was correlated to the concentration of the pathogen in the host to determine the limit of detection.

3. Results

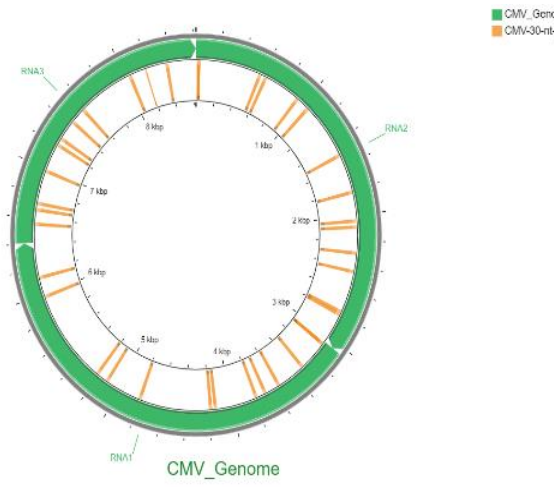
3.1 E-probes for virome detection

Highly specific E-probes sets were designed for 22 viruses infecting rose, 15 viruses infecting cucurbits and 80 plant-infecting waterborne viruses. The E-probe sets comprise all the RNA genomes of the targeted viruses (Figure 3.2), having an even distribution along each virus genome.

a.



b.



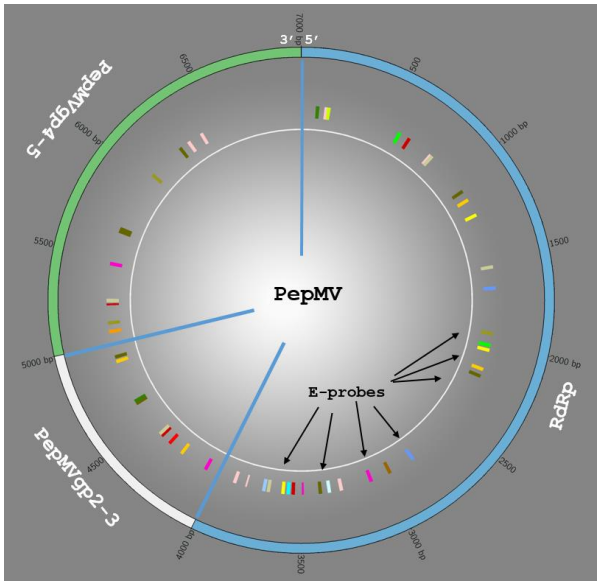
c. 

Figure 3. 2 E-probe set location along the three model virus genomes

a). artificial representation of the sever RNA that compose *Rose rosette virus* into a circular blue genome representation (blue) side by side with E-probe coverage (purple). *Rose rosette virus* is one example; however, all of the designed EDNA-Rose E-probe sets contain E-probes that will detect all 22 viruses infecting roses. b). Circular representation of the *Cucumber mosaic virus* (CMV) linear tripartite segmented genome. The green circle represents the segmented genome made of its three RNAs. The orange lines represent the E-probe coverage along the genome. c). Circular representation of the PepMV linear genome (6.45Kb), the three are shown by blue, white and green segments, the main encoding proteins. Generated E-probes are covering the complete genome and are represented by small lines of different colors distributed among the three encoding proteins of the virus particle.

The designed E-probes match all virus coding regions involved during active viral infection and symptom producing stage. The selected E-probe length was 20 – 30nt (Table 3.2). The E-probe sets detected rose viruses from raw metagenomic sequencing reads.

Table 3. 2 List of targeted viruses and genome size, and the number of Rose curated E-probes selected for the Rose Virome.

Virus	Genome size (Kb)	Number of E-probes	E-probe length	E-probe ID
Apple chlorotic leafspot virus	7.5	21	20	ACLSV-20
Apple mosaic virus	8.52	29	30	ApMV-30
Apple stem grooving virus	6.5	11	20	ASGV-20
Arabis mosaic virus	11.15	30	20	ArMV-20
	6.4	12	20	BCRV20
Impatiens necrotic spot virus	16.74	20	20	INSV-20
Iris yellow spot virus	16.83	21	30	IYSV-30
Prune dwarf virus	8.09	15	20	PDV-20
Prunus necrotic ringspot virus	7.88	23	20	PNRSV-20
Raspberry ringspot virus	11.85	29	20	RpRSV-20
Rose cryptic virus 1	4.69	5	30	RCV-1-30
Rose leaf curl virus	2.74	7	20	RoLCuV20
Rose leaf rosette-associated virus	17.66	59	20	RLRaV-20
Rose rosette virus	17.08	22	20	RRV-20
Rose spring dwarf-associated virus	5.81	19	20	RSDaV-20
Rose yellow leaf virus	3.9	10	20	RoYLV-20
Rose yellow mosaic virus	9.51	24	20	RoYMV-20
Rose yellow vein virus	9.31	17	20	RoYVV-20
Strawberry latent ringspot virus	11.34	34	20	SLRSV-20
Tobacco ringspot virus	11.44	22	20	TRSV-20
Tobacco streak virus	8.63	19	20	TSV-20
Tomato ringspot virus	15.48	48	20	ToRSV-20
Tomato spotted wilt virus	16.64	22	20	TSWV-20
Tomato yellow ring virus	0.82	4	20	TYRV-20

Similarly, E-probes were successfully generated for fifteen cucurbit infecting viruses in this study. Table 3.3 describes the number of E-probes per virus. The selected E-probe length was 30nt, which provide optimal specificity, sensitivity, and computing processing time. WSMoV and SLCuV are the two viruses with the lowest E-probes, four and seven respectively. The viruses with the larger number of e-probes were CYSDV with fifty E-probes and TRSV with forty-two E-probes.

Table 3. 3 List of Cucurbit reported viromes, their genome size, number of specific E-probes per virus, E-probe length, and E-probe ID for EDNA-Cucurbit MiFi™.

Virus	Genome size (Kb)	Number of E-probes	E-probe length	E-probe ID
Alfalfa mosaic virus	8.27	28	30	AMV-30
Bean pod mottle virus	9.66	32	30	BPMV-30
Cucurbit aphid-borne yellows virus	5.67	11	30	CABYV-30
Cucumber green mottle mosaic virus	6.42	20	30	CGMMV-30
Cucumber mosaic virus	8.63	34	30	CMV-30
Cucurbit yellow stunting disorder virus	17.1	50	30	CYSDV-30
Melon necrotic spot virus	4.27	10	30	MNSV-30
Papaya ringspot virus	10.33	38	30	PRSV-30
Squash leaf curl virus	5.24	7	30	SLCuV-30
Soybean mosaic virus	9.58	21	30	SMV-30
Squash mosaic virus	9.22	35	30	SqMV-30
Squash vein yellowing virus	9.84	20	30	SqVYV-30
Tobacco ringspot virus	11.44	42	30	TRSV-30
Watermelon mosaic virus	10.04	30	30	WMV-30
Watermelon silver mottle virus	17.33	4	30	WSMoV-30
Zucchini yellow mosaic virus	9.59	30	30	ZYMV-30

E-probes sets were also designed for 80 water-borne plant viruses (Table 3.4). The E-probe sets comprise all the ss (+) RNA genomes, having an even distribution of E-probes along their viral genomes. The selected E-probe length was 20 nt. These E-probe sets detected water-borne viruses from raw metagenomic sequencing reads.

Table 3. 4 List of water-borne plant viruses, their genome size, number of specific E-probes per virus, E-probe length, and E-probe ID for EDNA-Water MiFi™.

Virus	**Acronym**	**Genus**	**Genome size (Kb)**	**Number of E-probes**	**E-probe length**	**E-probe ID**
Allium virus X	AlVX	*Potexvirus*	7.12	30	20	AlVX-20-probes
Alstroemeria virus X	AlsVX	*Potexvirus*	7.01	27	20	AlsVX-20-probes
Alternanthera mosaic virus	AltMV	*Potexvirus*	6.61	39	20	AltMV-20-probes
Asparagus virus 3	AV-3	*Potexvirus*	6.93	34	20	AV-3-20-probes
Bamboo mosaic virus	BaMV	*Potexvirus*	6.37	34	20	BaMV-20-probes
Cactus virus X	CVX	*Potexvirus*	6.61	38	20	CVX-20
Cassava common mosaic virus	CsCMV	*Potexvirus*	6.38	35	20	CsCMV-20-probes
Cassava virus X	CsVX	*Potexvirus*	5.88	30	20	CsVX-20-probes
Clover yellow mosaic virus	ClYMV	*Potexvirus*	7.02	26	20	ClYMV-20-probes
Cymbidium mosaic virus	CymMV	*Potexvirus*	6.23	33	20	CymMV-20-probes
Foxtail mosaic virus	FoMV	*Potexvirus*	6.15	24	20	FoMV-20-probes
Hosta virus X	HVX	*Potexvirus*	6.53	40	20	HVX-20-probes
Hydrangea ringspot virus	HdRSV	*Potexvirus*	6.19	24	20	HdRSV-20-probes
Lagenaria mild mosaic virus	LaMMoV	*Potexvirus*	3.86	21	20	LaMMoV-20-probes
Lettuce virus X	LeVX	*Potexvirus*	7.21	42	20	LeVX-20-probes
Lily virus X	LVX	*Potexvirus*	5.82	28	20	LVX-20-probes
Malva mosaic virus	MaMV	*Potexvirus*	6.86	59	20	MaMV-20
Mint virus X	MVX	*Potexvirus*	5.91	30	20	MVX-20-probes
Narcissus mosaic virus	NMV	*Potexvirus*	6.96	37	20	NMV-20-probes
Nerine virus X	NVX	*Potexvirus*	6.58	47	20	NVX-20-probes
Opuntia virus X	OpVX	*Potexvirus*	6.65	30	20	OpVX-20-probes
Papaya mosaic virus	PapMV	*Potexvirus*	6.66	39	20	PapMV-20-probes
Pepino mosaic virus	PepMV	*Potexvirus*	6.45	49	20	PepMV-20-probes

Phaius virus X	PhVX	*Potexvirus*	5.82	44	20	PhVX-20-probes
Plantago asiatica mosaic virus	PlAMV	*Potexvirus*	6.13	41	20	PlAMV-20
Potato virus X	PVX	*Potexvirus*	6.44	14	20	PVX-20-probes
Schlumbergera virus X	SchVX	*Potexvirus*	6.63	24	20	SchVX-20-probes
Strawberry mild yellow edge virus	SMYEV	*Potexvirus*	5.97	29	20	SMYEV-20-probes
Tamus red mosaic virus	TRMV	*Potexvirus*	6.5	46	20	TRMV-20-probes
Tulip virus X	TVX	*Potexvirus*	6.06	26	20	TVX-20-probes
White clover mosaic virus	WClMV	*Potexvirus*	6.85	27	20	WClMV-20-probes
Yam virus X	YVX	*Potexvirus*	6.16	51	20	YVX-20-probes
Zygocactus virus X	ZyVX	*Potexvirus*	6.62	40	20	ZyVX-20-probes
Bell pepper mottle virus	BPeMV	*Tobamovirus*	6.38	34	20	BPeMV-20-probes
Brugmansia mild mottle virus	BrMMV	*Tobamovirus*	6.38	30	20	BrMMV-20
Cactus mild mottle virus	CMMoV	*Tobamovirus*	6.45	45	20	CMMoV-20-probes
Clitoria yellow mottle virus	ClitoriaYMoV	*Tobamovirus*	6.51	31	20	ClitoriaYMoV-20-probes
Cucumber fruit mottle mosaic virus	CFMMV	*Tobamovirus*	6.56	34	20	CFMMV-20-probes
Cucumber mottle virus	CuMoV	*Tobamovirus*	6.49	34	20	CuMoV-20-probes
Frangipani mosaic virus	FrMV	*Tobamovirus*	6.64	36	20	FrMV-20
Hibiscus latent Fort Pierce virus	HLFPV	*Tobamovirus*	6.43	5	20	HLFPV-20-probes
Hibiscus latent Singapore virus	HLSV	*Tobamovirus*	6.49	28	20	HLSV-20-probes
Maracuja mosaic virus	MarMV	*Tobamovirus*	6.79	31	20	MarMV-20-probes
Obuda pepper virus	ObPV	*Tobamovirus*	6.51	38	20	ObPV-20
Odontoglossum ringspot virus	ORSV	*Tobamovirus*	6.62	30	20	ORSV-20-probes
Paprika mild mottle virus	PaMMV	*Tobamovirus*	6.52	25	20	PaMMV-20-probes
Pepper mild mottle virus	PMMoV	*Tobamovirus*	6.36	38	20	PMMoV-20-probes

Plumeria mosaic virus	PLUMERIAMOSAICVIRUS	*Tobamovirus*	6.69	32	20	PLUMERIAMOSAICVIRUS-20-probes
Rattail cactus necrosis-associated virus	RattailCNaV	*Tobamovirus*	6.51	37	20	RattailCNaV-20
Rehmannia mosaic virus	RheMV	*Tobamovirus*	6.4	26	20	RheMV-20-probes
Ribgrass mosaic virus	RMV	*Tobamovirus*	6.31	23	20	RMV-20-probes
Sammons's Opuntia virus (Opuntia virus X)	OpVX	*Tobamovirus*	6.65	30	20	OpVX-20-probes
Streptocarpus flower break virus	SFBV	*Tobamovirus*	6.28	37	20	SFBV-20-probes
Tobacco mild green mosaic virus	TMGMV	*Tobamovirus*	6.36	33	20	TMGMV-20-probes
Tobacco mosaic virus	TMV	*Tobamovirus*	6.4	22	20	TMV-20
Tomato brown rugose fruit virus	ToBrRuFV	*Tobamovirus*	6.39	19	20	ToBrRuFV-20-probes
Tomato mosaic virus	ToMV	*Tobamovirus*	6.38	20	20	ToMV-20-probes
Tomato mottle mosaic virus	ToMoMV	*Tobamovirus*	6.4	17	20	ToMoMV-20-probes
Tropical soda apple mosaic virus	TSAMV	*Tobamovirus*	6.35	36	20	TSAMV-20-probes
Turnip vein-clearing virus	TVCV	*Tobamovirus*	6.31	12	20	TVCV-20-probes
Wasabi mottle virus	WMoV	*Tobamovirus*	6.3	17	20	WMoV-20-probes
Zucchini green mottle mosaic virus	ZGMMV	*Tobamovirus*	6.51	29	20	ZGMMV-20-probes
Artichoke mottled crinkle virus	AMCV	*Tombusvirus*	4.79	18	20	AMCV-20
Carnation Italian ringspot virus	CIRV	*Tombusvirus*	4.76	12	20	CIRV-20-probes
Cucumber Bulgarian virus	CBV	*Tombusvirus*	4.58	22	20	CBV-20-probes
Cucumber necrosis virus	CuNV	*Tombusvirus*	4.7	19	20	CuNV-20-probes
Cymbidium ringspot virus	CymRSV	*Tombusvirus*	4.73	11	20	CymRSV-20
Eggplant mottled crinkle virus	EMCV	*Tombusvirus*	4.77	4	20	EMCV-20-probes

Grapevine Algerian latent virus	GALV	*Tombusvirus*	4.73	11	20	GALV-20-probes
Havel River virus	HRV	*Tombusvirus*	2.09	7	20	HRV-20
Limonium flower distortion virus	LFDV	*Tombusvirus*	1.23	6	20	LFDV-20-probes
Moroccan pepper virus	MPV	*Tombusvirus*	4.77	6	20	MPV-20-probes
Neckar River virus	NRV	*Tombusvirus*	1.31	6	20	NRV-20-probes
Pelargonium leaf curl virus	PLCV	*Tombusvirus*	4.79	11	20	PLCV-20-probes
Pelargonium necrotic spot virus	PeNSV	*Tombusvirus*	4.74	8	20	PeNSV-20-probes
Petunia asteroid mosaic virus	PetAMV	*Tombusvirus*	1.24	3	20	PetAMV-20
Sitke waterborne virus	SWBV	*Tombusvirus*	1.19	10	20	SWBV -20
Tomato bushy stunt virus	TBSV	*Tombusvirus*	4.78	1	20	TBSV-20

The designed sets of E-probes are loaded and accessible for use at the EDNA MiFi™ server (http://www.edna2.okstate.edu).

3.2 *in-silico* simulation of metagenome query

A total of 117 E-probe sets were designed for three model systems, rose, cucurbits, and water. The optimal E-probe length for the targeted plant viruses varies between 20 to 30 nucleotides. Potential matches using E-probe sets of 20 to 30 bases provided optimal hit frequency at lower E-values. The queried database size, and the pairwise query-patches, are two determinants for the final BLAST E-values. Due to the genome size of the targeted plant viruses, a fixed length of 20-30nt E-probe sets increased plant virus detection by EDNA MiFi™. Resulting simulated metagenomes had the spiked percentage of target sequence (Tables 3.5 and 3.6) for the water-borne plant virus irrigation host model along with the computed p-value. The average limit of detection for *in-silico* sensitivity assay was 10^{-3}. The detection from other models host (rose and cucurbits) showed the same limit of detection.

Table 3. 5 EDNA sensitivity assay simulating single infection. Metagenomes were spiked with PepMV and detection performed using the EDNA MiFi™ detection pipeline.

METAGENOME ID	p-value	PepMV	# hits
PepMV-10-1	1.38E-47	+	49
PepMV-10-2	1.41E-47	+	49
PepMV-10-3	3.46E-09	+	49
PepMV-1-1	1.12E-09	+	49
PepMV-1-2	1.46E-10	+	49
PepMV-1-3	2.56E-12	+	49
PepMV-0.1-1	2.26E-21	+	49
PepMV-0.1-2	2.40E-23	+	49
PepMV-0.1-3	2.33E-08	+	49
PepMV-0.01-1	3.69E-10	+	41
PepMV-0.01-2	6.22E-81	+	49
PepMV-0.01-3	2.61E-02	+	40
PepMV-0.001-1	2.02E-08	+	28
PepMV-0.001-2	2.02E-08	+	20

PepMV-0.001-3	6.89E-07	+	10
PepMV-0.0001-1	8.54E-01	-	0
PepMV-0.0001-2	0.00E+00	S	0
PepMV-0.0001-3	9.62E-03	+	0
Negative	0.211203311	-	0

*Positive hits are presented by (+), negative by (-), and suspects by (S).

Table 3. 6 EDNA detection of simulated multiple infection spiked metagenomes with PepMV simulated metagenomes sensitivity assay using the EDNA MiFiTM detection pipeline.

	POTEXVIRUS			TOBAMOVIRUS			TOMBUSVIRUS		
METAGENOME ID	**p-value**	**PepMV**	**# hits**	**p-value**	**PMMoV**	**# hits**	**p-value**	**CuNV**	**# hits**
Multiple-10-1	1.38E-47	+	49	2.40E-23	+	38	2.02E-08	+	19
Multiple-10-2	1.41E-47	+	49	1.41E-47	+	38	1.12E-09	+	19
Multiple-10-3	3.46E-09	+	49	3.46E-09	+	38	3.46E-09	+	19
Multiple-1-1	1.20E-74	+	49	3.64E-55	+	38	2.70E-05	+	14
Multiple-1-2	1.03E-68	+	49	2.02E-56	+	38	2.16E-27	+	19
Multiple-1-3	1.38E-74	+	49	3.51E-55	+	38	6.11E-05	+	19
Multiple-0.1-1	6.10E-10	+	49	4.81E-08	+	38	8.67E-05	+	19
Multiple-0.1-2	8.13E-11	+	49	2.37E-09	+	38	3.08E-04	+	19
Multiple-0.1-3	1.93E-20	+	49	2.56E-08	+	38	8.67E-05	+	19
Multiple-0.01-1	3.43E-77	+	49	1.36E-54	+	38	5.84E-26	+	19
Multiple-0.01-2	2.37E-09	+	49	8.13E-11		38	3.08E-04	+	19
Multiple-0.01-3	7.62E-75	+	49	2.33E-52	+	38	3.14E-26	+	19
Multiple-0.001-1	1.78E-11	+	44	7.12E-09	+	20	6.86E-06	+	18
Multiple-0.001-2	1.26E-09	+	49	1.77E-07	+	38	4.77E-05	+	19

Multiple-0.001-3	4.59E-09	+	49	1.95E-07	+	38	3.69E-04	+	19
Multiple-0.0001-1	5.25E-01	S	7	6.22E-02	S	3	2.25E-01	S	3
Multiple-0.0001-2	3.27E-02	S	3	1.09E-01	-	8	6.40E-01	-	1
Multiple-0.0001-3	8.56E-03	S	0	7.10E-01	S	6	4.14E-02	S	5
Negative	0.331135	-	0	0.523113	-	0	0.231135	-	0

*Positive hits are presented by (+), negative by (-), and suspects by (S).

In rose and cucurbits, *in silico* simulated metagenomes with single and multiple infections of CMV and RRV showed a limit of detection of 10^{-3} (10-20 reads) in ten million reads. The limit of detection for multiple infections, the limit of detection was 10^{-3} for CMV, MNSV, PRSV, SqMV, WMV, and ZYMV. Multiple infections of RRV and PNRSV were detected in rose metagenomes to 10^{-3} reads of the pathogen in ten million reads. For limit of detection CMV and MNSV p-value of 0.08, PRSV p-value=0.02, SqMV PRSV p-value=0.001, WMV p-value=0.0006, and ZYMV p-value=0.00141 determined positive results. In the case of roses, the reported limit of detection generated suspects results at the p-value of 0.07 for RRV and PNRSV. For positive results for RRV p-value= $1.17673e^{-07}$, PNRSV=$4.81252e^{-09}$.

3.3 EDNA Water -MiFi™ validation with published SRA data

A total of nine biosample metagenomes were queried against E-probes sets designed for the three water-borne plant virus genera *Potexvirus*, *Tombusvirus*, and *Tobamovirus*. The agricultural biosample sequences generated from an Illumina HiSeq 2500 run from SRA SRR5995660-5995697 of viral metagenomics sequencing of freshwater samples SRX3151361 which were sourced from NCBI from Singapore irrigation water. Metagenomes queried by *Potexvirus* E-probe sets detected the presence of 11 plant viruses in this genus (Table 3.7). The virus with the highest number of hits (49) was PepMV in biosamples 2 and 3. Followed by CymMV with 22 hits in biosample 3, and 20 hits for YVX in biosample 5. Metagenomes queried by *Tobamovirus* E-probe sets detected the presence of 10 plant viruses in this genus (Table 3.8). The virus with the highest number of hits (38) was PMMoV in biosamples 1, 2 and 3, followed by TSAMV with 19-29 hits in biosamples 1, 2, and 3, and 20 hits for SFBV in biosample 1 and 9 hits in biosample 8. Metagenomes queried against the *Tombusvirus* E-probe sets showed positive hits to seven plant viruses in this genus (Table 3.9). The virus with the highest number of hits (8) was AMCV and HRV (7 hits) in biosample 1.

Table 3. 7 EDNA-Water MiFi™ for *Potexvirus* detection in raw SRA irrigation water metagenomics data

	POTEXVIRUS																																
METAGENOME ID	AlVX	AlsVX	AltMV	AV-3	BaMV	CVX	CsCMV	CsVX	ClYMV	CymMV	FoMV	HVX	HdRSV	LaMMoV	LeVX	LVX	MaMV	MVX	NMV	NVX	OpVX	PapMV	PepMV	PhVX	PlAMV	PVX	SchVX	SMYEV	TRMV	TVX	WClMV	YVX	ZyVX
Biosample1										10																							
Biosample2																				8			49										
Biosample3			19				16			22					15					14			49										
Biosample4							8																										
Biosample5																							9									20	
Biosample6																																	
Biosamaple7																							8										
Biosample8						5						3																					
Biosample9		5																										7					
Negative																																	

*see complete virus name at Table 3.4.

Table 3. 8 EDNA-Water MiFi™ for *Tobamovirus* detection in raw SRA irrigation water metagenomics data

	TOBAMOVIRUS																																
METAGENOME ID	BPeMV	BrMMV	CMMoV	ClitoriaYMoV	CFMMV	CuMoV	HLFPV	HLSV	MarMV	ORSV	PaMMV	PMMoV	PLUMERIAMOSAICVIRUS	RattailCNaV	RheMV	RMV	OpVX	SFBV	TMGMV	TMV	ToBrRuFV	ToMV	ToMoMV	TSAMV	TVCV	WMoV	ZGMMV						
Biosample1										16		38	16					20					13	29									
Biosample2												38												19									
Biosample3												38												19									
Biosample4																																	
Biosample5																																	
Biosample6																																	
Biosamaple7																9								12									
Biosample8		8		9														9	10														
Biosample9				13																													
Negative																																	

*see complete virus name at Table 3.4.

Table 3. 9 EDNA-Water MiFiTM for *Tombusvirus* detection in raw SRA irrigation water metagenomics data

	TOMBUSVIRUS													
METAGENOME ID	AMCV	CIRV	CBVCuNV	CymRSV	EMCV	GALV	HRV	LFDV	MPV	NRV	PLCV	PeNSV	PetAMV	SWBV	TBSV
Biosample1	8						7								
Biosample2									6	6					
Biosample3															
Biosample4															
Biosample5															
Biosample6											5				
Biosamaple7		5	5												
Biosample8			7												
Biosample9															
Negative															

*see complete virus name at Table 3.4.

3.4. Blind test using EDNA Rose-MiFi™ validation

A total of 23 samples were analyzed by EDNA-Rose MiFi™. The metagenomes showed low virus concentration, the maximum number of hits (Table 3.10) of 16 for RoYMV for metagenome R4_S8_R1_001.fastq.00.0_0.cor. The E-probe set for RMCV has a higher presence of hits in 13 analyzed metagenomes (Table 3.10). The negative control did not show hits in any of the E-probe sets.

Table 3. 10 Virus detection in raw Illumina HiSeq generated metagenomes sourced by Plant Foundation Service UC. Davis queried against EDNA-Rose MiFi™.

METAGENOME ID	ACLSV	ApMV	ASGV	ArMV	BCRV	INSV	IYSV	PDV	PNRSV	RpRSV	RMCV	RoLCuV	RLRaV	RRV	RSDaV	RoYMV	RoYLV	RYVV	SLRSV	TRSV	TSV	TSWV	TYRV
R1_S1_R1_001											1												
R10_S6_R1_001											1												
R11_S10_R1_001					2													9					
R12_S3_R1_001											5							12					
R13_S4_R1_001											5							9					
R14_S8_R1_001																							
R15_S9_R1_001																							
R16_S10_R1_001											5												
R18_S11_R1_001											5												
R2_S2_R1_001											1					1							
R21_S1_R1_001											5					1							
R22_S2_R1_001											1												
R24_S3_R1_001																							
R26_S4_R1_001											5												
R27_S1_R1_001															9								
R28_S2_R1_001											5												
R3_S5_R1_001											1					1	3						
R4_S8_R1_001.fastq.00.0_0.cor																16							
R5_S7_R1_001											1												
R6_S6_R1_001																							

R7_S1_R1_001			5			
R8_S5_R1_001	2			2		
R9_S4_R1_001			1		11	1
Negative						

3.6. *In vitro* sensitivity

In vitro analytical sensitivity of serially diluted plasmid carrying the ArMV RNA2 genome using Illumina NextSeq and EDNA MiFi™. The screening of the generated metagenomes was made with a single set of E-probes belonging to ArMV. Twenty-seven metagenomes were queried against the ArMV-60nt E-probes were analyzed for sensitivity. The metagenomes were parsed at the optimal e-value for viruses (eval5), and the recommended number of hits for quantification (250) was used. As expected, the higher number of hits was registered for the higher concentrations of 55 and 27 ng (Table 3.11) respectively. The quantified limit of detection was 10pg (10^{-3}) equivalent to 2 hits with 2, 1 and 0 hits in the three metagenomes of the same concentration respectively.

Table 3. 11 EDNA MiFi™ detection of ArMV RNA2 genome raw metagenomes in serially diluted plasmid sensitivity assay.

No.	Sample [ng]	Illumina Seq ID	ArMV-RNA2-60_hits
1	55 ng ArMV-RNA2	1_S1_LALL_R1_001.fastq.gz	500
2	55 ng ArMV-RNA2	2_S2_LALL_R1_001.fastq.gz	500
3	55 ng ArMV-RNA2	3_S3_LALL_R1_001.fastq.gz	500
4	27 ng ArMV-RNA2	4_S4_LALL_R1_001.fastq.gz	500
5	27 ng ArMV-RNA2	5_S5_LALL_R1_001.fastq.gz	500
6	27 ng ArMV-RNA2	6_S6_LALL_R1_001.fastq.gz	500
7	0.1 ng ArMV-RNA2	7_S7_LALL_R1_001.fastq.gz	35
8	0.1 ng ArMV-RNA2	8_S8_LALL_R1_001.fastq.gz	14
9	0.1 ng ArMV-RNA2	9_S9_LALL_R1_001.fastq.gz	27
10	0.01 ng ArMV-RNA2	10_S10_LALL_R1_001.fastq.gz	2
11	0.01 ng ArMV-RNA2	11_S11_LALL_R1_001.fastq.gz	1
12	0.01 ng ArMV-RNA2	12_S12_LALL_R1_001.fastq.gz	0
13	0.001 ng ArMV-RNA2	13_S13_LALL_R1_001.fastq.gz	0
14	0.001 ng ArMV-RNA2	14_S14_LALL_R1_001.fastq.gz	0
15	0.001 ng ArMV-RNA2	15_S15_LALL_R1_001.fastq.gz	0
16	0.0001 ng ArMV-RNA2	16_S16_LALL_R1_001.fastq.gz	0
17	0.0001 ng ArMV-RNA2	17_S17_LALL_R1_001.fastq.gz	0
18	0.0001 ng ArMV-RNA2	18_S18_LALL_R1_001.fastq.gz	1

19	0.00001 ng ArMV-RNA2	19_S19_LALL_R1_001.fastq.gz	0
20	0.00001 ng ArMV-RNA2	20_S20_LALL_R1_001.fastq.gz	0
21	0.00001 ng ArMV-RNA2	21_S21_LALL_R1_001.fastq.gz	0
22	0.000001 ng ArMV-RNA2	22_S22_LALL_R1_001.fastq.gz	0
23	0.000001 ng ArMV-RNA2	23_S23_LALL_R1_001.fastq.gz	0
24	0.000001 ng ArMV-RNA2	24_S24_LALL_R1_001.fastq.gz	0
25	0.0000001 ng ArMV-RNA2	25_S25_LALL_R1_001.fastq.gz	0
26	0.0000001 ng ArMV-RNA2	26_S26_LALL_R1_001.fastq.gz	0
27	0.0000001 ng ArMV-RNA2	27_S27_LALL_R1_001.fastq.gz	0

4. Discussion

International trade of goods and germplasm increases the chances of the introduction of some exotic pests. Agriculture, horticulture, and forestry are threatening by the introduction and establishment of plant viruses and their vectors. In the US, imported germplasm and live plants are the more likely pathways for introduction of non-native plant pathogens (Bradley et al., 2012). Since 2011, the United States Department of Agriculture – Animal and Plant Health Inspection Service (USDA-APHIS) implemented a risk-based inspection process to check the high-risk trade of plants and goods. Even though, this effort, to inspect each shipment is time-consuming for perishable goods (USDA-APHIS, 2011b). At present, molecular based detection protocols have addressed plant pathogen identification at the port of entry for detection of plant pathogens that do not show symptoms during the inspection. However, some of the molecular protocols cannot detect all pathogens, leaving undetected plant pathogens in the plant material (Liebhold et al., 2012). High Throughput Sequencing (HTS) of plant material and goods may facilitate a complete detection of complete phytobiome in imported germplasm, allowing a timely diagnosis before post entry quarantine introduction and establishment of a plant pathogen.

Moreover, biological questions regarding plant-pathogen-vector interactions can be addressed by analysis of HTS data. Metagenomic and transcriptomics can elucidate all the organisms and their expressed genes during an infection process (Tatineni et al., 2020). Notably, HTS not only offers a robust and accurate detection of all the pathogens in a sample, also allows researchers to infer

into the biological development of an infection. Over the last ten years, EDNA was developed from a proof-of-concept to a tangible detection and diagnostic tool. The EDNA-MiFi™ pathogen detection pipeline overcomes the discovery timely consuming traditional bioinformatic workflows, enabling the processing of raw reads. This is an advantage at the ports of entry to screen a suspect sample (Stobbe et al., 2013; Stobbe 2014; Blagden et al., 2016, Espindola et al., 2015; Espindola et al., 2018). This study provides the first validation of EDNA-MiFi™ for complete plant virome detection in three model systems, rose, cucurbits, and water. Reported plant viruses infecting rose, cucurbits, and water can be detected using specific designed E-probe sets. The detection process from the analyzed raw metagenomic reads is reduced in time due to the specificity of the E-probe datasets taking a few minutes to retrieve a detection report. Unassembled metagenomic or transcriptomic HTS datasets can be screened within minutes using EDNA.

Initially EDNA research focused on single pathogen detection, and was later modified to detect transcripts expression (Stobbe et al., 2014; Blagden et al., 2016; Espindola et al, 2018). The sensitivity of detection of host plant viruses was addressed in a related study with a different approach for sensitivity validation by Rydzak et al. (2018, not published). This study addresses the complete plant virome detection of a pathosystem and water source, as well as, a method to approach the analytical sensitivity of EDNA MiFi™ not previously described. In this study, the detection was made using transcriptomic datasets of known concentration that provide a proportional understanding of the pathogen within a host. EDNA- MiFi™ specific E-probes were generated for 22 plant viruses reported infecting rose, 15 viruses infecting cucurbit, and 80 known plant-waterborne viruses in three genera. The use of highly specific targeted E-probes designed for detection of each virus adds a technical advantage to EDNA because allows a rapid and efficient computational pathogen detection since it does not require assembling a large output of sequenced reads generated by the HTS sequencing.

EDNA MiFi™ was tested using simulated metagenomes *in silico*, published SRA data, blind metagenomic content, metagenomes generated *in-vitro* from serially diluted targets which provides estimates of E-probe detection sensitivity.

Simulated infected metagenomes of the three model tested systems have a limit of detection of 10^{-3} in ten million reads. The validation with *in-silico* metagenomes for sensitivity allowed to further test EDNA MiFi™ to detect plant viruses in published SRA metagenomic datasets and blind test comparison with provided metagenomes from the Plant Foundation Service (U.C. Davis). The blind test comparison of EDNA-MiFi™ between results from the traditional bioinformatic pipeline provides accurate detection of the rose virome. In a personal communication with Dr. Golino and Dr. Al Rwahnih (Foundation Plant Services, University of California Davis, CA), there was a correlation of results between their detection system and EDNA MiFi™. The Plant Foundation Service results provide insight into the new viruses that need to be added to the EDNA-Rose MiFi™ E-probe dataset since their approach allowed for virus discovery whereas EDNA MiFi™ queries the metagenomes against a specific E-probe dataset.

The sensitivity validation using controlled *in-vitro* system (plasmid pUC57 containing the RNA2 ArMV) is a robust attempt to determine the *in-vitro* limit of detection of EDNA-MiFi™. Comparing the limit of detection of *in silico* simulated metagenomes and *in vitro* obtained metagenomes (Illumina NextSeq) resulted in a similar limit of detection down to 10 pg (10^{-3}) indicating the sensitivity of the E-probes. These results provide a framework for comparison with standardized methods for plant virus detection. EDNA-MiFi™ is a valuable tool for infectious plant pathogens detection in a host. This research confirmed the extensive capability of the EDNA-MiFi™ for detection, and potentially to be deployed in ports of entry, use by biosecurity agencies.

References

Agrios, G. N. (2005). Plant diseases caused by nematodes. *Plant Pathology*, *4*, 565-597.

Blagden, T., Schneider, W., Melcher, U., Daniels, J., & Fletcher, J. (2016). Adaptation and validation of E-probe diagnostic nucleic acid analysis for detection of Escherichia coli O157: H7 in metagenomic data from complex food matrices. *Journal of food protection*, *79*(4), 574-581.

Bradley, B. A., Blumenthal, D. M., Early, R., Grosholz, E. D., Lawler, J. J., Miller, L. P., ... & Ibanez, I. (2012). Global change, global trade, and the next wave of plant invasions. *Frontiers in Ecology and the Environment*, *10*(1), 20-28.

Decker, C. J., & Parker, R. (2014). Analysis of double-stranded RNA from microbial communities identifies double-stranded RNA virus-like elements. *Cell reports*, *7*(3), 898-906.

Espindola, A., Schneider, W., Hoyt, P. R., Marek, S. M., & Garzon, C. (2015). A new approach for detecting fungal and oomycete plant pathogens in next-generation sequencing metagenome data utilizing electronic probes. *International journal of data mining and bioinformatics*, *12*(2), 115-128.

Espindola, A., & Cardwell, K. F. (2018, October). Third-generation sequencing and EDNA for detection of aflatoxin production in the soil. In *PHYTOPATHOLOGY* (Vol. 108, No. 10, pp. 69-69). 3340 PILOT KNOB ROAD, ST PAUL, MN 55121 USA: AMER PHYTOPATHOLOGICAL SOC.

Gu, X., Tay, Q. X. M., Te, S. H., Saeidi, N., Goh, S. G., Kushmaro, A., ... & Gin, K. Y. H. (2018). Geospatial distribution of viromes in tropical freshwater ecosystems. *Water Research*, *137*, 220-232.

Li, Z., Zhang, Z., Yan, P., Huang, S., Fei, Z., & Lin, K. (2011). RNA-Seq improves the annotation of protein-coding genes in the cucumber genome. *BMC genomics*, *12*(1), 540.

Liebhold, A. M., Brockerhoff, E. G., Garrett, L. J., Parke, J. L., & Britton, K. O. (2012). Live plant imports: the major pathway for forest insect and pathogen invasions of the US. *Frontiers in Ecology and the Environment*, *10*(3), 135-143.

Matthews, R. E. F. (2019). *Diagnosis of plant virus diseases*. CRC Press.

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., Škorić, D. J. F. i. m. (2017). A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. *8*, 45.

Mine, A., Sato, M., & Tsuda, K. (2014). Toward a systems understanding of plant-microbe interactions. *Frontiers in plant science*, *5*, 423.

Moreno, A. B., & López-Moya, J. J. (2020). When viruses play team sports: Mixed infections in plants. *Phytopathology*, *110*(1), 29-48.

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., ... & Kyrpides, N. C. (2016). Uncovering Earth's virome. *Nature*, *536*(7617), 425-430.

Richter, D.C., Ott, F., Auch, A.F., Schmid, R., and Huson, D.H. 2008. MetaSim—A Sequencing Simulator for Genomics and Metagenomics. PLoS ONE 3(10):e3373. doi:10.1371/journal.pone.0003373

Satya, R. V., Zavaljevski, N., Kumar, K., and Reifman, J. 2008. A high-throughput pipeline for designing microarray-based pathogen diagnostic assays. BMC Bioinformatics 9(1):1.

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., ... & Schneider, W. (2013). E-probe Diagnostic Nucleic acid Analysis (EDNA): a theoretical approach for handling of next-generation sequencing data for diagnostics. *Journal of microbiological methods*, *94*(3), 356-366.

Stobbe, A. H., Schneider, W. L., Hoyt, P. R., & Melcher, U. (2014). Screening metagenomic data for viruses using the e-probe diagnostic nucleic acid assay. *Phytopathology*, *104*(10), 1125-1129.

Tatineni, S., Stewart, L. R., Sanfaçon, H., Wang, X., Navas-Castillo, J., & Hajimorad, M. R. (2020). Fundamental Aspects of Plant Viruses− An Overview on Focus Issue Articles. *Phytopathology*, *110*(1), 6-9.

USDA-APHIS (US Department of Agriculture-Animal and Plant Health Inspection Service). 2011b. *Importation of plants for planting: risk-based sampling and inspection approach and propagative monitoring and release program (Docket No APHIS-2011-0092)*. Washington, DC: Government Printing Office. Federal Register Notice **76(** 203): 65165– 66.

CHAPTER IV

DETECTION OF THE GLOBAL ROSE VIROME OF GARDEN ROSES IN OKLAHOMA

Abstract

The rose industry is adversely affected by viral infections; twenty-six viruses have been reported to infect rose worldwide. Early viral infections in garden roses are generally not detected as the rose plant does not rapidly display symptoms. Moreover, similar symptomatology by multiple viral infections may be confusing and challenging during diagnosis. High Throughput Sequencing (HTS) using the portable sequencer MinION™ allows the rapid sequencing of metagenomes with broad detection capability to identify multiple viruses in a sample. Electronic Diagnosis Nucleic acid Analysis (EDNA) is a bioinformatic pipeline that can identify all the viruses reported to infect roses. Virus E-probes, ranging from 20 to 30 nucleotides in length, were designed using EDNA Mi/Fi™ software. Rose metagenomic searches *in-vitro* were validated and optimized using samples from a Rose Rosette varietal disease resistance trial located in Tulsa and Perkins, Oklahoma, USA. Additionally, E-probes proved to be a rapid diagnostic method for specific and discrete analysis of the metagenomic outputs of rose samples when compared with other exploratory bioinformatics analyses. The analyzed metagenomes showed mixed rose viral presence with a higher titer of *Rose rosette virus* (RRV) and *Tomato yellow ring virus* (TYRSV) among the varieties.

1. Introduction

Rose (*Rosa* spp.) is an economically important ornamental plant worldwide (Milleza et.al., 2013). This specialty crop worth about $400 million in the U.S. (AIPH, 2016). At present, 130 species of rose have been reported and about ten of them are asexually propagated as a genetic source in breeding programs (Leus et.al., 2018). Cultivars cegetatively propagated are susceptible to plant pathogens, and viruses are among the most detrimental to rose production worldwide. Diseased plants affect not only aesthetics but also photosynthesis, nutrient uptake, and flowering yield, while producing or gardening (Secor et.al., 1977).

Plant virus infection in roses produce symptoms that include excessive thorniness, yellow mosaic, general chlorosis, witches broom/rosette type growth, flecking of veins, curling leaves, leaf drop, shoot dieback, cracking of the leaf rachis, discoloration of the inner bark, pitting in the wood of mature canes, vein clearing and balling of leaves, fine line patterned rings on leaves, and necrosis (Di Bello, et.al., 2015; Laney, et.al., 2011; Windham, et.al., 2014). Moreover, some viral symptoms can be confused with nutrient deficiencies. There are leaf deformation, chlorosis, mottling, mosaic symptoms similarly caused by abiotic stress, reduced nutrition or chemical injury (Islam, 2017). Due to the extensive and related symptomatology within a virus infection, it is difficult to determine which virus is responsible for specific symptoms (Gomez et.al., 2009). Table 4.1 shows the list of plant viruses used for this study.

Table 4. 1 Reported plant viruses infecting roses and description of important virus features description.

Virus	Genus	Acronym	Nucleic acid	polyA	Description
Apple chlorotic leaf spot virus	Trichovirus	ACLSV	Linear, ssRNA(+) genome	Yes	Monopartite no-enveloped particle
Apple mosaic virus	Ilarvirus	ApMV	Tripartite, linear ssRNA(+) genome	No	Segmented, tripartite non-enveloped particle
Apple stem grooving virus	Capillovirus	ASGV	Linear, ssRNA(+) genome	Yes	Monopartite no-enveloped particle
Arabis mosaic virus	Nepovirus	ArMV	Segmented, bipartite linear ssRNA(+)	Yes	Bi-partite non-enveloped icosahedral particles
Blackberry chlorotic ringspot virus	Ilarvirus	BCRV	Tripartite, linear ssRNA(+) genome	No	Segmented, tripartite non-enveloped particle
Impatiens necrotic spot virus	Orthotospovirus	INSV	Tri-partite linear ssRNA(-)	No	Three genomic segments, encapsidated in one particle
Iris yellow spot virus	Orthotospovirus	IYSV	Tri-partite linear ssRNA(-)	No	Three genomic segments, encapsidated in one particle
Prune dwarf virus	Ilarvirus	PDV	Tripartite, linear ssRNA(+) genome	No	Segmented, tripartite non-enveloped particle
Prunus necrotic ringspot virus	Ilarvirus	PNRSV	Tripartite, linear ssRNA(+) genome	No	Segmented, tripartite non-enveloped particle
Raspberry ringspot virus	Nepovirus	RpRSV	Segmented, bipartite linear ssRNA(+)	Yes	Bi-partite non-enveloped icosahedral particles
Rose cryptic virus 1	Unclassified Deltapartitivirus	RCV-1	Ti-partite segmented dsRNA	Unknown	Monocistronic dsRNAs segmented particle
Rose leaf curl virus	Begomovirus	RoLCuV	Circular, ssDNA (+) genome	No	Bi-partite non-enveloped icosahedral DNA A-B particles
Rose leaf rosette-associated virus	Closterovirus	RLRaV	Linear, ssRNA(+) genome	No	Monopartite no-enveloped particle, 7 subgenomic RNAs
Rose rosette virus	Emaravirus	RRV	Linear, ssRNA(-) genome	No	Seven RNAs enveloped in a single particle
Rose spring dwarf-associated virus	Luteovirus	RSDaV	Linear, ssRNA(+) genome	No	Monopartite no-enveloped particle, 2 subgenomic RNAs

Rose yellow leaf virus	Tombusvirus	RoYLV	Linear, ssRNA(+) genome	No	Monopartite no-enveloped particle, 2 subgenomic RNAs
Rose yellow mosaic virus	Roymovirus	RoYMV	Linear, ssRNA(+) genome	Yes	Monopartite no-enveloped particle
Rose yellow vein virus	Rosadnavirus	RoYVV	Circular, dsDNA	Yes	Monopartite no-enveloped particle
Strawberry latent ringspot virus	Waikavirus	SLRSV	Linear, ssRNA(+) genome	Yes	Monopartite no-enveloped particle
Tobacco ringspot virus	Nepovirus	TRSV	Segmented, bipartite linear ssRNA(+)	Yes	Bi-partite non-enveloped icosahedral particles
Tobacco streak virus	Ilarvirus	TSV	Tripartite, linear ssRNA(+) genome	No	Segmented, tripartite non-enveloped particle
Tomato ringspot virus	Nepovirus	ToRSV	Segmented, bipartite linear ssRNA(+)	Yes	Bi-partite non-enveloped icosahedral particles
Tomato spotted wilt virus	Orthotospovirus	TSWV	Tri-partite linear ssRNA(-)	No	Three genomic segments, encapsidated in one particle
Tomato yellow ring virus	Tombusvirus	TYRV	Linear, ssRNA(+) genome	No	Monopartite no-enveloped particle, 2 subgenomic RNAs

In production environments plant health depends on the overabundance of synergy between microorganisms (Rodriguez et.al., 2018; Parasuraman et.al., 2019). Virus–host interaction may be asymptomatic due to a genomic association (Amrine, 1996; Epstein & Hill, 1999; Thomas & Scott, 1953; Roossinck, 2015; 2017) or overlapping symptoms may occur by higher titer of other infecting viruses in multiple infections (Syller, 2012; Gomez et.al., 2009). Plant virus-host ranges depends on a co-evolution based on interactions between the pathogen and their hosts (Pagan & García-Arenal, 2018). Viral diversity and synergistic associations with their plant hosts determined the host range of the virus (Tollenaere & Laine, 2016; Pagan & García-Arenal, 2018). Roses are hosts for 26 reported viruses worldwide (Horst, 2007; Converse & Bartlett, 1979; Fulton, 1970; Moury, et al., 2001; Tzanetakis, et. al., 2005, 2006, 2007; Rivera et.al., 2010; Golino, et.al., 2011; Di Bello, et.al., 2015; He, et.al., 2015). Vegetative propagation of infected plant material by budding or grafting of rootstocks is the main mechanism spreading rose viruses and establishment of viruses in new ecosystems and production areas (Secor et.al., 1977; Pscheidt & Rodriguez 2016; Byrne et.al., 2018). For example, *Rosa x damscena* was propagated by interspecific hybridization of stem grafting. This cultivar then was stabilized by vegetative propagation, but before being grafted none of the grafted tissue was tested to be virus-free. Therefore, the vegetative propagation led also to virus propagation in new cultivars (Tešanović et al., 2018).

The aim of rose breeders is to improve roses by introgression for genetic resistance or immunity to phytopathogens (Byrne et.al., 2018). The USDA National project to combat Rose rosette disease is a multidisciplinary project (17 scientists in six states) oriented to control the disease caused by a single virus, *Rose rosette virus* (RRV, genus *Emaravirus*) and its vector, the eriophyid mite *Phyllocoptes fructiphilus* Keifer. Field trials have been established to assess rose varieties resistance against RRV (Byrne et.al., 2017). The Combatting Rose Rosette Disease Research Team started to assess resistance to Rose rosette disease of 1252 roses, and to present 30 accessions appeared to be possess resistance, and 18 appeared to display intermediate level of resistance after

2 to 3-year trials of assesment at open fields (David Byrne, personal communication). Even though there are studies focused on testing resistance to a single pathogen (RRV) (Olson et al., 2017; Byrne et al., 2017; Novick et al., 2017; Byrne et al., 2018), roses in field trials may be infected by multiple viral infections in addition to RRV. Released marketable garden rose varieties and hybrid cultivars may have mixed virus infections (Debener & Byrne, 2014, Moreno & Lopez-Moya, 2020). There is a need to study the extend of multiple infections and the differences among rose cultivars and multiple plant-virus infections.

Multiple pathogen detection is challenging for plant disease diagnosticians (Abdulla et.al., 2017) and diagnostic methods that target multiple viruses are needed. Early pathogen detection is the main strategy for viral disease control (Sankaran et.al., 2010). At present, available detection methods for plant virus detection include serological (Jordan et.al., 2016), and nucleic–acid-based methods (Babu et al., 2017). Even though current detection standardized methods are sensitive, reliable, accurate and specific, they lack a multiple target detection and discrimination capability. Moreover, some of these methodologies detect no more than one to five virus genes at the time (Fang & Ramasamy, 2015, Dobhal et.al., 2016, Babu et al., 2016).

One of the objectives of this study is to overcome the current diagnosis paradigm where one diagnostic method is designed to detect one pathogen. Recent developments in *omics* technologies such as High, Throughput Sequencing (HTS) and bioinformatics pipelines are able to elucidate the components of viral communities in a host (Adams et.al., 2009; Wu et.al., 2010; Seguin et.al., 2014; He et. al., 2015; Massart et.al., 2017). In general bioinformatics pipelines require computer-intensive and trained bioinformaticians. Detection is time-consuming and cause delay diagnosis (Massart et.al., 2017). A molecular and genomic detection tool combined with bioinformatics for the detection of all known plant viruses reported to infect roses is proposed. Electronic Diagnostic Nucleic acid Analysis (EDNA MiFi^{TM}) is a phytobiome diagnostic tool. EDNA was designed to detect plant pathogens (viruses, bacteria, fungi, and oomycetes) (Stobbe et.

al., 2013; Espindola et.al., 2015; Blagden et.al., 2016). Electronic probes (E-probes) are pathogen-specific sequences used as genomic fingerprints that are carefully designed, validated and uploaded to the MiFi™ platform. EDNA is an applied method that combines HTS with bioinformatics pipelines for a graphical user interphase MiFi™ detection all pathogens associated with a crop at once. In this case, 22 virus infecting roses by generating specific E-probes. To provide a more comprehensive tool for diagnostics and biological characterization of rose virus infections, the following research objectives were advanced:

1. To assess the presence of plant viruses related to the garden rose cultivars (susceptible and resistant) by studying 12 garden rose varieties metagenomes (Oxford Nanopore-MinION™).

2. To test and validate two novel detection methods (EDNA-Rose MiFi™ and Multiple RT-qPCR) able to detect multiple plant viruses in 15 garden rose (susceptible and resistant) cultivars.

3. To compare the detection EDNA-Rose (MiFi ™) with traditional bioinformatics for multiple virus detection.

4. To quantify the titer of RRV among the 12 garden rose varieties by EDNA-Rose MiFi™ and to compare between resistant and susceptible cultivars.

2. Materials and methods

2.1 Field sample collection

Samples of plant tissue of garden rose with characteristic symptoms of viral infection were collected during the summer 2018. The samples were collected from field trials of the Combatting Rose Rosette project (Byrne et al., 2017). A field survey was done in Payne County at Perkins and Tulsa County, Oklahoma. Plant tissue from leaves and young stems were collected from 15 planted cultivars. Cultivar selection was based on symptomatology in the field. Viral symptoms were confirmed in leaves and by excessive thorns in the collected stems. A total of 50

mg of leaf tissue was stored in 1.5 mL nuclease-free tubes (3 tubes per cultivar) and saved at −80 °C until processing.

2.2. Primer design

Specific primers were designed to target a region of the viral genome. Viral genomes (Table 4.1) from genes that encode the RNA dependent RNA polymerase (RdRp), and the capsid protein genes of the virus from were retrieved the National Center for Biotechnology Information (NCBI). The sequences were aligned using MEGA6 (Tamura et al., 2013) and BioEdit (Hall et al., 2011) using the MUSCLE algorithm (Edgar, 2014). In order to group the virus sequences, a Maximum Likelihood phylogenetic tree was constructed with 1000 bootstrap values (Murshudov et al., 1997). Consensus sequences were used to design primers with specific range of detection, as follow:

Primer3 software was used for primer design, this application takes in consideration the thermodynamic features of the sequences (Untergasser et al., 2012). In order to check the primer specificity, their sequences were uploaded to PrimerBLAST (Ye et al., 2012) software. Thermodynamic features were analyzed in OligoAnalyzer (Kuulasmaa, 2002) and Mfold software (Zuker, 2003) in order to ensure less energy to form secondary and self-complementarity structures. Primer pairs were selected based on different amplicon sizes that allow product differentiation by High-Resolution Melting (HRM). uMelt was the software (Dwight et al., 2011) used to analyze and to predict the HRM.

2.3. Multiplex qRT-PCR HRM analysis

Multiplex RT-qPCR HRM assays were performed in 10 µl reaction volumes consisting of 5 µl HotStart. Master Mix (New England Biolabs, Ipswich, MA), 1 µl of LCgreen (Biofire, Salt Lake City, UT), 0.5 µl of each forward and reverse primer (7.5 µM), 2 µl of cDNA template, and 1 µl nuclease-free water (Ambion, Austin, TX, USA). Multiplex RT-PCR coupled with HRM was performed in a Rotor-Gene thermal cycler (QIAGEN, Hilden, Germany). The selected cycling parameters were: initial denaturation of 94ºC for 4 min followed by 35 cycles of denaturation at

94ºC for 20 s, annealing at 54ºC for 60 s, extension at 72ºC for 40 s, and a final extension at 72ºC for 3 min. Finally, a 10 µl of the amplified PCR product was denatured for HRM from 65 to 99 ºC, temperature range. Positive, negative (non-template control; water) controls and healthy tissue (25ng) were included. A sensitivity assay from 100ng to 1 fg was developed

2.4. Library preparation and sequencing

Twelve samples were selected based on the multiple viral infections found by RT-qPCR HRM, for Oxford Nanopore sequencing. RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). dscDNA was amplified using NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Kit (New England Biolabs, Ipswich, MA). dsDNA was quantified by Quant-iT™ PicoGreen™ (ThermoFisher Scientific, Waltham, MA).

The metagenome library was synthesized according to the protocol of Oxford Nanopore manufacturer's. Briefly, the library was prepared as the template strand was amplified by a transposase enzyme which synthesize long fragments with adapters ready to be sequenced by MinION™ (Oxford Nanopore Technologies, Oxford, UK).). The library was kept on ice (-20 ºC) until ready to load on the MinION device. Barcoding and ligation of the library were performed according to the Oxford Nanopore's protocol (SQK-LSK108). The barcoded libraries were equimolar-pooled before adding the adapters. The final library was stored at –20°C until MinION™ sequencing Oxford Nanopore Technologies, Oxford, UK).

Sequencing was performed with the MinION™ using the flow cell (FLO-Min106 R9.4; Oxford Nanopore Technologies, Oxford, UK). Platform Quality Check (QC) was performed to determine the number of active pores available in the flow cell. There were 1250 active pores after QC. After priming the flow cell following the per manufacturer's protocol, the pooled libraries were loaded onto the SpotON port of the flow cell using the Library Loading Beads (LLB; Oxford Nanopore Technologies).

2.5. Rapid virus detection using EDNA-Rose (MiFi™) database

EDNA-Rose MiFi™ tested the accurate identification of rose virus signatures within the metagenomic samples. The parameters selected for EDNA-Rose Mi/Fi™ use were 90% percent identity and query coverage for assessment of the twelve metagenomics databases generated by MinION™ sequencing. Hit frequencies between raw reads with E-probes were recorded for each of the selected varieties. Data on hit frequencies were analyzed with Tukey's HSD test and pairwise T-test at Pvalue= 0.05. EDNA-Rose aligned raw metagenomic reads and parsed those to match completely with the targeted infecting rose viruses E-probe sets in the database. Also, EDNA-Rose generated a statistical correlation between the number of matches in each sample as a semi-quantification tool of the virus presence in the host.

2.6. EPI2ME What is in my pot? detection database

A traditional standardized software was used to compare the obtained results with a commercial metagenomics identifier. Samples were demultiplexed into twelve barcodes using the Albacore algorithm onto fastq files. The EPI2ME software package (Oxford Nanopore Technologies, UK) was used to determine the taxonomic identity of the analyzed reads. For this purpose the software What is in my pot? (WIMP) was used. The EPI2ME database reported all the taxonomical nodes to form a phylogenetic tree based upon identification down to species level found in each barcode.

2.7. Virus HTS detectionby traditional pipeline

To validate EDNA-Rose MiFi™, a traditional bioinformatics approach for taxonomic identification and virus discovery of the reads was used. Figure 4.1 describes the workflow used for a traditional bioinformatic approach as well as EDNA-Rose MiFi™.

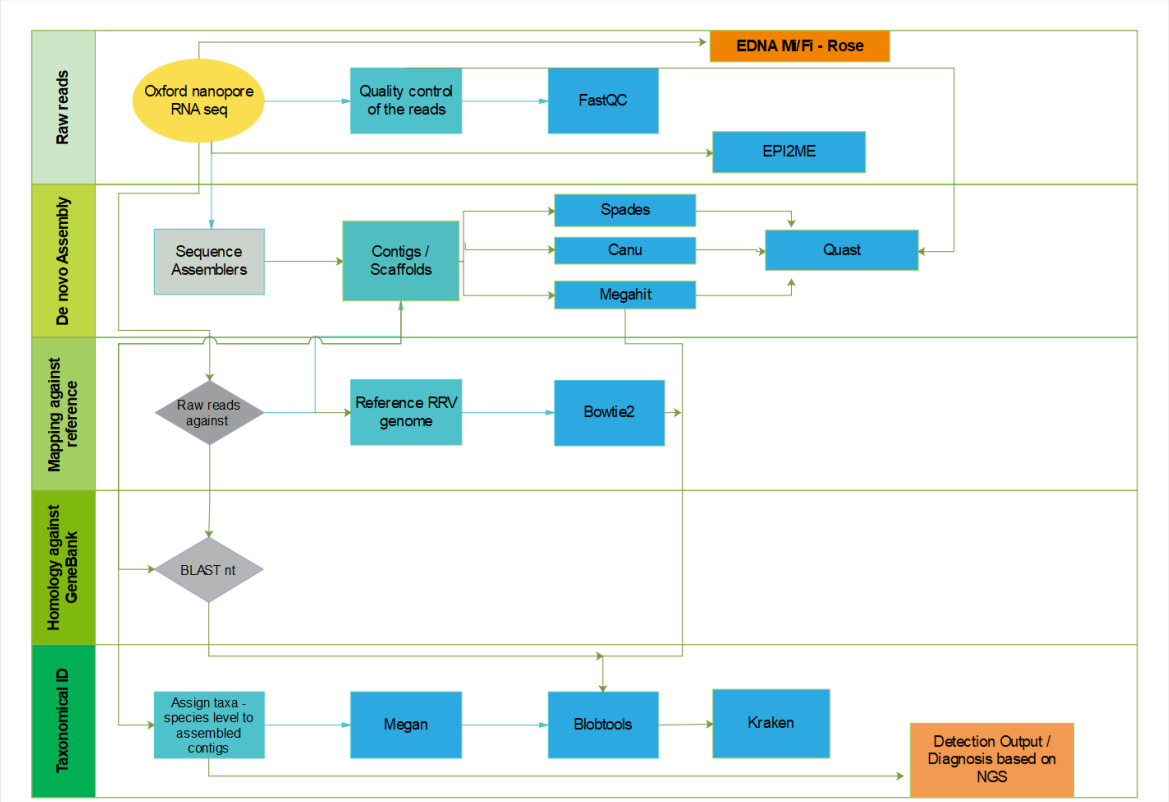


Figure 4. 1 Flowchart describing the traditional Bioinformatics approach for virus detection and discovery compared with the EDNA-Rose pipeline, and EPI2ME. Showing main steps including *de novo* assembly, mapping against the reference genome (RRV), homology exploration against GeneBank and finally taxonomical identification.

2.7.1. Quality control of raw reads

First, high quality reads were tested through a quality control step by using the FastQC program. Quality scores (Phred scores) were assessed by determining the sequence length distribution, adapter, and *k-mer* content. An HTML report was generated to visualize and to determine the quality scores across all databases.

2.7.2. Assembling

Oxford Nanopore reads were assembled using three commonly used assemblers. The assembling process was used to develop large contigs and scaffolds. Genome assemblers SPAdes v.3.11.1 (Bankevich et.al., 2012), MEGAHIT v.1. (Li et.al., 2016), and Canu v.1.8 (Koren et.al., 2017) was used to generate long contigs. The use of different assemblers allowed the selection of the best *k-mer* size as well as the N50 of the generated contigs. The quality of the generated contigs and scaffold was verified by Quast (Gurevich et.al., 2013).

2.7.3. Mapping reads against the reference genome

Generated contigs were mapped using the program Bowtie2 (Langmead & Salzberg, 2012) selecting the highest sensitivity setting. Two approaches were followed. First, mapping the raw reads against the generated contigs. Second, contigs were mapped against the reference RRV reported genome (NC_015298.1, NC_015299.1, NC_015300.1, NC_015301.1, NC_034979.1, NC_034980.1 NC_034981.1). A multifasta file containing the seven previously mentioned RRV RNAs 1-7 was used for mapping. Finally, a SAM file was generated.

2.7.4. BLAST homology against the GenBank

Blastn was used to compare generated contigs against the complete nucleotide (nt) database of GeneBank. A tab-separated file was generated.

2.7.5. Taxonomic identification

Three taxonomic identifiers BlobTools (Laetsch & Blaxter, 2017), MEGAN v.6. (Huson et.al., 2007), and Kraken (Wood & Salzberg, 2014) were compared using obtained results of EDNA-Rose MiFi™. All of the taxonomical identifiers were designed for metagenomics data and long reads generated by Oxford Nanopore.

3. Results

3.1. Primer design

Eight sets of primers, designed on the genomic RNA segment of the RNA dependent RNA polymerase (RdRp), and the capsid protein genes were designed for Multiplex RT-qPCR-HRM. Table 4.2 describes the thermodynamic features of all the primers sets. Analysis of primer sets *in silico* showed they have a 97–100% identity and a query coverage of 100% to the target using BLAST. The average mean E-value was 3e-17. No non-specific matches were detected with other viruses or the rose genome.

Table 4. 2 List of a resume of thermodynamic features of the designed primers for rose infecting viruses.

Targeted species	Gene	PRIMER ID	SEQUENCE	Tm	GC%	Amplicon (bp)	MELTING PICK AT	δg (Kcal/mol)	
Impatiens necrotic spot virus	RdRp	INSV-F1	GACTTCATTTGGGCATCCT	59.932	47.368			FORWARD	REVERSE
		INSV-R1	CTGCACTGAATCGTCACATA	59.876	45	97	75	0.9	0.8
Tomato spotted wilt virus	RdRp	TSWV-F1	GTTATGAACCTCGGGAAAGG	60.066	50				
		TSWV-R1	TACAGACCCGGTGACATT	60.105	50	99	79.25	0.7	1
Tomato ring spot virus	RdRp	ToRSV-F1	TGTTCCGAAGGATGAGAGA	59.89	47.368				
		ToRSV-R1	TACGCAGCAGCAAGTTATAG	59.931	45	101	90.5	0.2	0
Blackberry chlorotic ringspot virus	RdRp	BCRV-F1	CGATCGGTCTCTTGAGTTATG	59.849	47.619				
		BCRV-R1	GCGTCTTTGTCTGCTCTT	59.849	50	99	78	1	1
Rose rosette virus	RdRp	RRV-F1	CCTCTCAGTGGGTGTGATATT	61.179	47.619				
		RRV-R1	CGAACTGATTACGGTGCATT AG	61.211	45.455	108	75.25	0.9	1
Tobacco streak virus	RdRp	TSV-F1	GTGGCGAACAGGATGAAA	59.938	50				
		TSV-R1	CGAGAGTCAAAGTGAGGAAC	59.85	50	107	77	0	0.9
Prunus necrotic spot virus	RdRp	PNRSV-F1	CACTCAGATTTCCACCGAAT AC	60.741	45.455				
		PNRSV-R1	GACCCATCTCGGATCCTATA AA	60.839	45.455	131	79	0.9	1
Apple mosaic virus	RdRp	ApMV-F1	CCGATTCCTTCGGACTTTAC	60.009	50				
		ApMV-R1	GGTGTCTTCACGATCTTTCC	60.229	50	137	83	0	0.9

3.2. Multiplex RT-qPCR –HRM analysis

Multiplex RT-qPCR –HRM is used as the reference method to standardize pathogen detection method, therefore this detection tool was used to validate EDNA MiFi^{TM} specificity detection. Viral pathogen detection with multiplex RT-qPCR combined with High-Resolution Melting (HRM) allows accurate multiple pathogen detection. Three multiplex reactions RT-qPCR-HRM showed high specificity at detecting the viral target sequence. All three multiplex reactions detected 1ng to 1fg of the virus in rose leave samples. This validation method allowed fast screening for detection of multiple infections in a single sample. The multiplex RT-qPCR detects three and two viruses simustaneously as shown in Figure 4.2.

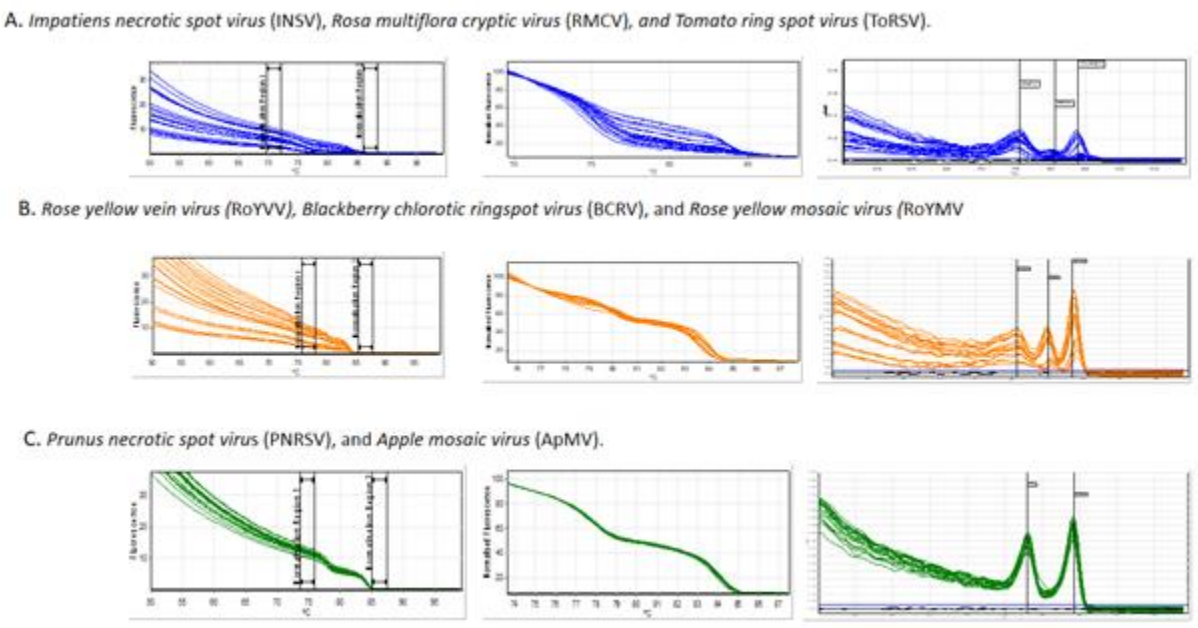


Figure 4. 2 Analysis showing the non-normalize and normalized loss of fluorescence graphs (left and center), and multiplex melting curve (right) of common viral pathogens infecting roses.

Sensitivity results from 1ng to 1 fg are showed for multiplex assays: A. INSV, RMCV, and ToRSV; B. RoYVV, BCRV, and RoYMV. C. PNRSV and ApMV the left column of each row. High Resolution Melting profiles from the positive controls are in the central column of each row.

Normalized fluorescence data derived from raw data plots showing the TM breaking point of fluorescence of the viruses at the Tm. the right column of each row. The low-resolution melt profile derivative plot (-dF/dT against T), shows the steepest slope is easily visualized as the melting peak.

In order to confirm the virus presence in ds cDNA, a qPCR was amplified in the samples before the library preparation. The qPCR detected low viral titter down to femtograms, as shown in Table 4.3 and Figure 4.3.

Table 4. 3 Summarized plant virus concentration determined by qPCR in rose varieties ds cDNA.

VARIETY	**Barcode**	**INSV**	**TSWV**	**TSV**	**PNRSV**
Kiss me, Rose	RLB07		1.00E-06	1.00E-07	
Pink surprise	RLB02	1.00E-06			1.00E-06
5-13 hybrid	RLB11			1.00E-06	
Como Park	RLB04			1.00E-06	
Champlain	RLB05			1.00E-06	
Caroline Hunt	RLB06			1.00E-06	
Dulchen	RLB01			1.00E-06	
5-21 hybrid	RLB08			1.00E-07	
Lemon Splash	RLB09	1.00E-06		1.00E-06	
Top Gun	RLB10				
Rosa Seitigera	RLB03				
Apricot drift	RLB12		1.00E-06	1.00E-06	

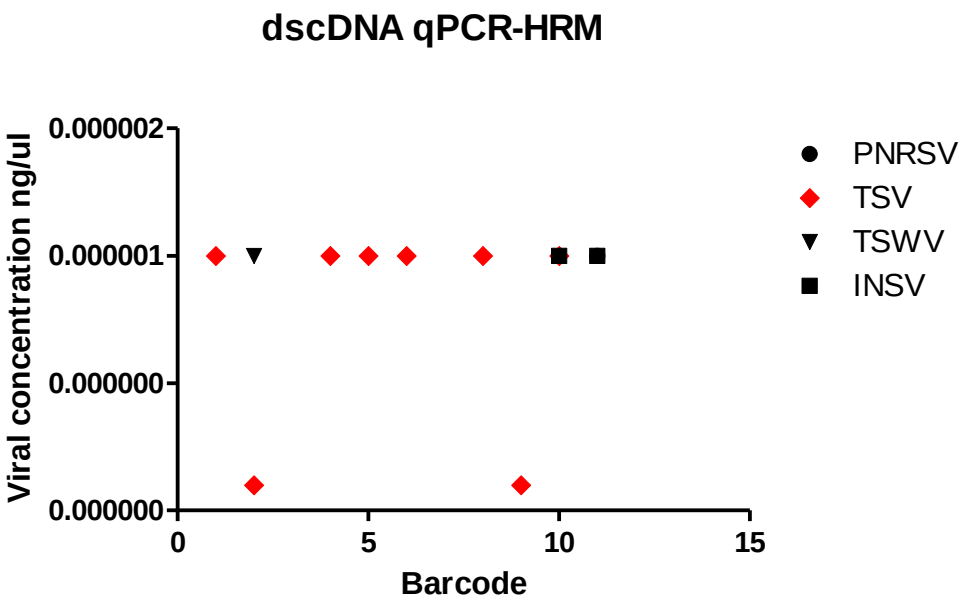


Figure 4. 3 Virus titer accumulated in the selected rose cultivars. The box at the right represents the positive viruses at ds cDNA qPCR- HRM test.

3.3. RNA sequencing analysis

The sequence results of twelve garden rose cultivars showed multiple viral infections by qPCR screening. Sequencing performed in the MinION flow cell (FLO-Min106 R9.4; Oxford Nanopore Technologies, UK) was stopped after 48h. Raw reads retrieved after the demultiplexing step using Albacore script (Oxford Nanopore Technologies, UK) generated the working metadata (.fastq files). The RNA sequenced from 12 barcoded rose varieties yielded 1,593,655 reads of reads per sequencing run (Table 4.4). The average length of reads is ~ 2,801 nt. A total of 409, 560 reads were unclassified due to a lack of attachment with the barcode adaptor. Viral detection was possible, each barcoded variety showed the expected virus target.

Table 4. 4 A summarized report of metagenomes generated by MinIONTM. Sequencing output from rose varieties and E-probe tested for each barcode

ROSE VARIETY	DESCRIPTION	METAGENOME ID	Read Count	SL (nt)	Probe Length (nt)
Kiss me Rose	Susceptible	Rose_barcode07	318	241-1814	20 - 30
Pink surprise	Susceptible	Rose_barcode02	172,412	170-6821	20 - 30
5-13 hybrid	Susceptible	Rose_barcode11	77,433	67-4293	20 - 30
Como Park	Susceptible	Rose_barcode04	76,675	83-4052	20 - 30
Champlain	Susceptible	Rose_barcode05	108,798	149-4548	20 - 30
Caroline Hunt	Susceptible	Rose_barcode06	62,564	136-4710	20 - 30
Dulchen	Susceptible	Rose_barcode01	142,194	164-5315	20 - 30
5-21 hybrid	Susceptible	Rose_barcode08	197,597	169-4176	20 - 30
Lemon Splash	Susceptible	Rose_barcode09	296,115	136-5519	20 - 30
Top Gun	Susceptible	Rose_barcode10	45,988	169-3125	20 - 30
Rosa Seitigera	Resistant	Rose_barcode03	183,861	134-5435	20 - 30
Apricot drift	Susceptible	Rose_barcode12	229,700	159-4896	20 - 30

SL, sequence length

3.4. Complete plant virus detection and hit frequency of rose cultivars

The analyzed metagenomes showed single and multiple viral infections. Cultivars *Kiss me Rose*, *Dulchen*, and *Apricot drift* presented single virus infections (Table 4.5). All other nine rose varieties presented a mixed viral infection (Table 4.5). Semi-quantitative results are determined by the number of hits of the E-probes in the tested cultivars (Table 4.5). *Lemon Splash* was found to harbor a co-infection of 5 viruses is, with a higher concentration of *Rose Rosette Virus* (RRV). The rest of the co-infecting virus are *Tomato yellow ring virus* (TYRV), *Arabis mosaic virus* (ArMV), *Blackberry chlorotic ringspot virus* (BCRV), and *Tobacco ringspot virus* (TRSV). The result suggests *Lemon Splash* is the more susceptible variety in the Rose Rosette Disease Field Trial.

Table 4. 5 List of analyzed metagenomes in EDNA-Rose MiFiTM results of sequenced rose varieties analyzed. A list of tested viruses acronym is represented at the top row of the table. The number of hits is reported in a red scale (pink to intense red) based on lower to higher virus titer.

VARIETY	ACLSV*	ApMV	ASGV*	ArMV	BCRV	INSV	IYSV	PDV	PNRSV	RpRSV*	RMCV	RoLCuV	RLRaV	RRV	RSDaV	RoYMV*	RoYLV	RoYVV*	SLRSV*	TRSV*	TSV	ToRSV*	TSWV	TYRV
Kiss me Rose														0										26
Pink surprise														64										2
5-13 hybrid														47								18	15	26
Como Park					12								1	48										27
Champlain														37							17			29
Caroline Hunt														41							17			
Dulchen														110										
5-21 hybrid														132										28
Lemon Splash				13	15									409						13				27
Top Gun														27									11	15
Rosa Seitigera								3						159					25					14
Apricot drift														2										

* VIRUSES THAT HAVE A POLY A TAIL

As expected, the MiFi™ Rose E-probe database allowed a complete plant virus detection in susceptible and resistant rose varieties. Moreover, a significant hit frequency of E-probes against viral targets in rose varieties was found associated directly to the viral concentration in the sample (Figure 4.4)

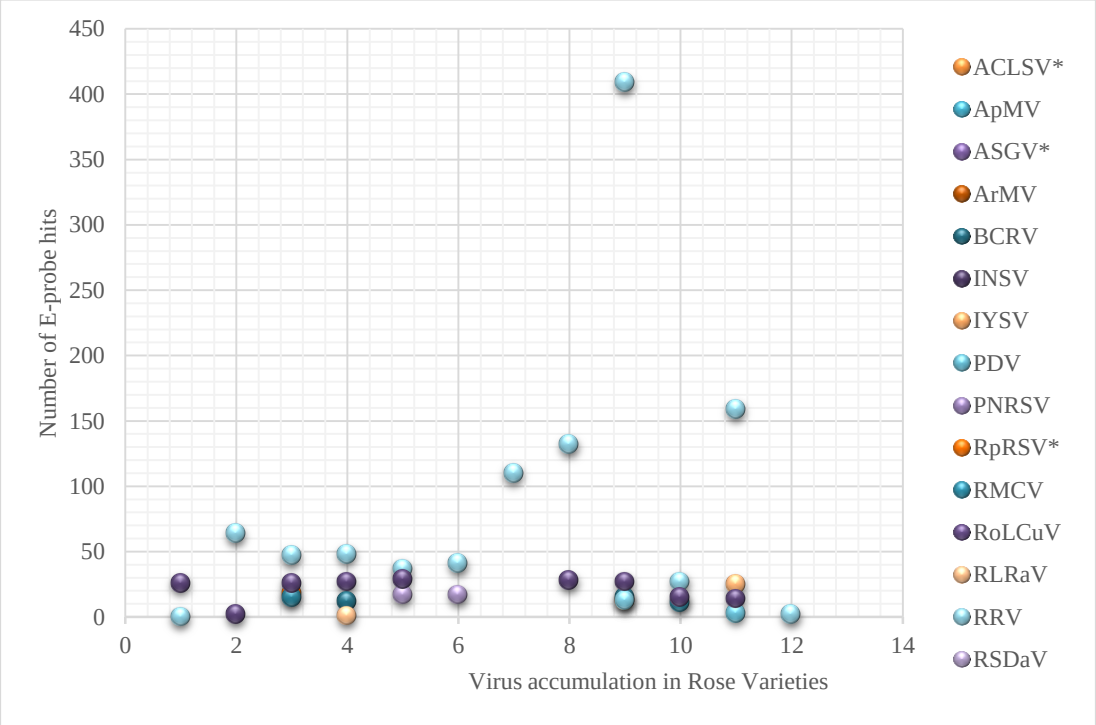


Figure 4. 4 Virus detaction of rose viral metagenomic reads hit the frequency of rose varieties of the Rose Rosette Disease Trial tested against the MiFi™ Rose E-probe database. Metagenomic read hit frequency equates with a viral concentration in the sample. The X-axis represents how the virus accumulated in the host by means of number of hits. The Y-axis correlates the viral accumulation with the number of E-probes that hit a raw metagenomics read. The distribution of color dots represents the tested virus in the selected rose variety.

EDNA-Rose MiFi™, quantified the number of hits per each virus E-probe dataset. *Lemon splash* cultivar is an example of hit frequency for RRV accumulation. In Figure 4.5, the hit frequency accumulation was represented. For a hit be considered positive, the threshold of percent of identity

was calculated over 90%. The graph shows less variance in hits over the selected threshold with complete alignment length at 20nt matching the E-probe length, those considered positive for RRV. Although, there were multiple hits under the threshold (<80 percent of identity), these were not considered for the final quantification output. The EDNA-Rose MiFi^{TM} pipeline considers the thresholds mentioned above for all rose virome quantification when queried the metagenomes against the E-probe sets.

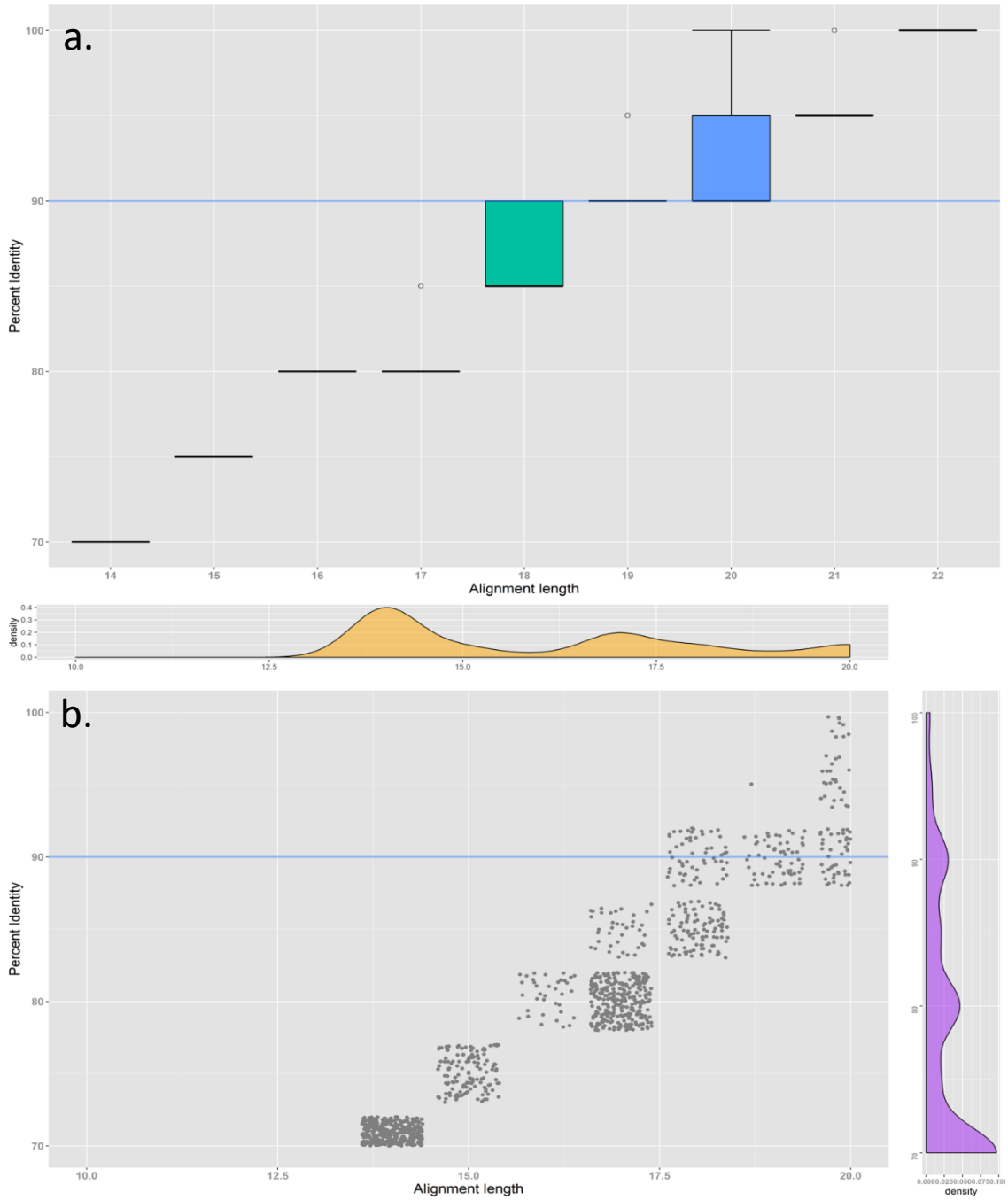


Figure 4. 5 EDNA-Rose MiFiTM E-probe hit distribution and frequencies for RRV detection in *Lemon Splash* cultivar from metagenomics sequences obtained from a Oxford Nanopore sequencing (MinION^{TM}). The X-axis shows the alignment length while the Y-axis represents the percent of identity. a. Hit frequencies of the RRV E-probe set in the sequencing library of the

Lemon Splash dataset. Positive hits are reported only over 90% identity. b. the complete distribution of RRV hits, gray dots represent the aligned hits, the positive hits are over the threshold (>90% identity).

3.5. Standardized (Multiplex RT-qPCR –HRM Analysis) correlation with EDNA (MiFi™) detection analysis

Multiplex RT-qPCR –HRM results were comparable and correlated with High-throughput sequencing and EDNA-Rose MiFi™, which reflected the robustness of the proposed screening method.

The analysis by EDNA-Rose MiFi™ revealed multiple viral infections that correlates with results obtained by Multiplex RT-qPCR –HRM. Not surprisingly, the use of HTS combined with a curated database detection (EDNA-Rose MiFi™) allows rapid multiplex plant viral detection.

3.6. EPI2ME results

EPI2ME workflow through What is in my pot? (WIMP) revealed few reads per barcode corresponding to plant infecting viruses. Only two rose varieties tested positive for RRV. *Lemon splash* and *Apricot drift* resulted in 6 and 5 reads identified as RRV (Table 4.6). Other plant viruses were detected, however, none of them were previously reported in rose cultivars (*Dasheen mosaic virus, Maize dwarf mosaic virus, Garlic common latent virus, Allium virus X, Pepper chlorotic spot virus,* and *Colombian potato soil-borne virus*). *Kiss me Rose, 5-13 hybrid, Como Park, Champlain, Caroline Hunt, 5-21 hybrid,* and *Top Gun* did not present a plant virus presence.

Table 4. 6 EPI2ME Results related to plant viruses per barcode analyzed.

County	Sample date	Sample ID	Barcode	Read per barcode	Genus	Significant Viral Results	# matching reads
Tulsa	Summer2018	*Kiss me Rose*	RLB07	18	No genus identified	No plant virus identified	0
						Dasheen mosaic virus	1
					Potyvirus	*Maize dwarf mosaic virus*	1
Tulsa	Summer2018	*Pink surprise*	RLB02	56557	*Carlavirus*	*Garlic common latent virus*	1
Tulsa	Summer2018	*5-13 hybrid*	RLB11	2002	No genus identified	No plant virus identified	0
Tulsa	Summer2018	*Como Park*	RLB04	5913	No genus identified	No plant virus identified	0
Tulsa	Summer2018	*Champlain*	RLB05	3888	No genus identified	No plant virus identified	0
Tulsa	Summer2018	*Caroline Hunt*	RLB06	9783	No genus identified	No plant virus identified	0
Tulsa	Summer2018	*Dulchen*	RLB01	57960	*Carlavirus*	*Garlic common latent virus*	1
Tulsa	Summer2018	*5-21 hybrid*	RLB08	23748	No genus identified	No plant virus identified	0
					Emaravirus	*Rose rosette virus*	6
Tulsa	Summer2018	*Lemon Splash*	RLB09	20468	*Potexvirus*	*Allium virus X*	1
Perkins	Summer2018	*Top Gun*	RLB10	1458	No genus identified	No plant virus identified	0
					Othotospovirus	*Pepper chlorotic spot virus*	1
Perkins	Summer2018	*Rosa Seitigera*	RLB03	43323	*Pomovirus*	*Colombian potato soil-borne virus*	1
Perkins	Summer2018	*Apricot drift*	RLB12	9058	*Emaravirus*	*Rose rosette virus*	5

3.7. Data analysis using traditional virus identification and discovery bioinformatics

The downstream analysis of the FastQC results showed a low Phred score for all the barcodes, the FastQC software was designed for short reads. Phred scored reported from FastQC were designed for Illuimina® short reads, and in this study were used to visualize the reads that did not contain adapters on the sequences and are ready for further analysis. The mean Phred score was 17 ranging in the low-quality portion on pink (Figure 4.6 a). Sequence length distribution ranged ~0.1-5.3 Kb (Figure 4.6 b). All the barcodes follow the same quality score and read length distribution pattern. %GC had a mean value of 43 for all the analyzed barcodes.

a. b.

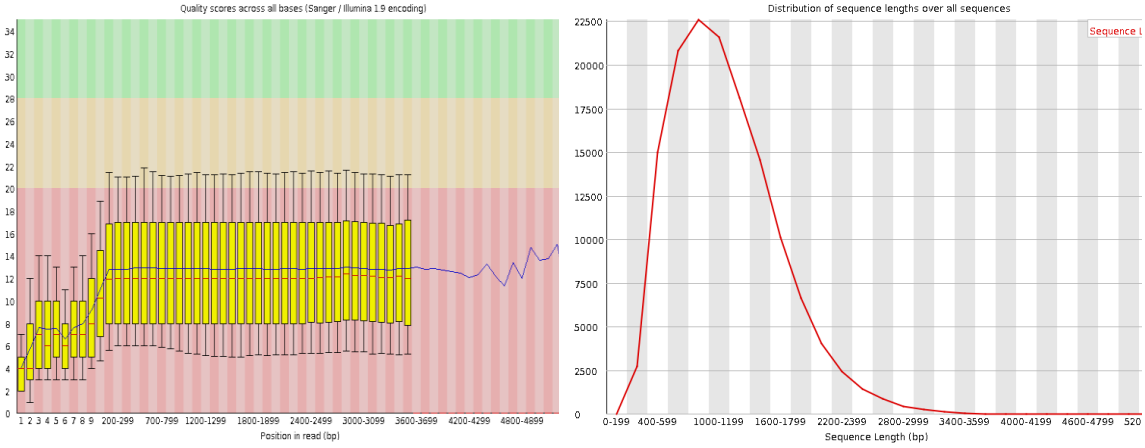


Figure 4. 6 a. Phred score from barcode1 representing the low quality of the reads. b. Read length distribution from barcode1 (*Kiss me Rose* variety).

The best assembler for rose metagenomics data was MEGAHIT, in comparison with SPADES and CANU. MEGAHIT generated contigs analysis with QUAST, it gave higher quality and quantity (Table 4.7). MEGAHIT provided better assembly contiguity. Overall, results on MEGAHIT had the best quality besides barcode07, corresponding to the variety *Kiss me Rose*. The N50 was between 517 to 676 for all rose cultivars. The contigs generated have an optimal length (length ≥1000bp) (Figure 4.6) and there is a correlation for %GC previously found in

FastQC 43% of the mean value. No mismatches were produced after the assembly (Table 4.7). Table 4.8 represents a summary of a comparison of the used assemblers' results per barcode.

Table 4. 7 QUAST Results for MEGAHIT assembler

All statistics are based on contigs of size >= 100 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs).

Worst Median Best ☑ Show heatmap

Statistics without reference	final.contigs1	final.contigs10	final.contigs11	final.contigs12a	final.contigs2	final.contigs3	final.contigs4	final.contigs5	final.contigs6	final.contigs7	final.contigs8	final.contigs9
# contigs	46 391	17 520	21 146	58 342	41 143	49 080	21 658	27 107	16 647	95	46 623	60 501
# contigs (>= 0 bp)	46 391	17 520	21 146	58 342	41 143	49 080	21 658	27 107	16 647	95	46 623	60 501
# contigs (>= 1000 bp)	5741	950	805	3913	1727	4337	1234	890	588	4	2758	2443
# contigs (>= 5000 bp)	0	0	0	0	0	0	0	0	0	0	0	0
# contigs (>= 10000 bp)	0	0	0	0	0	0	0	0	0	0	0	0
# contigs (>= 25000 bp)	0	0	0	0	0	0	0	0	0	0	0	0
# contigs (>= 50000 bp)	0	0	0	0	0	0	0	0	0	0	0	0
Largest contig	3713	2787	2951	3423	2987	4281	3084	2783	2434	1641	3504	2755
Total length	28 677 208	9 831 287	10 906 796	32 724 180	21 943 515	28 630 296	11 869 713	13 883 641	8 545 125	55 212	25 858 192	31 783 658
Total length (>= 0 bp)	28 677 208	9 831 287	10 906 796	32 724 180	21 943 515	28 630 296	11 869 713	13 883 641	8 545 125	55 212	25 858 192	31 783 658
Total length (>= 1000 bp)	7 659 690	1 177 594	998 217	4 871 416	2 086 247	5 669 634	1 534 936	1 094 159	737 725	4917	3 390 338	2 957 924
Total length (>= 5000 bp)	0	0	0	0	0	0	0	0	0	0	0	0
Total length (>= 10000 bp)	0	0	0	0	0	0	0	0	0	0	0	0
Total length (>= 25000 bp)	0	0	0	0	0	0	0	0	0	0	0	0
Total length (>= 50000 bp)	0	0	0	0	0	0	0	0	0	0	0	0
N50	676	593	521	595	554	620	567	519	517	589	585	542
N75	456	441	406	429	422	441	428	406	406	458	429	413
L50	13 953	5997	7438	19 251	14 432	15 607	7373	9641	5884	34	15 635	21 162
L75	26 928	10 798	13 391	35 504	25 770	29 368	13 426	17 220	10 567	61	28 586	37 981
GC (%)	43.1	42.85	43.22	43.22	42.77	43.44	43.09	43.06	43.17	42.69	42.54	43.07
Mismatches												
# N's	0	0	0	0	0	0	0	0	0	0	0	0
# N's per 100 kbp	0	0	0	0	0	0	0	0	0	0	0	0

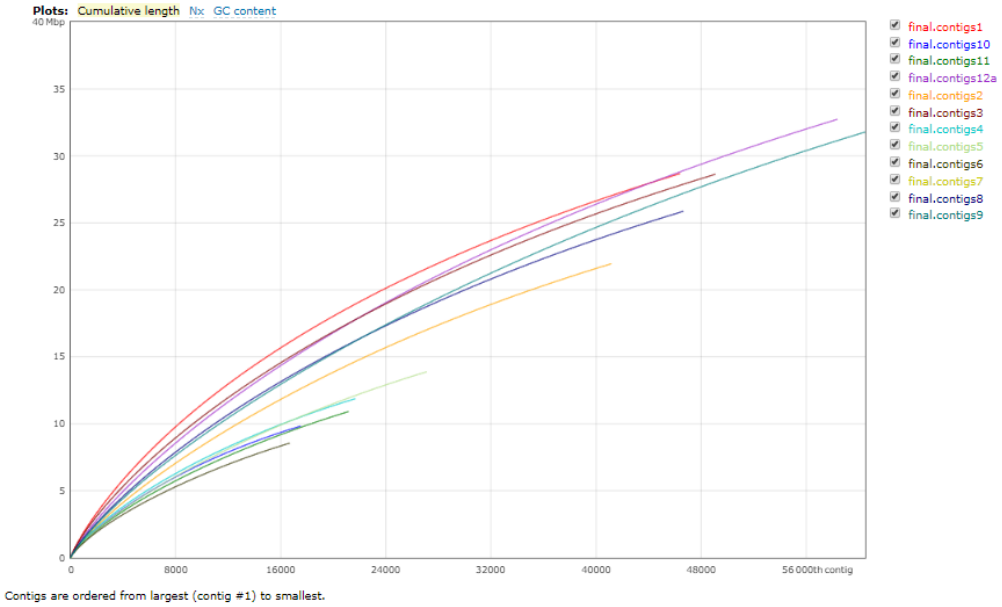


Figure 4. 7 Plot representation of the contig length per barcode and GC content.

The largest contigs from *Lemon Splash* cultivar are represented on turquoise, and the smallest from *Kiss me Rose cultivar* is light green.

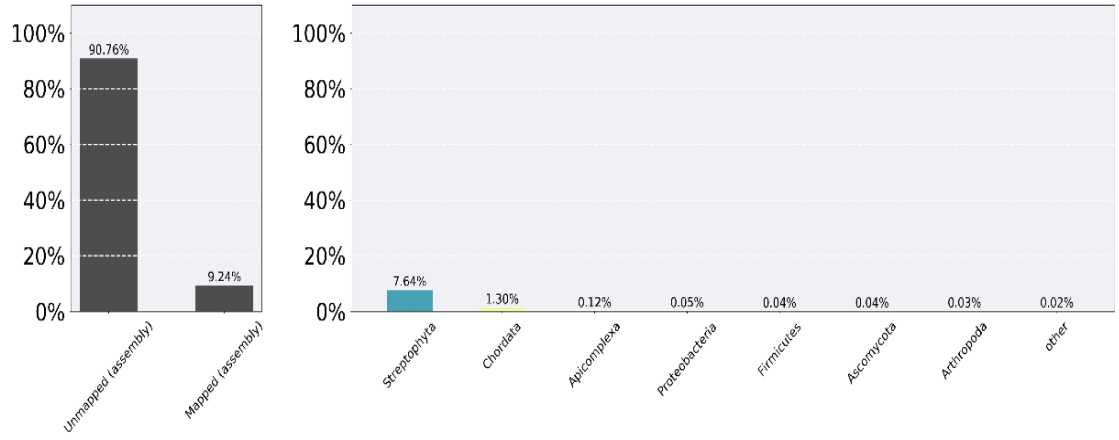
Table 4. 8 Analysis of assembled contigs obtained with the assemblers MEGAHIT, SPADES, and CANU for each garden rose cultivars.

Sample	Barcode	MEGAHIT	SPADES	CANU
Kiss me Rose	RLB07	4 Contigs	2 Contigs - 0 Scaffolds	No Contigs - No Scaffolds
Pink surprise	RLB02	1727 Contigs	3 Contigs - 3 Scaffolds	
5-13 hybrid	RLB11	805 Contigs	1 Contig - 1 Scaffold	
Como Park	RLB04	1234 Contigs	1 Contig - 1 Scaffold	
Champlain	RLB05	890 Contigs	12 Contigs - 12 Scaffolds	
Caroline Hunt	RLB06	588 Contigs	3 Contigs - 3 Scaffolds	
Dulchen	RLB01	5741 Contigs	5 Contigs - 5 Scaffolds	
5-21 hybrid	RLB08	2758 Contigs	627 Contigs - 627 Scaffolds	
Lemon Splash	RLB09	2443 Contigs	5 Contigs - 5 Scaffolds	
Top Gun	RLB10	950 Contigs	2 Contigs - 2 Scaffolds	
Rosa Seitigera	RLB03	4337 Contigs	594 Contigs -595 Scaffolds	
Apricot drift	RLB12	3913 Contigs	5 Contigs - 5 Scaffolds	

The inclusivity estimation of the assembly alignment as a post-quality control of the contigs allowed to estimate the accuracy of the assemblies. Raw reads were mapped back to the generated contigs and mapped against the reference RRV genome. The final mapped reads generated an output, .sam files (Bowtie 2), that were plotted in BlobTools.

Taxonomic identification is based on the number of positive matches with the BLASTn database. Based on Phylum Distribution presented in Figure 4.8, it depicts the read coverage distribution, and, as expected, most of the mapped reads belong to Phylum Streptophyta. BlobTools showed significant matches with the *Rosa chinensis* (host) genome (Figure 4.9). A virus cluster appears in Figure 4.10. These results suggest the presence of plant viruses are in low titer with respect to the host genome. Virus assemblies are positioned on the X-axis based on the length of the contig, confirmed low coverage for virus taxonomy.

a.



b.

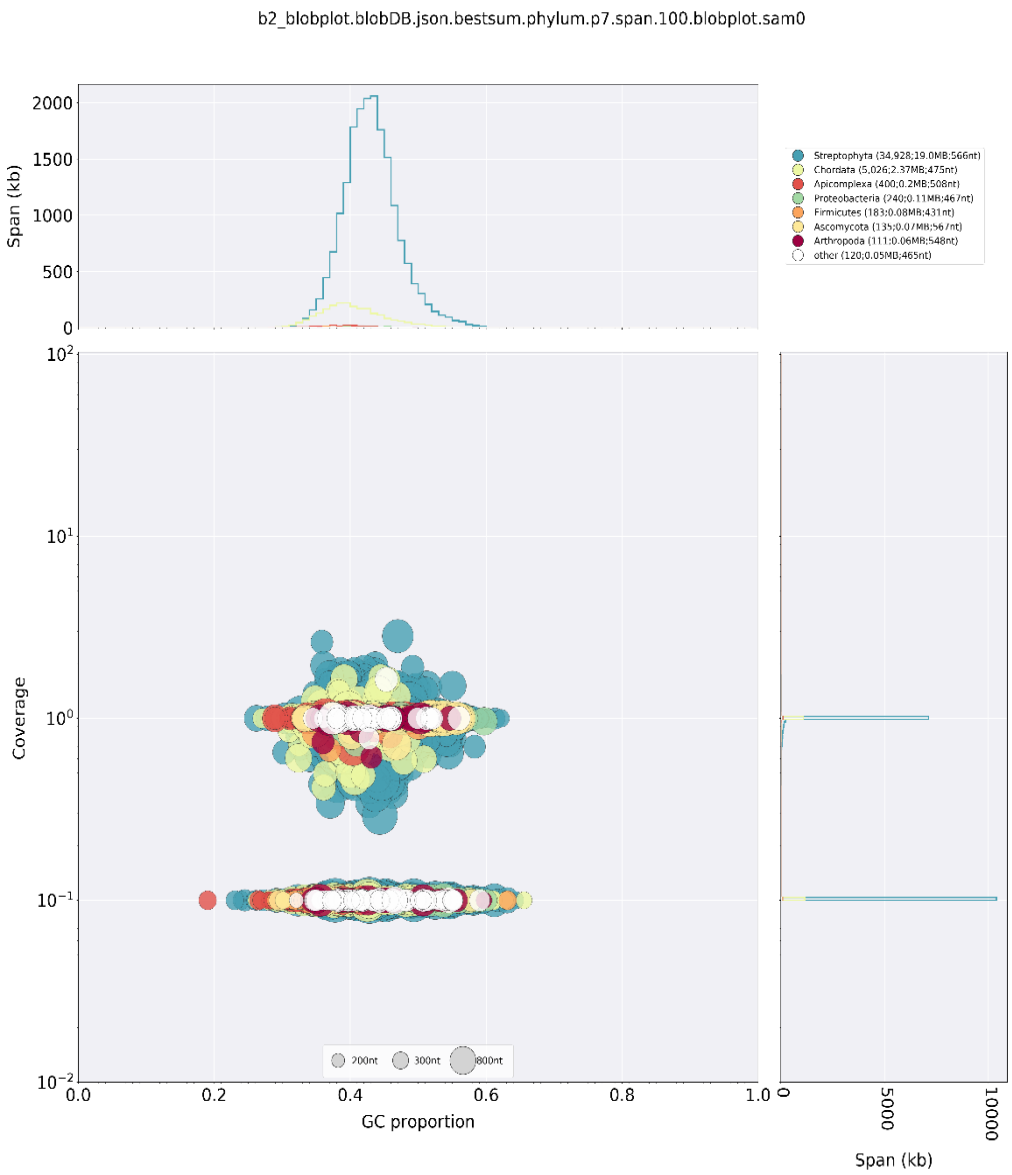


Figure 4. 8 a. Read distribution percentage mapped to the assembly. b. Coverage Plot of the contigs belonging to the Phylum taxonomical identification obtained from cultivar *Pink surprise*. Sequences in the assembly are depicted as colored circles, its diameter is in scale proportional to the length of the contigs. On the X-axis GC proportion and in the Y-axis genome coverage.

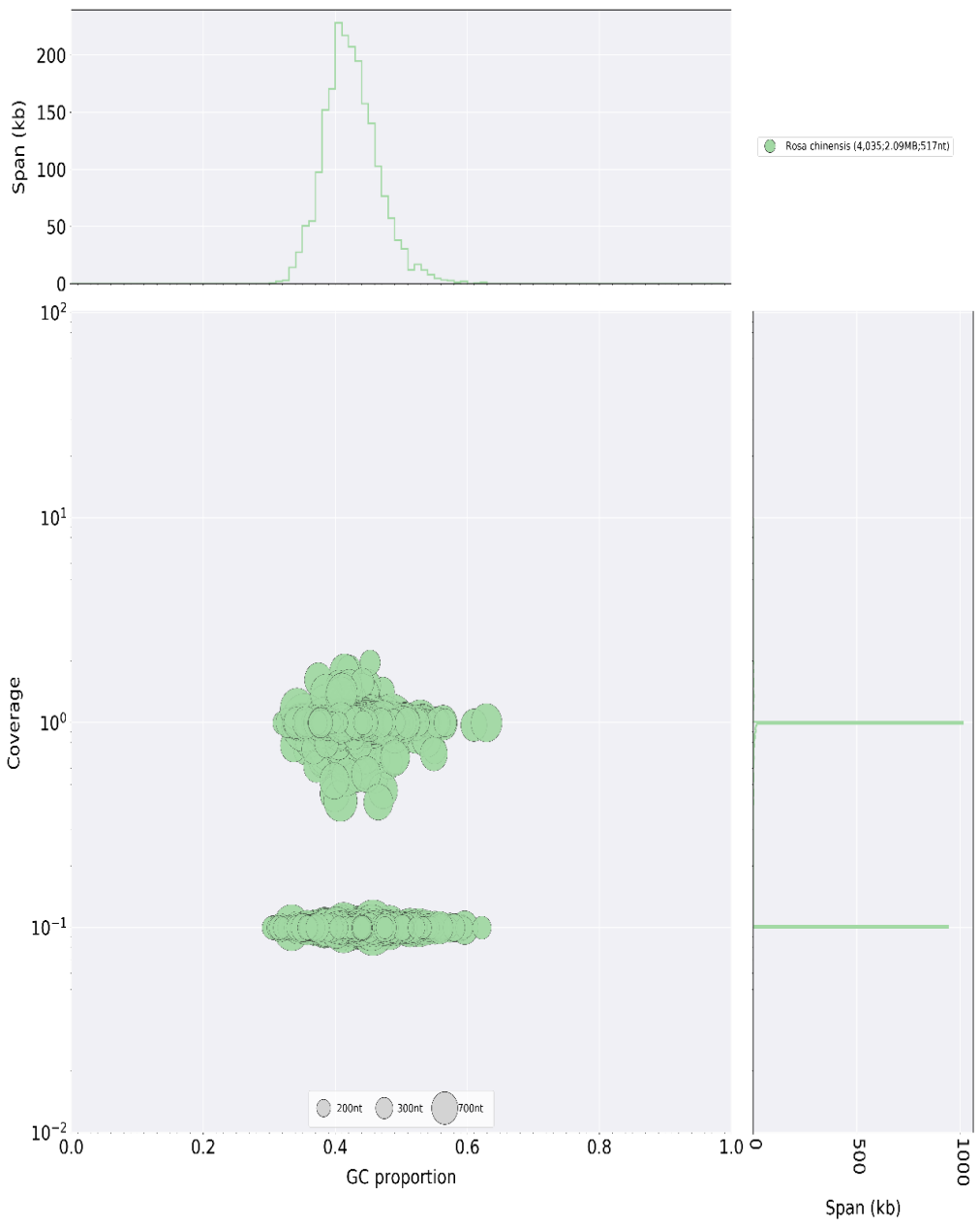


Figure 4. 9 Coverage Plot of the host (rose) contigs. Sequences in the assembly are depicted as green circles, its diameter is in scale proportional to the length of the contigs. On the X-axis GC proportion and in the Y-axis genome coverage.

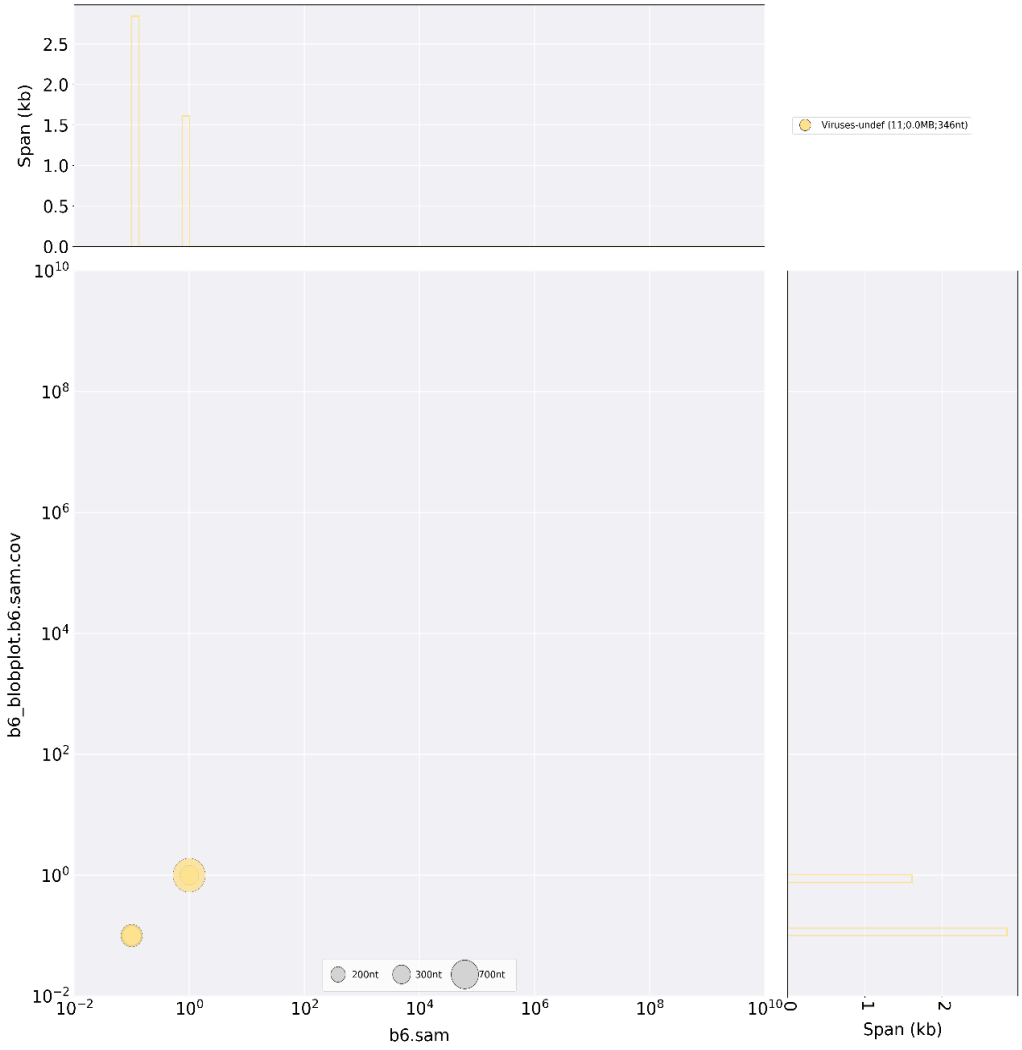


Figure 4. 10 Coverage Plot of contigs belonging to viruses present in the cultivar *Caroline Hunt* (barcode 06). Sequences in the assembly are depicted as yellow circles, its diameter is in scale proportional to the length of the contigs. On the X-axis GC proportion an d in the Y-axis genome coverage.

BlobTools taxonomically identified few contigs belonging to plant viruses as showed in Table 4.9. Only *Lemon splash* and *Rosa seitigera* cultivars showed positive matches for Rose Rosette Emaravirus with 26 and 17 matches respectively.

Table 4. 9 BlobTools summarized results of positive hits belonging to plant viruses present in the tested rose varieties.

Sample	Barcode	Plant Virus positive	# of matches
Kiss me, Rose	RLB07	No plant virus found	0
		Beet chlorotic virus	1
		Cherry necrotic rusty mottle virus	2
		Turnip mosaic virus	1
Pink surprise	RLB02	*Apple stem grooving virus*	1
		Chilli veinal mottle virus	1
5-13 hybrid	RLB11	*Soybean mosaic virus*	1
		Citrus tristeza virus	1
		Sugarcane mosaic virus	1
Como Park	RLB04	*Zucchini yellow mosaic virus*	1
		East african cassava mosaic virus	1
		Turnip mosaic virus	1
		Potato virus Y	1
		Watermelon mosaic virus	1
Champlain	RLB05	*Sweet potato feathery mottle virus*	1
		Cassava brown streak virus	1
		Tomato spotted wilt virus	1
		Grapevine fanleaf virus	1
		Cotton leaf curl multan virus	1
		Grapevine leafroll-associated virus	3
Caroline Hunt	RLB06	*Tomato chlorosis virus*	1
		Papaya ringspot virus	1
Dulchen	RLB01	*Banana bunchy top virus*	1
		Grapevine pinot gris virus	1
		Banana bunchy top virus	1
		Potato virus Y	1
5-21 hybrid	RLB08	*Blackberry yellow vein-associated virus*	1
		Bean common mosaic virus	1
Lemon Splash	RLB09	*Fig mosaic virus*	1

		Tobacco necrosis virus	1
		Turnip mosaic virus	1
		Rose rosette emaravirus	26
		Wheat dwarf virus	1
		Tomato spotted wilt virus	1
Top Gun	RLB10	No plant virus found	0
Rosa Seitigera	RLB03	No plant virus found	0
		Cherry virus A	1
		Tomato spotted wilt virus	3
		Cucumber green mottle mosaic virus	1
		Sugarcane mosaic virus	1
		Fig mosaic virus	1
		Rose rosette emaravirus	17
Apricot drift	RLB12	*Banana bunchy top virus*	1

Kraken detected positive RRV presence in *Lemon splash* and *Apricot drift* rose varieties, as summarized in Table 4.10. None of the other sequenced rose varieties showed plant virus infection when compared with the Kraken detection pipeline.

Table 4. 10 Summarized Kraken taxonomically identification for tested rose varieties.

Sample	**Barcode**	**Kraken**
Kiss me, Rose	RLB07	No plant virus identified
Pink surprise	RLB02	No plant virus identified
5-13 hybrid	RLB11	No plant virus identified
Como Park	RLB04	No plant virus identified
Champlain	RLB05	No plant virus identified
Caroline Hunt	RLB06	No plant virus identified
Dulchen	RLB01	No plant virus identified
5-21 hybrid	RLB08	No plant virus identified
Lemon Splash	RLB09	*Rose rosette virus*
Top Gun	RLB10	No plant virus identified
Rosa Seitigera	RLB03	No plant virus identified
Apricot drift	RLB12	*Rose rosette virus*

3.8. RRV quantification among garden rose cultivars by EDNA-Rose MiFi™

EDNA-Rose MiFi™ semi-quantifed the number of positive hits (Figure 4.11). Only the *Kiss me Rose* cultivar did not show positive hits against the RRV E-probe set. This is due to the number of reads that were produced after sequencing (18 reads) which belong to the host. The cultivar with the highest number of hits was *Lemon Splash* with 409 hits, followed by *Rosa seitigera* with 159 hits. Cultivar *Apricot Drift* showed the lower RRV titer with 2 hits. All virus concentrations are shown in Table 4.5, column 15 corresponding to RRV.

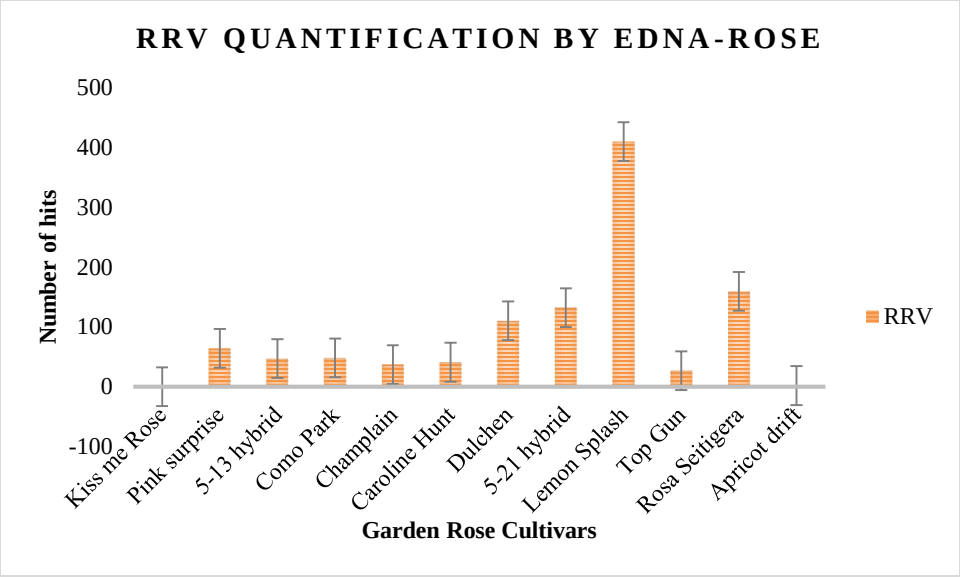


Figure 4. 11 Rose viral metagenomic reads hit the frequency plot for *Rose Rosette Virus* semi-quantification tested against the MiFi™ Rose E-probe database. Metagenomic read hit frequency equates with a viral concentration in the sample. The X-axis represents how the virus accumulated in the host by means of concentration. The Y-axis correlates the viral accumulation with the number of E-probes that hit a raw metagenomics reads.

4. Discussion

Detection of virus diseases and discrimination of viruses co-infecting roses constitute a major threat. Reliable pathogen diagnostics is essential for disease control strategies. Therefore, early pathogen detection brings benefits for viral disease management. Growers scout for reported symptoms before submitting samples to plant diagnostic clinics. Since most viral symptoms are visually similar to each other the diagnosis may be inconclusive. In this study, results demonstrated EDNA-Rose MiFi™ can be used for accurate detection of the reported virome infecting roses (22 in total). EDNA previously was proposed as a pathogen detection tool (Stobbe, et.al., 2013). Following the application of EDNA fundamental concept, this research aims to present a new way plant virus detection using HTS. Previous attempts portrayed EDNA as a tool designed for single pathogen detection (Espindola, et.al.,2016; Blagden, et.al., 2016). In this study, EDNA is presented

as a prototype for broad reliable detection screening of germplasm able to successfully detect viruses infecting rose varieties.

The current use of High Throughput Sequencing (HTS) has become extensively applied as a newly DNA-based pathogen detection technology applied to metagenomic samples (Massart, et.al., 2017). This study describes the screening of twelve rose varieties quantifying hit frequencies obtained for each rose variety queried against the E-probes hits. In order to make a detection, the number of hits of the specific E-probe sets was compared to the hits resulting from the decoy E-probes. The significant p-value for the positive results in metagenomes queried against RRV and TYRSV ($p \leq 0.05$) demonstrated that both viruses are present in most of the analyzed rose cultivars.

The common bioinformatic pipeline is a complete cumputer intensive approach for virus detection and discovery. Similar results among taxonomic classifiers and EDNA-Rose MiFi™ were obtained, only cultivar *Lemon Splash* was the one showing high virus titter for *Rose rosette virus*. This is because the virus is in high concentration and this variety shows to be highly susceptible to RRV infection, and collected specimen showed characteristic symptoms of RRV infection. *Rosa seitigera* was the only rose spp. reported resistant to RRV (Table 4.2). However, after EDNA-Rose MiFi™ analysis was the second one with more number of hits, this may be the result of either resistance or a recent event of overcoming resistance. All other cultivars were reported susceptible correlating with the result hit frequency of the virus among the cultivars.

Di Bello et al., (2015) demonstrated that Rose Rosette Disease is the result of the presence of only RRV infection in a cultivar. This study showed that there is a synergy of mixed infection in the studied cultivars reflecting the presence of other viruses in addition to RRV. For instance, analysis of the cultivar with a higher RRV titer, *Lemon splash*, which showed evidence of presence of four additional viruses infecting with RRV (409 hits). The hit frequency detected by EDNA-Rose MyDetect™ pipeline of ArMV (13hits), BCRM (15 hits), TRSV (13 hits), and TYRSV (27 hits),

suggests RRV may be masking the symptomatology of other co-infecting viruses and probably lowering their replication rate. RRV and the genus Emaravirus have been determined as a complex evolutionary virus group (DiBello et al, 2015). The co-evolution of RRV with the host was determined by the complex genome plasticity of the Emaravirus genus by reassortment and duplication of pathogenicity proteins (Tatineni at al., 2011) similarly this may lead to overcoming resistance in *Rosa seitigera* spp.

This study demonstrates multiple virus infections the HTS obtained metagenomes. Besides *Kiss me Rose* and *Apricot drift* cultivars, the other 10 metagenomes showed infections with two or more viruses part of the reported rose virome. Susceptible cultivars host more than one virus. Rose infecting viruses showed different accumulation of hits among resistant and susceptible cultivars. *Rosa seitigera* was the reported resistant species against RRV (Byrne et al., 2015), however, EDNA-Rose MiFi™ demonstrated that this variety was infected by four viruses (PDV, RRV, SLRSV, and TYRSV). This rose species accumulated high RRV titer of with 159 hits in Oklahoma fields.

Deepness of the sequencing platform used for viral detection is important (Pecman et.al., 2017). In this research the Oxford Nanopore platform was selected because its field-deployable potential, also because it generates long reads. Long reads allow a full virus genome sequence. The plant virus genomes are small if compared with the host rose genome (Table 4.1). Therefore, if considered the complete rose genome is 88781.76 Mb and the genome of a single virus, for instance, RRV 17.8 Kb to be able to find reads from the virus. There is a no need to use deep sequencing platforms to detect rose virome. Another alternative to find more reads of viral particles within a metagenome is to deplete the host RNA. Noteworthy, total RNA was extracted in this study, therefore most of the mRNA of the host was amplified in a higher ratio than the genome of the targeted virus. Poly A tail termination of some rose virus is found within the analyzed metagenomes. The selected amplification method was ds cDNA before library preparation that will

selectively amplify sequences with a Poly A, terminus. Further research has to use random hexamers in order to equally enrich all virus presence.

Even though traditional molecular diagnostics require previous knowledge of a small region of the target pathogen. EDNA database is built upon specific electronic probes generated from the complete reported pathogen genome sourced by the NCBI GeneBank or alternative repository. It is also demonstrated the usefulness and further applicability of Multiplex RT-qPCR –HRM which was applied in this research as a validation checkpoint before sequencing. This validation step determined the accuracy of EDNA Rose MiFi^{TM} during detection using HTS. The multiplex RT-PCR developed in this study detects and discriminates the most common virus reported in the US.

Although true positive hits were found in EDNA-Rose MiFi^{TM} analysis when compared with Multiplex RT-qPCR –HRM, some of them correlate between both detection methods. However, since the Multiplex RT-qPCR –HRM targets a small region (~90-350bp) some of the expected viruses were not detected by HTS. This is suggesting, a ribosomal depletion method should be applied before sequencing must be executing prior HTS (Kim, et.al., 2012).

Rose is a multispecies complex and is vulnerable to pathogen introduction, even more in cultivated systems. Modern rose cultivars are a product of interspecific hybridization. Most of the garden roses are derived from hybrids with a genetic association of wild rose parentages. Breeding selection and hybridization resulting in cultivars of 2x to 8x chromosomal cultivars focus on disease resistance (Horst & Cloyd, 2007). Achieving high levels of disease protection is cornerstone and the main target in rose breeding programs. Once a viral infection establishes in a plant there is no remediation method available rather than eradication of the infected plant. Breeding programs focus on cloning rose genes with putative functions for disease resistance in order to avoid the infection of the virus to cultivated areas similarly the molecular marker-assisted breeding and host-virus interaction (Debener & Byrne, 2014). The Rose rosette disease trial at Oklahoma State University

in Payne county (Perkins) is one of the nationwide research replicates of newly released rose cultivars with continuous disease monitoring. The sampled cultivars were varieties that are now in the market as well as two hybrids. All of the samples besides *Kiss me Rose* have a high concentration of RRV (Figure 4.11). The hybrids 5-13 and 5-21 showed 47 and 132 hits against RRV showing virus accumulation and less RRV characteristic symptomatology. As expected, the use of HTS as diagnostic tools combined with bioinformatics will benefits and hasten outright reported viral disease screening.

EDNA-Rose MiFi^{TM} provides a new framework for entire plant virus detection. Due to the multiple pathogen detection capabilities, EDNA has strengthened since its creation as a diagnosis method and has the potential to be applied widely at the border by biosecurity systems, quarantine laboratories, greenhouses, and plant diagnostic clinics. Future direction and application of this technology will be used for germplasm evaluation and in this way ensuring rose virus-free cultivars. Even more, in the near future, a database that includes reported pathogens infecting a host will provide a complete and integrated diagnosis. Also, deep sequencing of the vector *Phyllocoptes fructiphilus* may give some answers to the scientific community of the host-virus-vector interaction into the studied pathosystem.

From a biosecurity perspective, the developed technology (EDNA-Rose MiFi^{TM}) has described capabilities for monitoring, control and avoid foreign virus introduction to an economically horticultural crop (Mumford, et.al, 2016). This study aims to accelerate plant diagnosis with cut edge technologies needed for risk assessment and agricultural biosecurity

References

Abdullah, A. S., Moffat, C. S., Lopez-Ruiz, F. J., Gibberd, M. R., Hamblin, J., & Zerihun, A. (2017). Host–multi-pathogen warfare: pathogen interactions in co-infected plants. *Frontiers in plant science*, *8*, 1806.

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., ... & Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular plant pathology*, *10*(4), 537-545.

Amrine J. (1996) Phyllocoptes fructiphilus and biological control of multiflora rose p. 741–749. In: E.E. Lindquist M.W. Sabelis and J. Bruins (eds.). Eriophyid mites—Their biology natural enemies and control. Elsevier Amsterdam. The Netherlands.

Babu, B., Jeyaprakash, A., Jones, D., Schubert, T. S., Baker, C., Washburn, B. K., ... & Paret, M. L. (2016). Development of a rapid, sensitive TaqMan real-time RT-PCR assay for the detection of Rose rosette virus using multiple gene targets. *Journal of virological methods*, *235*, 41-50.

Babu, B., Washburn, B. K., Miller, S. H., Poduch, K., Sarigul, T., Knox, G. W., ... & Paret, M. L. (2017). A rapid assay for detection of Rose rosette virus using reverse transcription-recombinase polymerase amplification using multiple gene targets. *Journal of virological methods*, *240*, 78-84.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... & Pyshkin, A. V. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, *19*(5), 455-477.

Barrero, R. A., Napier, K. R., Cunnington, J., Liefting, L., Keenan, S., Frampton, R. A., ... & Whattam, M. (2017). An internet-based bioinformatics toolkit for plant biosecurity diagnosis and surveillance of viruses and viroids. *BMC bioinformatics*, *18*(1), 26.

Blagden, T., Schneider, W., Melcher, U., Daniels, J., & Fletcher, J. (2016). Adaptation and validation of E-probe diagnostic nucleic acid analysis for detection of Escherichia coli O157: H7 in metagenomic data from complex food matrices. *Journal of food protection*, *79*(4), 574-581.

Byrne, D., Klein, P., Hall, C., Windham, M., Ochoa-Corona, F., Olson, J., & Hammond, J. (2017). Combating Rose rosette disease US national project. In *VII International Symposium on Rose Research and Cultivation 1232* 203-212.

Byrne, D. H., Klein, P., Yan, M., Young, E., Lau, J., Ong, K., & Novick, D. (2018). Challenges of breeding rose rosette–resistant roses. *HortScience*, *53*(5), 604-608.

Converse, R. H., & Bartlett, A. B. (1979). The occurrence of viruses in some wild Rubus and Rosa species in Oregon. *Plant Disease Reporter*, *63*(6), 441-444.

Debener, T., & Byrne, D. H. (2014). Disease resistance breeding in rose: current status and potential of biotechnological tools. *Plant Science*, *228*, 107-117.

Di Bello, P. L., Ho, T., & Tzanetakis, I. E. (2013). Identification of the rose rosette disease agent. In *VI International Symposium on Rose Research and Cultivation 1064* 295-298.

Dobhal, S., Olson, J. D., Arif, M., Suarez, J. A. G., & Ochoa-Corona, F. M. (2016). A simplified strategy for sensitive detection of Rose rosette virus compatible with three RT-PCR chemistries. *Journal of virological methods*, *232*, 47-56.

Dwight, Z., Palais, R., & Wittwer, C. T. (2011). uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics*, *27*(7), 1019-1020.

Epstein A.& Hill J. (1999) Status of rose rosette disease as a biological control for multiflora rose *Plant Dis.*83292101.

Espindola, A., Schneider, W., Hoyt, P. R., Marek, S. M., & Garzon, C. (2015). A new approach for detecting fungal and oomycete plant pathogens in next-generation sequencing metagenome data utilizing electronic probes. *International journal of data mining and bioinformatics*, *12*(2), 115-128.

Fang, Y., & Ramasamy, R. (2015). Current and prospective methods for plant disease detection. *Biosensors*, *5*(3), 537-561.

Fulton, R. W. (1970). Prunus necrotic ringspot virus. CMI/AAB Descript. *Plant Viruses No. 5*.

Golino DA, Sim ST, Cunningham M, Rowhani A (2011) Evidence of root graft transmission of two rose mosaic viruses, Prunus necrotic ringspot virus and Apple mosaic virus in rose rootstocks. Phytopathology 101: S62

Gómez, P., Sempere, R. N., Elena, S. F., & Aranda, M. A. (2009). Mixed infections of Pepino mosaic virus strains modulate the evolutionary dynamics of this emergent virus. *Journal of Virology*, *83*(23), 12378-12387.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: a quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072-1075.

Hall, T., Biosciences, I., & Carlsbad, C. (2011). BioEdit: an important software for molecular biology. *GERF Bull Biosci*, *2*(1), 60-61.

He, Y., Yang, Z., Hong, N., Wang, G., Ning, G., & Xu, W. (2015). Deep sequencing reveals a novel closterovirus associated with wild rose leaf rosette disease. *Molecular plant pathology*, *16*(5), 449-458.

Horst, K. (2007). Compendium of Rose Diseases. APS Press, St. Paul, MN. pp 83.

Horst, R. K., & Cloyd, R. A. (2007). *Compendium of rose diseases and pests*. American Phytopathological Society (APS Press).

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, *17*(3), 377-386.

International Association of Horticultural Producers AIPH (2016) International statistics flowers and plants. Vol 64 International Association of Horticultural Producers.

Islam, W. (2017). Management of plant virus diseases; farmer's knowledge and our suggestions. *Hosts and Viruses*, *4*(2), 28.

Jordan, R., Guaragna, M. A., & Hammond, J. (2016). Development of polyclonal and monoclonal antibodies to Rose rosette virus nucleoprotein. In *XIV International Symposium on Virus Diseases of Ornamental Plants 1193* 77-82.

Kim, J., Park, J. H., Lim, C. J., Lim, J. Y., Ryu, J. Y., Lee, B. W., ... & Kim, D. (2012). Small RNA and transcriptome deep sequencing proffer insight into floral gene regulation in Rosa cultivars. *BMC genomics*, *13*(1), 657.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722-736.

Kuulasmaa, T. (2002). Oligo Analyzer Primer Design Tool. *Version 1.0. 2. Kuopio: distributed by the author*.

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*.

Laney, A. G., Keller, K. E., Martin, R. R., & Tzanetakis, I. E. (2011). A discovery 70 years in the making: characterization of the Rose rosette virus. *Journal of General Virology*, *92*(7), 1727-1732.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357.

Leus, L., Van Laere, K., De Riek, J., & Van Huylenbroeck, J. (2018). Rose. *Ornamental Crops*. Springer, Cham, 719 -767.

Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., ... & Lam, T. W. (2016). MEGAHIT v1. 0: a fast and scalable metagenome assembler is driven by advanced methodologies and community practices. *Methods*, *102*, 3-11.

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., ... & Vainio, E. J. (2017). A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. *Frontiers in microbiology*, *8*, 45.

Milleza, E.J.M., Ward, L.I., Delmiglio, C., Tang, J.Z., Veerakone, S. and Perez-Egusquiza, Z. (2013) A survey of viruses infecting *Rosa* spp. in New Zealand. *Australas. Plant Pathol.* 42, 313– 320.

Moury, B., Cardin, L., Onesto, J. P., Candresse, T., & Poupet, A. (2001). Survey of Prunus necrotic ringspot virus in Rose and Its Variability in Rose and Prunus spp. *Phytopathology*, *91*(1), 84-91.

Mumford, R. A., Macarthur, R., & Boonham, N. (2016). The role and challenges of new diagnostic technology in plant biosecurity. *Food Security*, *8*(1), 103-109.

Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography*, *53*(3), 240-255.

Novick, D. N. (2017). *Evaluation of rose germplasm for resistance to rose rosette disease, and studies of disease transmission and vector management* (Doctoral dissertation, University of Delaware).

Olson, J., Rebek, E. J., & Schnelle, M. A. (2017). Rose rosette disease.

Pagan, I., & García-Arenal, F. (2018). Population genomics of plant viruses. In *Population Genomics: Microorganisms* (pp. 233-265). Springer, Cham.

Parasuraman, P., Pattnaik, S., & Busi, S. (2019). Plant-Microbe Interactions in Ecosystems Functioning and Sustainability. In *New and Future Developments in Microbial Biotechnology and Bioengineering* (pp. 255-266). Elsevier.

Pecman, A., Kutnjak, D., Gutiérrez-Aguirre, I., Adams, I., Fox, A., Boonham, N., & Ravnikar, M. (2017). Next-generation sequencing for detection and discovery of plant viruses and viroids: comparison of two approaches. *Frontiers in microbiology*, *8*, 1998.

Pscheidt, J. W., & Rodriguez, T. G. (2016). Diseases of Rose. *Handbook of Florists' Crops Diseases*, 1-29.

Rivera PA, Engel EA (2010) Presence of rose spring dwarf-associated virus in Chile: partial genome sequence and detection in roses and their colonizing aphids. Virus Genes 41(2):295–297

Rodriguez, M. J., Stansell, J. S., & Schneider, J. P. (2018). *U.S. Patent Application No. 15/981,206*.

Roossinck MJ (2015) Plants, viruses and the environment: Ecology and mutualism. 243 Virology 479–480:271–277.

Roossinck MJ, Bazán ER (2017) Symbiosis: Viruses as Intimate Partners. 247 doi:10.1146/annurev-virology.

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, *72*(1), 1-13.

Secor, G., Kong, M., & Nyland, G. (1977). Rose virus and virus-like diseases. *California agriculture*, *31*(3), 4-7.

Seguin, J., Rajeswaran, R., Malpica-Lopez, N., Martin, R. R., Kasschau, K., Dolja, V. V., ... & Pooggin, M. M. (2014). De novo reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PloS one*, *9*(2), e88513.

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., ... & Schneider, W. (2013). E-probe Diagnostic Nucleic acid Analysis (EDNA): a theoretical approach for handling of next-generation sequencing data for diagnostics. *Journal of microbiological methods*, *94*(3), 356-366.

Syller, J. (2012). Facilitative and antagonistic interactions between plant viruses in mixed infections. *Molecular plant pathology*, *13*(2), 204-216.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, *30*(12), 2725-2729.

Tatineni, S., Robertson, C. J., Garnsey, S. M., & Dawson, W. O. (2011). A plant virus evolved by acquiring multiple nonconserved genes to extend its host range. *Proceedings of the National Academy of Sciences*, *108*(42), 17366-17371.

Tešanović, M., Bonić, Ž., & Bošković, J. (2018). ORIGIN OF CULTIVATED ROSES AND APPROACHES USED TO STUDY IMPORTANT ROSE TRAITS. *Journal of Agronomy*, *26*.

Thomas E. & Scott C. (1953) Rosette of rose *Phytopathology* 43218219

Tzanetakis, I.E., Reed, J. and Martin, R.R. (2005) Nucleotide sequence, genome organization and phylogenetic analysis of Strawberry pallidosis associated virus, a new member of the genus crinivirus. *Arch. Virol.* 150, 273– 286.

Tzanetakis, I.E., Susaimuthu, J., Gergerich, R.C. and Martin, R.R. (2006b) Nucleotide sequence of Blackberry yellow vein associated virus, a novel member of the *Closteroviridae*. *Virus Res.* 116, 196– 200.

Tzanetakis, I.E., Halgrena, A. and Martin, R.R. (2007) Strawberry chlorotic fleck: identification and characterization of a novel *Closterovirus* associated with the disease. *Virus Res.* 124, 88– 94.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic acids research*, *40*(15), e115-e115.Windham, M., Windham, A., Hale, F., & Amrine Jr, J. (2014). Observations on rose rosette disease. *Amer. Rose*, *42*(9), 56-62.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), R46.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for a polymerase chain reaction. *BMC bioinformatics*, *13*(1), 134.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, *31*(13), 3406-3415.

CHAPTER V

DETECTION AND DISTRIBUTION OF CUCURBIT VIRUSES IN EASTERN OKLAHOMA

Abstract

Cucurbits such as cantaloupe, cucumber, melon, pumpkin, squash, and watermelon are cultivated in Oklahoma, and has decreased mainly caused by the prevalence of viral diseases. Quick identification of all the plant viruses infecting cucurbits may facilitate management, breeding and will allow control strategies to prevent the cross-border introduction of new detrimental viruses to the US. This research aims to detect cucurbit infecting viruses by using High Throughput Sequencing (HTS) combined with EDNA-Cucurbits MiFi™, a bioinformatics pipeline. This broad detection method was validated with Multiplex RT-qPCR combined with High-Resolution Melting (HRM), allowing sensitive detection, confirmation, and discrimination of the screened plant viruses. Unique and pathogen-specific Electronic probes (E-probes) were generated and uploaded directly to EDNA-Cucurbits MiFi™. The EDNA-Cucurbits MiFi™, rapidly detect and quantify plant viruses infecting cucurbits. This detection method has the potential to be applied in epidemiological, breeding, and management studies of viral diseases in cucurbit crops.

1. Introduction

Cucurbits are one of the largest horticultural commodities in the United States. Oklahoma produces cucumber, melon, squash, pumpkin, and watermelon, which are grown in about 5,500 acres annually (USDA, 2017). The total production in the U.S. is 26.1 million tons (FAO, 2013). Watermelon and cantaloupe had the highest value of specialty crops reported during 2019 (USDA, 2019). Cucurbits are produced during the summer in areas with abundant humidity (Mc Creight, 2016).

Cucurbits are susceptible to a number of diseases and insect pests. Fungi, bacteria, nematodes, and viruses are the most common pathogens that cause damage to this crop. The cultivated acreage of cucurbits in Oklahoma had decreased mainly due to viral diseases generating a marketable loss of up to three percent in watermelon only in 2016 (USDA, 2016). Moreover, the total value of this crop was $ 4.71 U.S. million dollars for 2016, down 13 percent respect the previous year (USDA, 2016-2017). Disease control is not always successful and is done using cultural practices, including chemical protectants and host plant resistance (Mc Creight, 2016). Cultivated cucurbit germplasm has been greatly improved by common breeding techniques. Disease resistance is one of the most important features for crop improvement with other marketable traits, and finding new alleles and genes to increase production is also important (Grumet et al., 2017).

The natural incidence of viral infections in Oklahoma has been reported in field trials. *Papaya ringspot virus* (PRSV), *Cucumber mosaic virus* (CMV I and CMV II), and *Squash mosaic virus* (SqMV) in mixed infections of more than two viruses were detected among cucurbit crops, at different plant growing stages (Ali, 2012). Early viral infection in cucurbit plants can significantly reduce yield (Fletcher, 2000). More than 20 viruses have been discovered causing severe viral infection in cucurbits (Keinath, et. al., 2017). *Cucurbit green mottle mosaic virus* (CGMMV), *Cucumber mosaic virus* (CMV I and CMV II), *Cucurbit aphid-borne yellows virus* (CABYV), *Cucurbit yellows stunting disorder virus* (CYSDV), *Melon necrotic spot virus* (MNSV), *Papaya*

ringspot virus (PRSV), *Squash mosaic virus* (SqMV), *Watermelon mosaic virus* (WMV), and *Zucchini yellows mosaic virus* (ZYMV), are among the most common viruses in nature (Ali & Abdalla, 2012).

The family Cucurbitaceae comprise 1000 plant species in 96 genera, which are divided into five phylogenetically and statistically well-supported clades. The cucurbit genomes are well annotated and have homologous linkage groups for economically important genera belonging to the genera Cucurbiteae and Benincaseae (Schaefer & Renner, 2011). The sample collection for this study was focused on members corresponding to these two genera (Figure 5.1). The large genetic difference in the family Cucurbitaceae depends on the diversity of monoploid chromosome numbers within the genera Cucurbitaceae (Grumet et al., 2017).

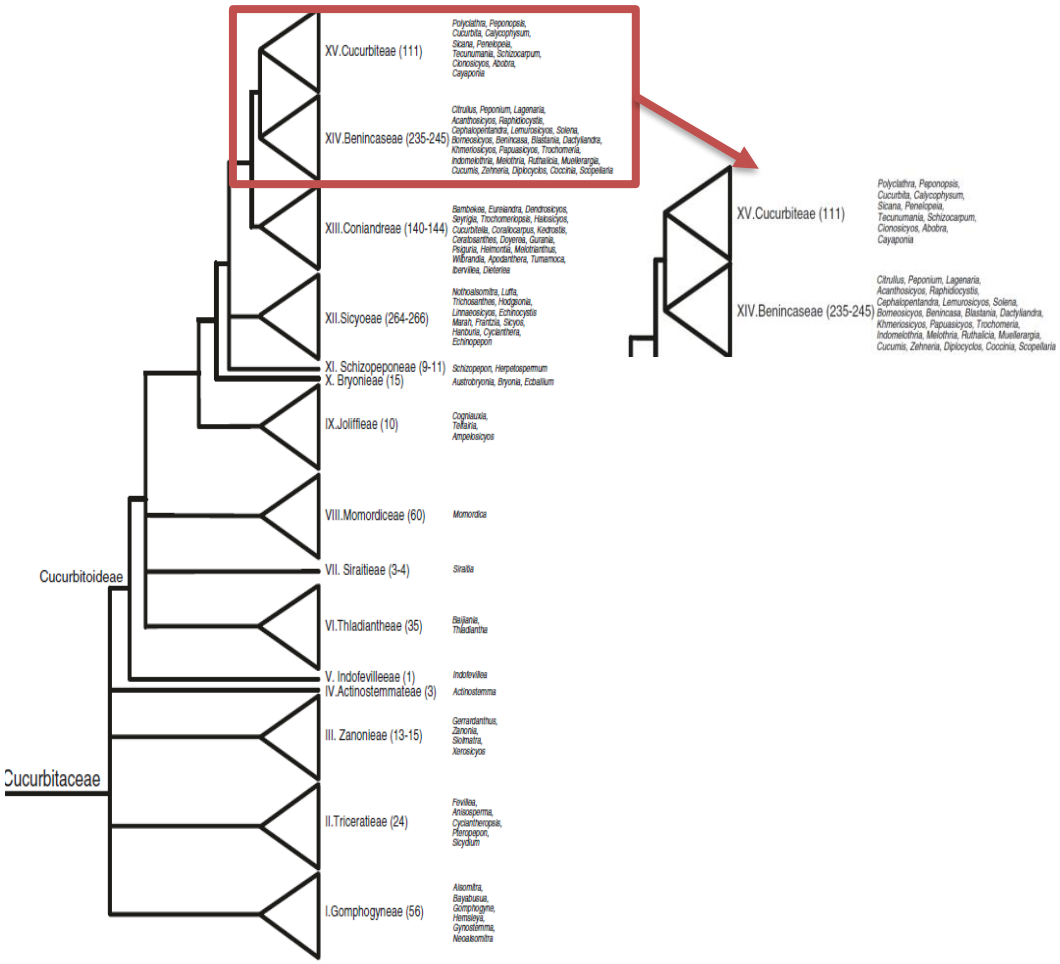


Figure 5. 1 Maximum likelihood Bayesian phylogenetic cladogram of Cucurbitaceae modified from Schaefer & Renner (2011) elaborated considering chloroplast, nuclear and mitochondrial DNA sequences. The orange box contains two clades of the main species sampled in Oklahoma counties during the summers of 2016-2018.

Plant viruses co-evolve with their specific host. Virus-host, virus-vector, and plant-virus-vector interactions studies had identified host and vector proteins involved in disease development and transition among cultivars (Wang, 2015). Mixed infections are determined by an antagonist or synergist virus-virus interaction within a host. A synergistic interaction may lead to an economically damaging disease. In contrast, antagonistic interaction may lead to cross-protection of a host. Breeding strategies for cucurbit virus-resistant cultivars often relay on the introgression of resistance genes found in wild related species (Lecoq et al., 1998). The selection of resistance genes is based in host-virus interaction; genetic resistance corresponds to a level of virulence of the pathogen. Localized viral infections are associated with dominant alleles in the host. Systemic lesions correlate to recessive alleles or deficient dominant genes with few symptom expressions. Finally, complete immunity is derived from fully recessive alleles (Fraser, 1992). Viral disease management by breeding programs are successful by the deployment of host resistance in the field (Mascia & Gallitelli, 2016). Virulence determinants of plant viruses can be mapped in the viral genome to control loss of pathogenic fitness-enhancing durable resistance (Fraser, 1992).

Evolutionary changes in the coat protein of plant viruses (HC-Pro protein of plant-infecting RNA genome viruses) are influenced by agriculture, farming, grafting, and germplasm trading. This may enable effective transmission and host jump within a close genetic relative (Gibbs at al., 2020). Plant-virus prevalence and molecular diversity studies in mixed viral infection revealed similar symptomatology expression in the host (Ali et al., 2012). The family Cucurbitaceae is a complex of species with high genetic differences. In Oklahoma, cucurbit production fields are located into Cucurbitaeae and Bernicaceae, and mainly into *Cucurbita* and *Citrullus*. Although

there are many studies done to analyze the dynamic changes within viral infections in cucurbit crops (Grumet et al., 2017), most of them are general not taking into account the genetic differences among the hosts. Therefore, there is a need to analyze virus abundance in mixed infections within samples collected from these two groups (*Cucurbita* and *Citrullus*), which were accomplished over the summer of 2016 and 2018.

Among the existing methods for viral detection and diagnostics, some are limited in their capacity and versatility. Therefore, there is a need for a broad, reliable, and relatively rapid and simultaneous detection and discrimination of cucurbit viruses. Advances in High Throughput Sequencing (HTS) technologies currently allow to study microbial communities and pathogen discovery (Massart et.al., 2017). Due to the deep sequencing capability of the entire amplification of microbes genomes present in a host, HTS had become the most suitable available technology for multiple pathogen detection. On the other hand, Electronic Diagnosis Nucleic acid Analysis (EDNA-MiFi™) is a multiple pathogen detection bioinformatics software. Unique and specific electronic probes (E-probes) targeting just the infective pathogen can be precisely designed (Stobbe et.al., 2013; Espindola et.al., 2018) to be combined to work with HTS.

The validation of HTS detection requires a side to side comparison with accepted and approved, standardized and validated detection methods. Multiplex reverse transcription – quantitative polymerase chain reaction (RT-qPCR) is a sensitive, rapid and reliable method for the detection of RNA viruses (Elnifro et al., 2002). In this research, a double pathogen detection system is presented. The validation of a Multiplex RT-qPCR will support the positive results obtained by EDNA-Cucurbits. The selected viruses for RT-qPCR are the ones reported a presence in Oklahoma fields along with the ones of mayor concern of introduction (Ali & Abdalla, 2012). A brief description of the selected viruses for Multiplex RT-qPCR is detailed in Table 5.1 which also includes the complete reported virome that infects Cucurbitaceae.

Table 5.1 summarizes the main characteristics and genomic features of the selected viruses in this research.

Table 5. 1 The reported Cucurbit virome and main virus features.

Virus	Genus	Acronym	Nucleic acid	polyA	Main features
Alfalfa mosaic virus	Alfamovirus	AMV	Segmented, tripartite linear ssRNA (+)	No	Three genomic and one subgenomic segments are encapsidated in distinct particles
Bean pod mottle Virus	Comovirus	BPMV	Segmented, bipartite linear ssRNA(+)	Yes	Two genomic RNAs encapsidated separately into two different particles
Cucurbit aphid-borne yellows virus	Polerovirus	CABYV	Monopartite, linear, ssRNA(+)	No	Monopartite no-enveloped particle
Cucumber green mottle mosaic virus	Tobamovirus	CGMMV	Monopartite, linear, ssRNA(+)	No	Monopartite no-enveloped helical particle
Cucumber mosaic virus	Cucumovirus	CMV	Segmented, tripartite linear ssRNA(+)	No	Three genomic segments, encapsidated in one particle
Cucurbit yellow stunting disorder virus	Crinivirus	CYSDV	Bi-partite linear ssRNA(+)	No	Non-enveloped, bipartite filamentous particles with four subgenomic RNAs at RNA2
Melon necrotic spot virus	Gammacarmovirus	MNSV	Linear, ssRNA(+)	No	Monopartite non-enveloped, spherical particle
Papaya ringspot virus	Potyvirus	PRSV	Monopartite, linear, ssRNA(+)	Yes	Monopartite non-enveloped, flexus, filamentous particle
Squash leaf curl virus	Begomovirus	SLCuV	Bi-partite circular, ssDNA genome (+)	No	Bi-partite non-enveloped icosahedral particles
Soybean mosaic virus	Potyvirus	SMV	Monopartite, linear, ssRNA(+)	Yes	Monopartite non-enveloped, flexus, filamentous particle
Squash mosaic virus	Comovirus	SqMV	Segmented, bipartite linear ssRNA(+)	Yes	Two genomic RNAs encapsidated separately into two different particles
Squash vein yellowing virus virus	Begomovirus	SqVYV	Bi-partite circular, ssDNA genome (+)	No	Bi-partite non-enveloped icosahedral particles
Tobacco ringspot virus	Nepovirus	TRSV	Segmented, bipartite linear ssRNA(+)	Yes	Bi-partite non-enveloped icosahedral particles
Watermelon mosaic virus	Potyvirus	WMV	Monopartite, linear, ssRNA(+)	Yes	Monopartite non-enveloped, flexus, filamentous particle

Watermelon silver mottle virus	Othotospovirus	WSMoV	Tri-partite linear ssRNA(-)	No	Three genomic segments, encapsidated in one particle
Zucchini yellow mosaic virus	Potyvirus	ZYMV	Monopartite, linear, ssRNA(+)	Yes	Monopartite non-enveloped, flexus, filamentous particle

Highlighted cells belong to polyA tail viruses.

To achieve the highest target specificity and sensitivity during molecular detection assay design primer thermodynamics are considered a cornerstone. Primer design software facilitates this process and the gene walking for efficient oligonucleotide primer selection (Pasin et al., 2019). Software for primer design uses accurate thermodynamic models, and algorithms to predict and reduce primer secondary structures and primer dimers. Primer design software also allows multiple primer target recognition. Polymerase chain reaction (PCR) requires specific and sensitive primers for accurate and sensitive detection (Untergasser, 2012).

uMelt™ is a bioinformatic tool for predicting DNA melting curves and fluorescent denaturation curves of PCR products (Dwight, 2011). The software allows us to calculate and visualize the mean helicity and the dissociation probability of the PCR product sequences at a temperature range represented as prediction curves.

High resolution melting (HRM) is a post-PCR analytical technique that allows identifying differences in the melting points of amplified PCR products. The thermocycler can be program to increase the temperature from 60-100 ^{o}C after PCR, which causes a loss of fluorescence of the PCR obtained product. The measurement of the decrease of fluorescence is quantified with software assistance (~ every 0.06 – 0.25 ^{o}C) and the specific Temperature of melting (Tm) is calculated (Donna, 2016). This method has several advantages over an Endpoint PCR since HRM does not require post-PCR gel processing, does not use dye-labeled primers and detects the presence of DNA in a multiplex assay (Elkins, 2016).

The specific objectives of this study are: 1. To test *in-vitro* EDNA- Cucurbits MiFi™, to confirm and validate EDNA outputs with multiplex RT-qPCR HRM to identify plant viruses infecting cucurbits. 2. To determine the prevalence of cucurbit related viruses *Cucurbita* and *Citrullus* hosts in summer 2016 - 2018 in Oklahoma.

2. Materials and methods

2.1 Source of viruses and infected plant material

Seven lyophilized reference positive controls *Cucurbit green mottle mosaic virus* (CGMMV), *Cucumber mosaic virus* (CMV I and CMV II), *Melon necrotic spot virus* (MNSV), *Papaya ringspot virus* (PRSV), *Squash mosaic virus* (SqMV), *Watermelon mosaic virus* (WMV), and *Zucchini yellows mosaic virus* (ZYMV) were sourced from Agdia, Inc (Agdia, Elkhart, IN). Two other viruses *Cucurbit aphid-borne yellows virus* (CABYV), *Cucurbit yellows stunting disorder virus*(CYSDV) were provided by USDA (Fort Pierce, FL).

Lyophilized reference controls, including near-neighbor viruses and viruses commonly infecting cucurbits, were used in an exclusivity panel. Healthy cucumber (*Cucumus sativus)* seeds were planted and plants harvest to be used as the negative control healthy tissue. The plant tissue was stored in −80 °C at the Institute of Biosecurity and Microbial Forensics (IBMF), Oklahoma State University.

In the field cucurbits related viruses represent complex and constantly changing pathosystem (Lecoq, 1998) Field surveys were conducted in summer 2016 and 2018 to detect cucurbit virus presence in cucurbit-growing counties. Samples of plant tissue with characteristic symptoms of viral infection were collected as part of the annual 2016 and 2018 survey of cucurbit infections in Muskogee and Tulsa County, Oklahoma respectively. Table 5.2 describes the number of samples used in this study as well as the location of the sampling, and symptomatic host tissue. Plant tissue was stored in −80 °C until processing.

Table 5. 2 Summary of collected field samples, indicating year of survey, sampled county and cucurbit host.

Summer	No. Samples	Sampled County	Host
2016	74	Muskogee	Pumpkin, Watermelon, and Squash
2018	15	Tulsa	Pumpkin and Watermelon.

2.2 Library preparation and sequencing

Theoretically, extracting total nucleic acid from a plant sample followed by deep sequencing allows the detection of all microbes present in a sample for further identification. HTS can be used as a phytobiome detection tool, also capable for microbe discovery.

High-quality total RNA was extracted with the RNeasy Plant Mini Kit (QIAGEN, USA), ds cDNA was amplified by NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Kit (NEB, USA). Library preparation uses a transposase enzyme that fragments the ds cDNA into long reads and attaches adapters for downstream analysis. Library preparation was performed according to the Nanopore manufacturer's protocol. The processed libraries were kept on ice -20 ºC until ready to load on the MinION device. Barcoding library preparation was performed according to Oxford Nanopore's protocol for the kit and the Ligation Sequencing kit (SQK-LSK108). Barcoded libraries and dsDNA were quantified by Qubit4™ (ThermoFisher Scientific, USA), the six barcoded libraries were equimolar-pooled prior to adding the adapters. The final library was stored at –20°C until MinION sequencing.

Sequencing was performed in the MinION loaded in the flow cell (FLO-Min106 R9.4; Oxford Nanopore Technologies, UK). Platform Quality Check (QC) was performed to determine the number of active pores available in the flow cell. A total of 412 active pores after QC were reported. The priming of the flow cells was done as per manufacturer's recomended procedure. The pooled library mixed with Library Loading Beads (LLB; Oxford Nanopore Technologies) was loaded onto the SpotON port of the flow cell to initialize the run.

2.3. Rapid virus detection using EDNA-Cucurbits (MiFi^{TM}) database

EDNA-Cucurbits Mi/Fi^{TM} parameters were 100% percent identity and query coverage to assess six metagenomes separated by barcodes generated by MinION sequencing. Hit frequencies between raw reads with E-probes were recorded for each of the selected varieties. Data on hit frequencies were analyzed with Tukey's HSD test and pairwise T-test at P value= 0.05. The EDNA-Cucurbits bioinformatic pipeline aligns not assembled metagenomic reads and quantifies the matches with the Cucurbit E-probe database. A semi-quantification of the number of matches of EDNA-Cucurbits generated a statistical correlation of the tested samples to the designed validation method (Multiplex RT-qPCR-HRM).

2.4. Comparison with standard bioinformatic pipeline EPI2ME

To validate further detection from EDNA-Cucurbits, resulting metagenomes from Oxford Nanopore sequencing were separated into barcodes (demultiplexed) as .fastq files. EDNA-Cucurbits results from six samples collected during the summer of 2018 were compared with the EPI2ME (Oxford Nanopore) taxonomical identifier in order to compare and determine the taxonomic ID of the analyzed reads. The EPI2ME database reports all of the taxonomic nodes to form a phylogenetical tree down to the species level.

2.5. Primer design for Multiplex RTq-PCR HRM

The first step in developing an EDNA validation assay is designing primer sequences to query the results found by EDNA with a standardized detection method in this case a Multiplex RT-qPCR HRM. Virus sequences were retrieved from the National Center for Biotechnology Information (NCBI). Targeted genes were the RNA dependent RNA polymerase (RdRp), and the capsid protein. The retrieved sequences were aligned using Mega 6.0 and BioEdit, the Muscle tool for grouping the viruses. The aligned sequences were analyzed by a Maximum likelihood phylogenetic tree was made.

Software Primer3, PrimerQuest®, and Primaclade were used for primer design, taking into account the specified primer design features described. In order to check primer specificity, primers were uploaded to PrimerBLAST. To analyze the thermodynamic features of primers Olygoanalyzer and mFOLD™ were used. In general, lower free energy to avoid secondary and self-complementarity structures were pursued. Primer pairs were selected with different amplicon sizes to allow product differentiation in agarose gel. uMelt™ was used to predict the HRM curve. Consensus sequences of predicted PCR products were used to design specific primers.

Once primer design was completed and validated individually with each targeted virus a Multiplex RT-qPCR-HRM was tested for the detection of the targeted positive controls in a validation process.

2.6. HRM uMelt™ prediction curves

uMelt™ prediction of HRM curves was obtained by using the uMelt™ batch option. Multiple sequences of the PCR products were uploaded. Selected settings were: mono cations 50mM, free [Mg^{++}] 1.8 mM based on PCR chemistry of the TaqPolymerase to be used, thermodynamic library Blake & Decourt (Nucleic Acids, 1998), temperature range 65 to 95 ^{o}C with high resolution of 0.25 ^{o}C.

2.7. Direct trapping of virions to plastic

Direct trapping of virions onto PCR tubes was performed as described by Babu, et. al., 2016. This method was the one selected for the first screening of the field samples. Vials containing cucurbit virus-positive controls (100mg) were reconstituted in 450 uL of phosphate-buffered saline (PBS) pH 7.4 solution containing 0.05% of Tween-20 (1X PBS-T) buffer. An aliquot of fifty microliters of the crude sap was aliquoted in a 0.2 uL PCR tube, avoiding bubbles. The tubes were incubated on ice for two minutes. Then, the sap was completely removed, and the tubes washed twice with 50uL of 1X PBS-T buffer. Thirty microliters of DEPC-water, containing RNA-sin (100U/mL)

was added to the PCR tube, followed by immediate denaturation at 95 ºC for one minute and cooled on ice for one minute. The resulting solution was used for cDNA synthesis.

2.8. cDNA synthesis

cDNA was synthesized directly into PCR trapping tubes from the positive controls and symptomatic or non-symptomatic plant tissues. Four microliters of RNA was used to synthesize the first-strand cDNA. Random hexamer primers and Moloney murine leukemia virus (MML-V) reverse transcriptase (Promega, Madison, WI, USA) were used following the manufacturer's instructions.

2.9. Single RT-qPCR HRM

Single virus RT-qPCR assays were performed in 10 µl reaction consisting of 5 µl HotStart. Master Mix (Biolabs), 1 µl of LCgreen (Biolabs), 0.5 µl of each forward and reverse primer (7.5 µM), 2 µl of cDNA template, and 1 µl nuclease-free water (Ambion, Austin, TX, USA). Individual virus RT-qPCR coupled with High-Resolution Melting was performed in a Rotor-Gene thermal cycler (QIAGEN, USA) as follows: initial denaturation of 94∘C for 4 min followed by 35 cycles of denaturation at 94∘C for 20 s, annealing at 54∘C for 60 s, extension at 72∘C for 40 s, and a final extension at 72∘C for 3 min. Finally, the amplified PCR product was settled for HRM from 65 to 99ºC, temperature range. Positive, negative controls, non-template control, and healthy tissue were included. A sensitivity assay starting at 1 ng to 1 fg was performed following previously described settings above.

2.10. Multiplex RT-qPCR HRM

Multiplex RT-qPCR assays were performed in 10 µl reaction volumes consisting of 5 µl HotStart. Master Mix (Biolabs), 1 µl of LCgreen (Biolabs), 0.5 µl of each forward and reverse primer (7.5 µM), 2 µl of cDNA template, and 1 µl nuclease-free water (Ambion, Austin, TX, USA). Multiplex RT-PCR coupled with High-Resolution Melting was performed in a Rotor-Gene thermal cycler (QIAGEN, USA) as follows initial denaturation of 94◦C for 4 min followed by 35 cycles of

denaturation at 94∘C for 20 s, annealing at 54∘C for 60 s, extension at 72∘C for 40 s, and a final extension at 72∘C for 3 min. The amplified PCR product continued in the thermocycler for further HRM from 65 to 99 ºC, with increments of 0.2 seconds. Positive, negative controls, non-template control, and healthy tissue were included. A sensitivity assay starting at 1 ng to 1 fg was performed following previously detailed settings and specificity assays were determined.

2.11. Inclusivity and exclusivity panel

The specificity of the primers was tested *in silico* and *in vitro* testing an inclusivity and exclusivity panel (Table 5.1). The alignment of primer sequences against the GenBank database using BLASTn was tested at 100% query coverage and 100% identity. The *in vitro* specificity assays were performed by RT-qPCR-HRM with the individual primer set against a panel of taxonomically near-neighbors. The cross-reactivity of all primer sets also was tested against the cDNA of three cucurbit species zucchini 'Easy Pick Gold', yellow watermelon and crookneck squash from leaf tissue.

3. Results

3.1. Rapid virus detection using EDNA-Cucurbits (MiFi™) database

De-multiplexed unassembled reads tested against EDNA-Cucurbits (MiFi™) database of curated E-probes showed the presence of positive hits to PRSV, TRSV, and WMV. Nonetheless, ZYMV (*Potyvirus*) was not detected. According to the result with Table 5.3, higher numbers of hits (Figure 5.2) were affiliated with *Potyvirus* and *Nepovirus*. All positive hits were identified of sequences that match up with specific virus sequences

containing a Poly A tail.

Table 5. 3 Summary of positive semi-quantitative results for cucurbit samples summer 2018 tested against EDNA-Cucurbits (MiFi™) database.

Sample	Virus	Number of hits
	PRSV	28
	TRSV	1650
Barcode01	ZYMV	8
	PRSV	731
	TRSV	40
Barcode02	WMV	324
	PRSV	76
	TRSV	2614
Barcode03	WMV	104
	CABYV	12
	PRSV	2850
	SMV	41
	TRSV	2
	WMV	5734
Barcode04	ZYMV	5482
	CABYV	12
	PRSV	2850
	SMV	41
	TRSV	2
	WMV	5734
Barcode05	ZYMV	5482
	PRSV	5683
	SMV	52
	TRSV	10
	WMV	3158
Barcode06	ZYMV	4462

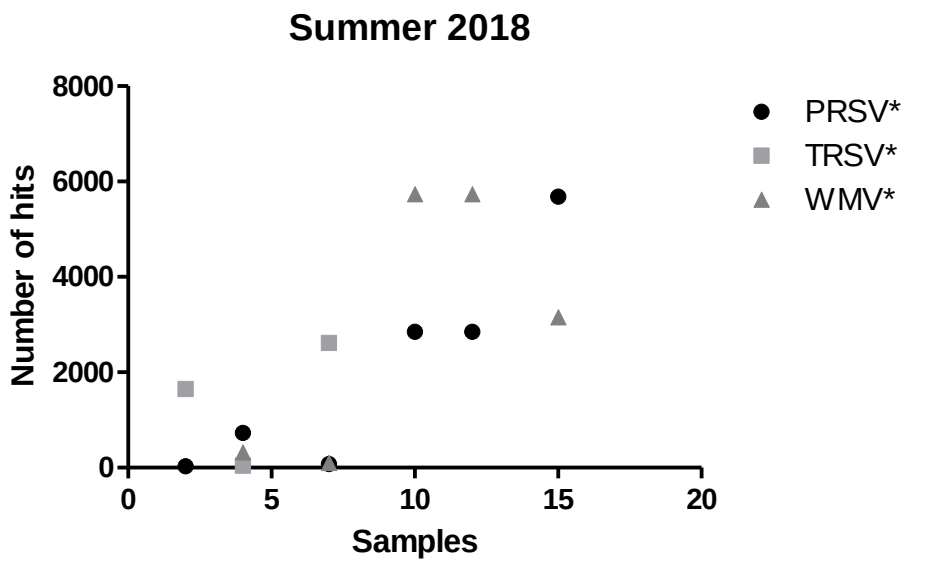


Figure 5. 2 Reported positive hits from each barcoded cucurbit samples. Y-axes show the number of hits corresponding to the virus that was reported as positive by EDNA-Cucurbits (MiFi™). Samples are represented in Y axes.

3.2. Oxford nanopore sequencing and EPI2ME Analysis

Oxford Nanopore sequencing reported 140 K reads. A total of 51059 reads were analyzed. Table 5.4 shows a detailed summarized report from the matching virus per barcode. EPI2ME analyzed reads did not match with any host background in any of the sequenced barcodes. However, a BLASTn against the non-redundant database showed high similarity with *Cucurbita maxima* the host mRNA (GTP binding nuclear protein Ran-3-like LOC11491953). BLASTn showed 81% Query coverage, 89.21% of Identity, and an E-value of $7e^{-73}$, belonging to the accession XM_023141057.1.1. As expected we found cucurbit infecting viruses belonging to the *Potyvirus* and *Nepovirus* genus. Also, the presence of other related viruses present in the sample, but with few reads were assigned.

Table 5. 4 Summarized EPI2ME taxonomical classification results for cucurbit barcodes summer 2018.

County	Sample date	Sample ID	Barcode	Read per barcode	Genus	Significant Viral Results	# matching reads
Tulsa	Summer2018	2T	BC01	1397	Nepovirus Potyvirus	775 *Tobacco ringspot virus*- 2 *Potato black ringspot virus* 10 *Papaya ringspot virus*	777 10
Tulsa	Summer2018	4T	BC02	2399	Nepovirtus Potyvirus Caferiavirus	9*Tobacco ringspot virus* 188 *Papaya ringspot virus-86 Watermelon mosaic virus - 2Tomato necrotic stunt virus-1Telosma mosaic virus* 1 *Cafeteria roenbergensis virus BV-PW1*	9 277 1
Tulsa	Summer2018	7T	BC03	2563	Nepovirus Potyvirus Mamastrovirus Furovirus	594 *Tobacco ringspot virus*- 1 *Potato black ringspot virus* 23 *Watermelon mosaic virus -19 Papaya ringspot virus -1Cowpea aphid-borne mosaic virus-1Zucchini yellow mosaic virus* 1 *Porcine astrovirus 3* 1 *Sorghum chlorotic spot virus*	595 44 1 1
Tulsa	Summer2018	10T	BC04	3104	Potyvirus Ichnovirus Clorovirus	571 *Papaya ringspot virus-2Zucchini yellow mosaic virus - 1 Moroccan watermelon mosaic virus* 1 *Glypta fumiferanae ichnovirus* 1 *Paramecium bursaria Chlorella virus 1*	574 1 1
Tulsa	Summer2018	15T	BC05	15896	Potyvirus Polerovirus	7151 *Zucchini yellow mosaic virus - 2967 Watermelon mosaic virus - 1064 Papaya ringspot virus - 35 Tomato necrotic stunt virus - 24 Calla lily latent virus 23 Soybean mosaic virus - 6Telosma mosaic virus - 4 Yambean mosaic virus - 3 Bean common mosaic necrosis virus - 3 Potato virus Y* 9 *Cucurbit aphid-borne yellows virus - 1 Melon aphid-borne yellows virus - 1 Pepo aphid-borne yellows virus*	7319 13

					Triatovirus	2 *Homalodisca coagulata virus-1*	2
					Nepovirtus	1 *Tobacco ringspot virus*	1
					Felixo1virus	1 *Erwinia phage phiEa104*	1
					Orthotospovirus	1*Tomato zonate spot virus*	1
					Carlavirus	1*Garlic common latent virus*	1
					Mamastrovirus	1*Qinghai Himalayan marmot astrovirus 2*	1
					Marafivirus	1*Grapevine Syrah virus 1*	1
					Gammacoronavirus	1*Turkey coronavirus*	1
					Hepatovirus	1Phopivirus	1
Tulsa	Summer2018	12T	BC06	25700	Potyvirus	*8513 Papaya ringspot virus-8184 Zucchini yellow mosaic virus - 1131Watermelon mosaic virus - 70 Yambean mosaic virus- 24 Potato virus Y - 13 Bean common mosaic necrosis virus - 11 Yam mosaic virus - 8 Colombian datura virus - 6 Moroccan watermelon mosaic virus - 5 Bean common mosaic virus - 4 Cowpea aphid-borne mosaic virus - 2Leak yellow stripe virus - 2 Soybean mosaic virus -2 Calla lily latent virus- 2 Zucchini tigre mosaic virus - 1 Hardenbergia mosaic virus- 1Jasmine virus T - 1Narcissus late season yellows virus- 1Pennisetum mosaic virus- 1 Cowpea aphid-borne mosaic virus 4 - 1 Leek yellow stripe virus -2 Soybean mosaic virus -2 Calla lily latent virus- 2 Zucchini tigre mosaic virus -2 Hardenbergia mosaic virus -1Jasmine virus T - 1Narcissus late season yellows virus -1 Pennisetum mosaic virus - 1 Telosma mosaic virus- 1 Tobacco etch virus- 1Impatiens flower break potyvirus*	18019

			Nepovirtus	6 *Tobacco ringspot virus*- 1 S*oybean latent spherical virus*	7
			Felixo1virus	2 *Erwinia phage phiEa104*	2
			Orthotospovirus	2 *Pepper chlorotic spot virus*	2
			Betapartitivirus	1 *Rosellinia necatrix partitivirus 6*	1
			Bromovirus	1*Broad bean mottle virus*	1
	Total reads	51059			

3.3. *In silico* analysis of the primers

Nine sets of primers, designed to amplify a genomic RNA segment from the nucleocapsid RNA dependent RNA polymerase (RdRp), and the capsid protein genes were designed to perform in Multiplex RT-qPCR-HRM. Table 5.5 describes the thermodynamic features of all the primers sets. *In silico* analysis of these primer sets showed they have 97–100% identity and a query coverage of 100% after using BLAST. The average means E-value was 3e-17. No unspecific match was detected with other viruses and the three host genomes, including viruses listed in the exclusivity and inclusivity panels.

Table 5. 5 Multiple infection species-specific primer sets showing the amplification product size, and thermodynamic features.

Target Species	Gene	Primer Name	Start	Primer Sequences (5'-3')	Length (bp)	GC%	*Tm	**Δg (Kcal/mol)	^any	^^3'	Product (bp)
Cucurbit aphid-borne yellow virus (CABYV)	RdRp	CABYV_F1	5004	CGACGTTGATGATGACGATTTATG	24	41.67	63.59	0.1	4	0	166
		CABYV_R1	5169	TCCAGGATTGAACTCGGTAATG	22	45.45	61.22	0.8	5	0	
Cucurbit green mottle mosaic virus (CGMMV)	RdRp	CGMMV_F1	4544	TGGCTAGTATGTTGCCGTTAG	21	47.62	58.02	0	4	0	114
		CGMMV_R1	4657	GGTTGGCAGTAGCCTGTATATC	22	50	57.8	0	4	2	
Cucumber mosaic virus (CMV I- II)	Coat protein	CMV_noflap_F1	1953	TCTCTCTCTCTTCTCCCTCCGATTCCTGTG	30	53.3	63	1	2	0	124
		CMV_noflap_R1	2075	TCTCTTCACACACCACGACTGACCATTTTAG	31	45	61	1	0	0	
Cucurbit yellows stunting disorder virus (CYSDV)	RdRp	CYSDV_F1	2213	GATTCAGGTGAGGCGTACTATC	22	50	57.84	0.7	4	1	234
		CYSDV1-R1	2446	TCCCTCGACAGGTACTCTTATT	22	45.45	57	0.3	4	1	
Melon necrotic spot virus (MNSV)	RdRp	MNSV_F1	3257	CAGATTTACTCGGCTCCGTATC	22	50	60.12	0	4	2	284
		MNSV_R1	3540	TGCTACCAGCACCAGAATAAG	21	47.62	58.07	0.1	5	0	
Papaya ringspot virus (PRSV)	RdRp	PRSV_F1	6685	GGATTCAGACAGCTTCGATAGG	22	50	62,03	0.7	0	0	76
		PRSV_R1	6760	CGCGGAGTTTCGGACATAATA	21	47,61	62,11	0.7	2	1	
Squash mosaic virus (SqMV)	RdRp	SqMV_F1	1342	GGACTCCTTGGAGGATCTTATTT	20	50	60.07	0	6	1	206
		SqMV_R1	1547	GGACTCCTTGGAGGATCTTATTT	23	43.48	59.02	0	6	0	
Watermelon mosaic virus (WMV)	RdRp	WMV_F1	8467	GATGAGGATGTGTGGCTGTATG	22	50	60.41	0.9	2	0	48

WMV_R1	8514	CTCAGCGAATGAAGCACTTAGA	22	45.45	59.79	0.7	4	1

Sequences of cucurbit infecting virus-specific primers and their thermodynamic features used for Multiplex RT-qPCR combined with HRM. *The melting temperature of the primer calculated using Primer 3. **Plot δG value in plot calculated by mFOLD. ^ The self-complementarity score of the oligo (a tendency of the oligo to anneal to itself or form a secondary structure) calculated using Primer 3; ^^3’ self-complementarity of the oligo (a tendency to form a primer-dimer with itself) calculated using Primer3.

3.4. Direct virion trapping analysis

Recovery of the viral nucleic acid was obtained in all PCR treated tubes at pH 7.4 as reported by Babu, et. al. (2016). This technique was used as the trapping method of choice for virion and RNA recovery in this research (Figure 5.3).

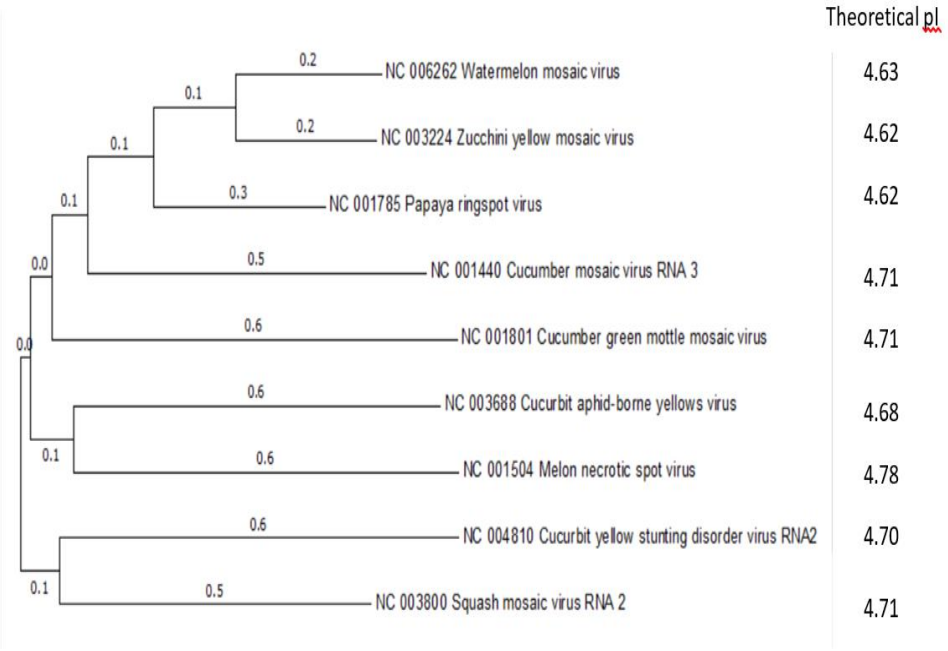


Figure 5. 3 Maximum likelihood phylogenetic relationship of amino acid sequences of virus capsid proteins and their theoretical isoelectric point of the studied viruses. The theoretical isoelectric point ranged from 4.62 to 4.78.

3.5. Individual RT-qPCR

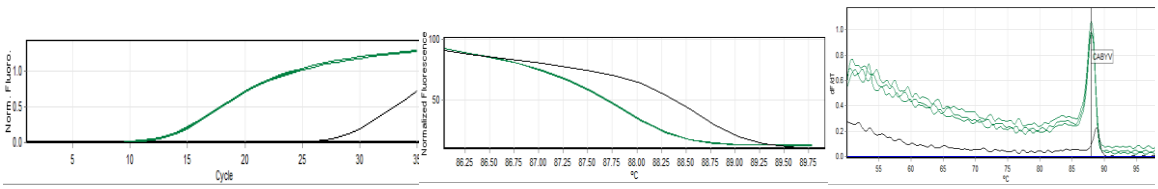
The performance of all primer sets was assessed in single virus reactions using End-Point PCR, and RT qPCR-HRM. The optimal annealing temperature was 54 ^{o}C and the obtained PCR product sequences after BLASTn showed a high percentage of identity and query coverage. Single End-point reactions with each of the primer set amplified the expected amplicon size of each target (~90-250bp).

The predicted melting temperature (Tm) of each primer set varies +/- 2ºC from the Tm obtained *in-vitro* PCR products. Figure 5.4 shows the individual amplification of the expected targets listed in Table 5.5. The Tm of the obtained PCR products used for virus discrimination is shown in Table 5.6.

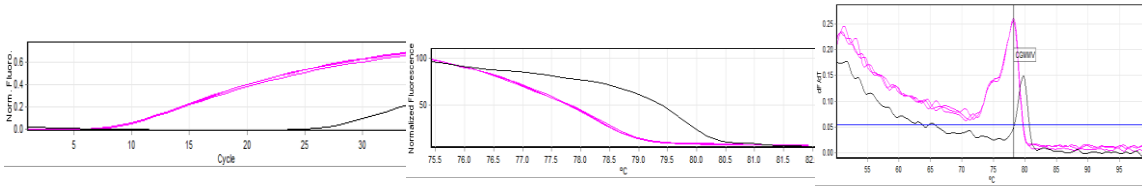
Table 5. 6 The difference between uMelt^{TM} predicted Tm of obtained amplicons compared with the final obtained Tm as measured after RT-qPCR-HRM

VIRUS	PREDICTED Tm (°C)	FINAL MEASURED Tm (°C)
CMV	83	84.6
CYSDV	80	81.35
MNSV	85.8	85.08
PRSV	76.9	77.1
SqMV	83.3	83.5
ZYMV	82	82.8
WMV	78.9	76.2
CABYV	86.5	87.93
CGMMV	81	78.7

a.



b.



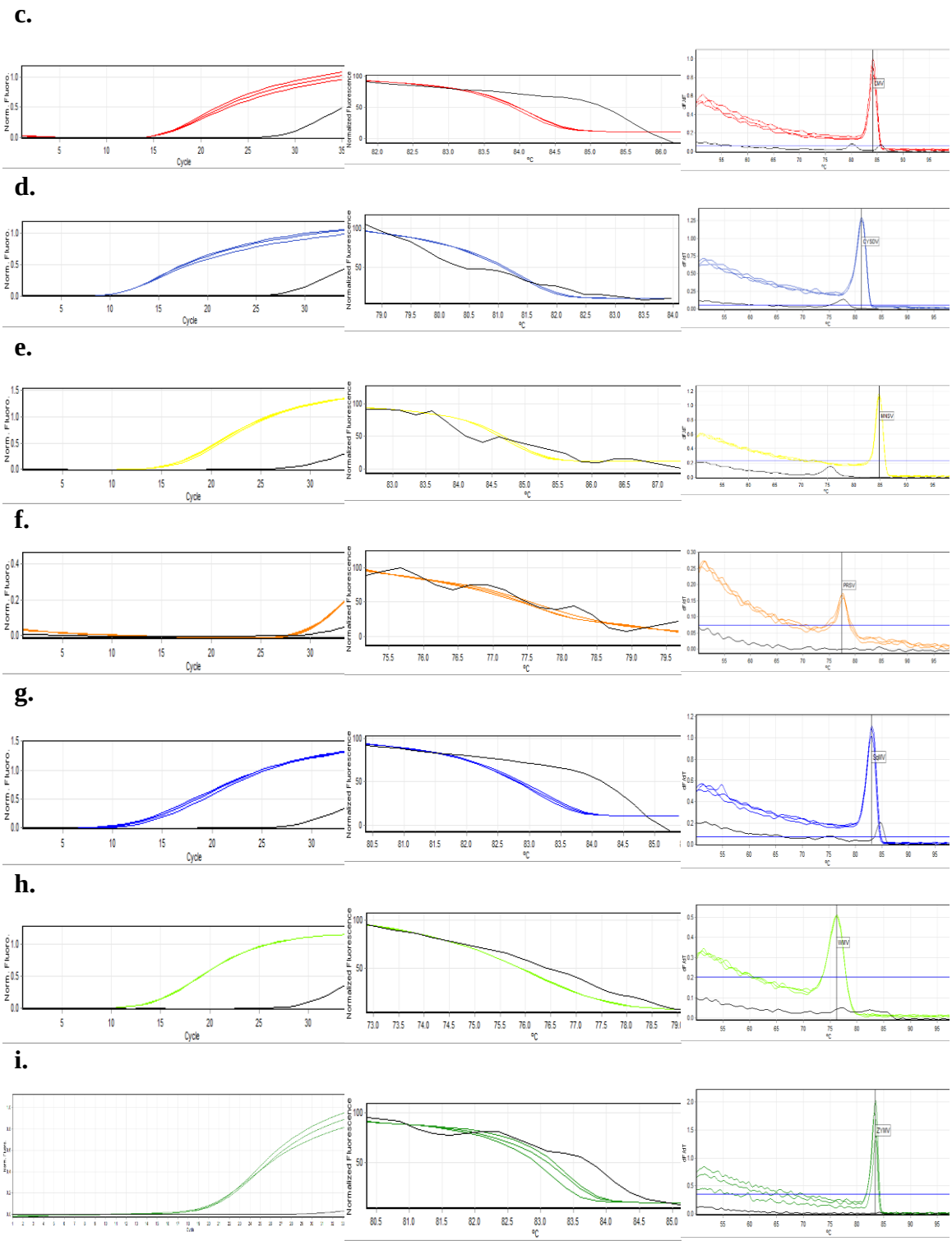


Figure 5. 4 Results of individual amplification RT-qPCR for cucurbit viruses

a. *Cucurbit aphid-borne yellows virus (CABYV)* **b.** *Cucurbit green mottle mosaic virus (CGMMV)*

c. *Cucumber mosaic virus (CMV)* **d.** *Cucurbit yellows stunting disorder virus (CYSDV)* **e.** *Melon*

necrotic spot virus (MNSV) **f.** *Papaya ringspot virus (PRSV)* **g.** *Squash mosaic virus (SqMV)* **h.** *Watermelon mosaic virus (WMV)* **i.** *Zucchini yellow mosaic virus (ZYMV)* **left the column.** RT-qPCR from the positive control. **central column.** Normalized fluorescence data derived from raw data plots showing the Tm breaking point of fluorescence at the Tm. **right column.** Low-Resolution Melt profile derivative plot (-dF/dT against T). The steepest slope is easily visualized as a melting peak.

3.6. Multiplex RT-qPCR HRM

Five multiplex and two triplex virus reactions were performed and are described in Table 5.7. All reference viral positive controls amplified the expected diagnostic targets using Multiplex RT-qPCR. Amplification was not detected in the NTC. This result confirmed that primer sets do not cross-react in multiplex reactions. A graphic representation of the multiplex selected controls is shown in Figure 5.5.

Table 5. 7 Multiplex discrimination panel design based on product Tm reaction for multiple detections of most common Cucurbit viruses.

REACTION	VIRUS Tm (ºC)
A.	WMV, CYSDV, SqMV, MNSV, CABYV
	76.36 81.2 83.1 84.94 87.8
B.	PRSV, CYSDV, CMV
	77.76 81.2 84.24
C.	CGMMV, ZYMV, CABYV
	78.35 83.82 87.8

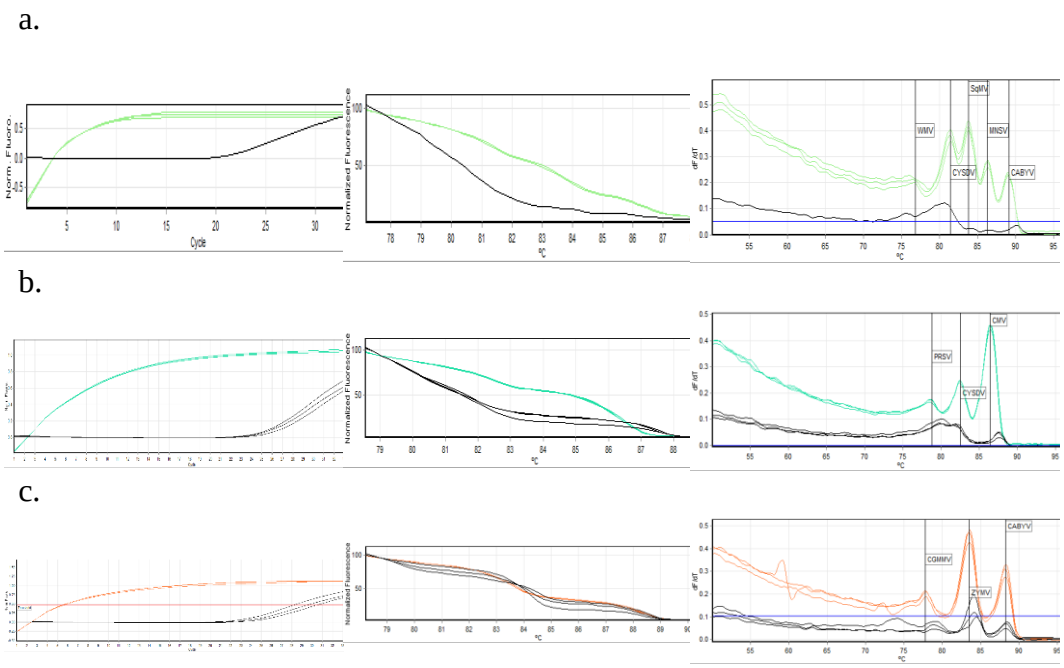


Figure 5. 5 Results of multiplex arrangement for multiple detections

a Shows Multiplex for five targets b. triplex B, and c. triplex C. **Left.** RT-qPCR amplification of the positive control. **Center.** Normalized fluorescence data derived from raw data plots showing the TM breaking point of fluorescence of the viruses at the Tm. **Right.** Low-Resolution Melt profile derivative plot (-dF/dT against T). The steepest slope is easily visualized as a melting peak.

3.7. Specificity and sensitivity assay

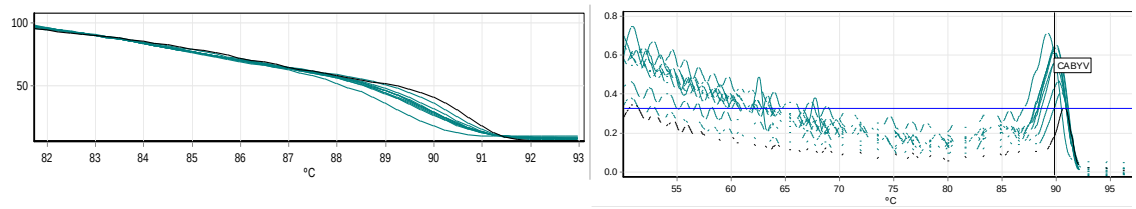
RT-qPCR HRM individual analysis of the nine virus-positive controls of the exclusive and inclusive panel did not amplify in cross-reaction. The positive control of each virus (1 ng/µl RNA transcript as well as the total RNA from the infected plant sample) produced a positive reaction, while the RNA from the healthy tissue as well as the NTC did not produce any reaction (Table 5.8).

Table 5. 8 Tested exclusivity panel for infecting Cucurbit viruses primer sets.

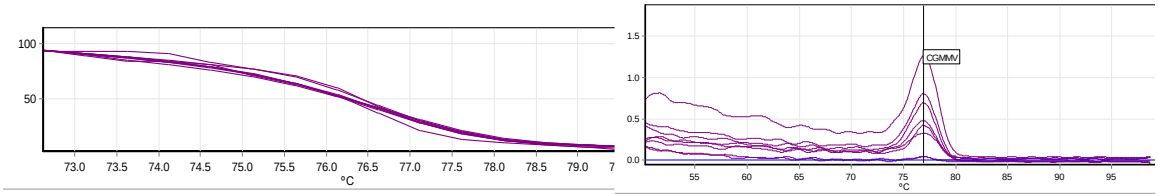
TARGET VIRUS	TAXONOMICALLY RELATED VIRUSES
Cucumber green mottle mosaic virus	*Kyuri green mottle mosaic virus*
(TOBAMOVIRUS)	
Cucumber mosaic virus (CUCUMOVIRUS)	*Peanut stunt virus*
Cucurbit aphid-borne yellows	*Beet western yellows virus*
virus (POLEROVIRUS)	
Papaya ringspot virus (POTYVIRUS)	*Zucchini yellow mosaic virus*
Watermelon mosaic virus (POTYVIRUS)	*Soybean mosaic virus*
Zucchini yellow mosaic virus (POTYVIRUS)	*Potato virus Y*
Squash mosaic virus (COMOVIRUS)	*Bean pod mottle virus*

The sensitivity assays of RT-qPCR-HRM allows an accurate detection down to femtograms. Individual sensitivity assays and the normalized fluorescence data derived from raw data plots with the Tm breaking point of fluorescence from 10 ng/µl down to 1 fg/µl as shown in Figure 5.6. The multiplex RT-qPCR HRM analysis Figure 5.7, indicated that the designed primer set could detect the diluted transcripts up to a concentration of 1 fg/µL in most of the cases. Similar trends were observed in the five multiplexes and two designed triplex. Sensitivity assay using virus RNA prepared by the direct virion-capture method also showed that the primer sets could detect the virus up to a concentration of 1 fg/µL. Figure 5.6 shows the sensitivity assays.

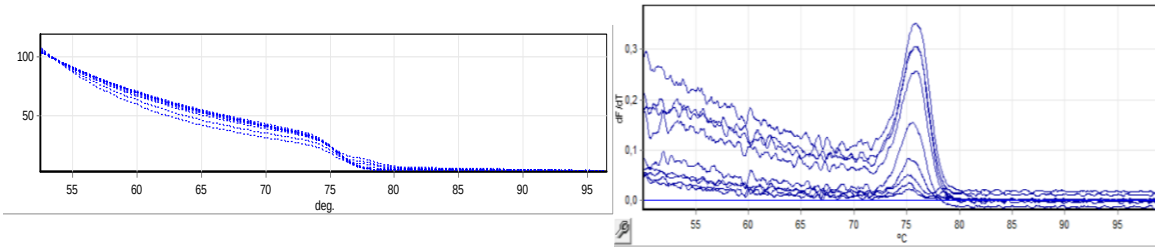
a.



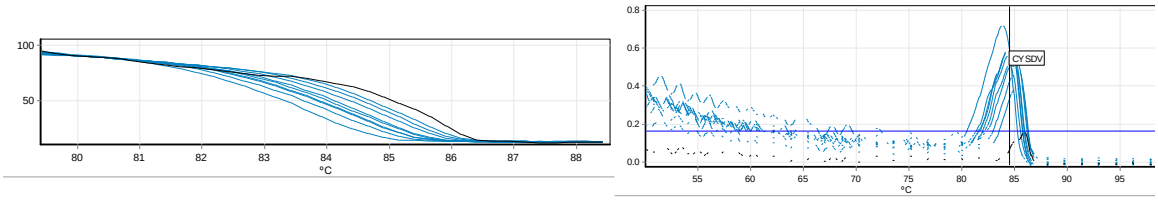
b.


100
50
73.0 73.5 74.0 74.5 75.0 75.5 76.0 76.5 77.0 77.5 78.0 78.5 79.0
°C
1.5
1.0
0.5
0.0
CGMMV
55 60 65 70 75 80 85 90 95
°C

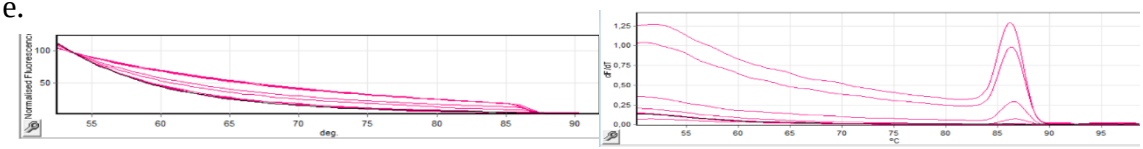
c.


100
50
55 60 65 70 75 80 85 90 95
deg.
dF /dT
0.3
0.2
0.1
0.0
55 60 65 70 75 80 85 90 95
°C

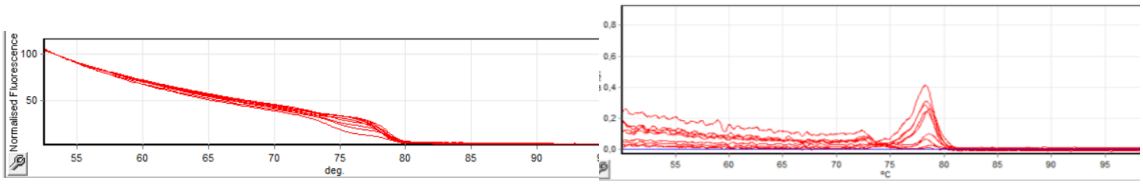
d.


100
50
80 81 82 83 84 85 86 87 88
°C
0.8
0.6
0.4
0.2
0.0
CY SDV
55 60 65 70 75 80 85 90 95
°C

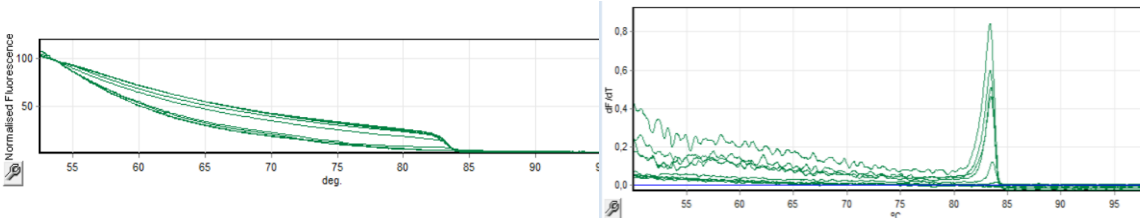
e.


Normalised Fluorescence
100
50
55 60 65 70 75 80 85 90
deg.
dF /dT
1,25
1,00
0,75
0,50
0,25
0,00
55 60 65 70 75 80 85 90 95
°C

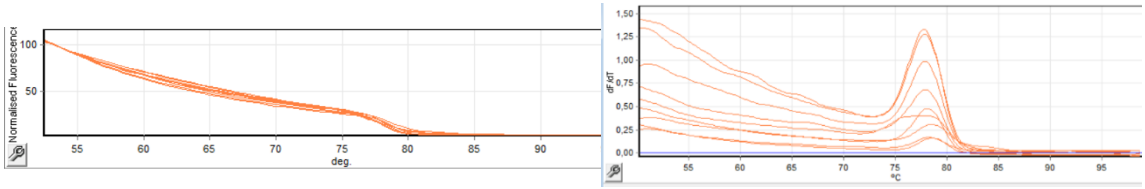
f.


Normalised Fluorescence
100
50
55 60 65 70 75 80 85 90
deg.
0,8
0,6
0,4
0,2
0,0
55 60 65 70 75 80 85 90 95
°C

g.


Normalised Fluorescence
100
50
55 60 65 70 75 80 85 90
deg.
dF /dT
0,8
0,6
0,4
0,2
0,0
55 60 65 70 75 80 85 90 95
°C

h.



i.

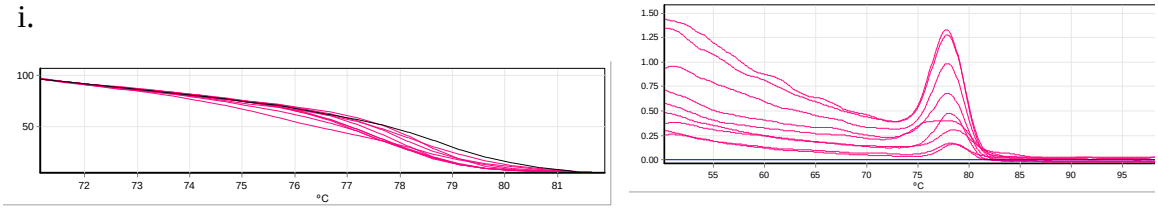
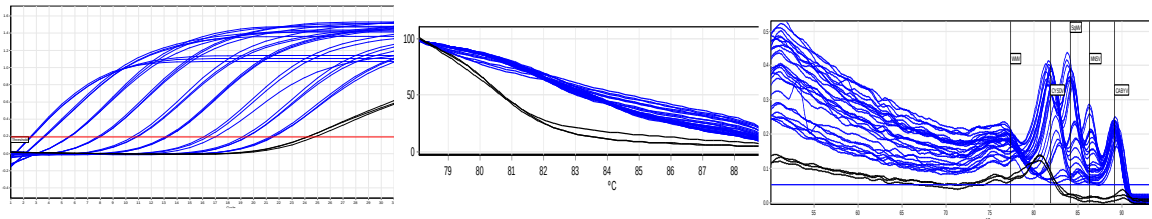


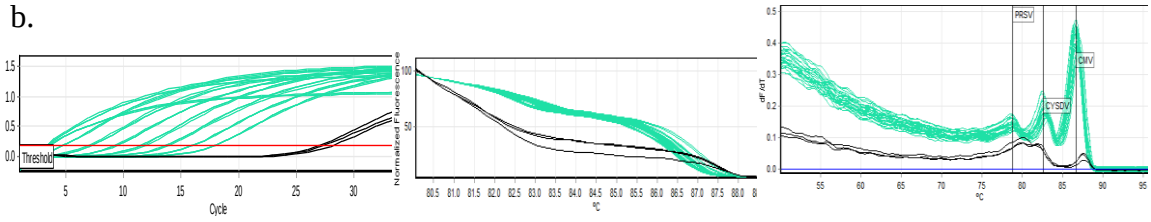
Figure 5. 6 Results of individual sensitivity amplification

a. *Cucurbit aphid-borne yellows virus (CABYV)* **b.** *Cucurbit green mottle mosaic virus (CGMMV)* **c.** *Cucumber mosaic virus (CMV)* **d.** *Cucurbit yellows stunting disorder virus (CYSDV)* **e.** *Melon necrotic spot virus (MNSV)* **f.** *Papaya ringspot virus (PRSV)* **g.** *Squash mosaic virus (SqMV)* **h.** *Watermelon mosaic virus (WMV)* **i.** *Zucchini yellow mosaic virus (ZYMV)* **left the column.** Normalized fluorescence data derived from raw data plots showing the TM breaking point of fluorescence at the Tm. **right column.** Low-Resolution Melt profile derivative plot (-dF/dT against T). The steepest slope is easily visualized as a melting peak.

a.



b.



c.

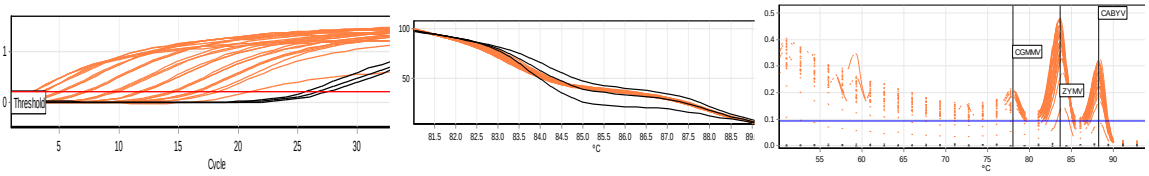


Figure 5. 7 Sensitivity results for each the three multiplex assays

a. WMV, CYSDV, SqMV, MNSV, CABYV. **b.** PRSV, CYSDV, CMV. **c.** CGMMV, ZYMV, CABYV. **Left.** RT-qPCR from positive controls. **Center.** Normalized fluorescence data derived from raw data plots showing the Tm breaking point of fluorescence of the viruses at the Tm. **Right.** Low-Resolution Melt profile derivative plot (-dF/dT against T). The steepest slope is easily visualized as a melting peak.

3.8. Screening of field samples

The Multiplex RT-qPCR-HRM analysis of the collected samples during the summer of 2016 and 2018 from Muskogee and Tulsa Counties, Oklahoma are described in Table 5.9. A serial dilution of the positive control (from 1 ng/µl down to 1fg of RNA transcript) produced a positive reaction in Multiplex RT-qPCR-HRM, while the RNA from the healthy tissue as well as the NTC did not produce any reaction.

Table 5. 9 Summarized results of RT-qPCR-HRM from field samples screening in the summer of 2016 - 2018 at Muskogee and Tulsa Counties, Oklahoma, USA.

County	Year	Virus	Positive	Negative
Muskogee	Summer 2016	CABYV	4	70
Muskogee	Summer 2016	CGMMV	52	22
Muskogee	Summer 2016	CMV	59	15
Muskogee	Summer 2016	CYSDV	5	69
Muskogee	Summer 2016	MNSV	33	41
Muskogee	Summer 2016	PRSV	9	65

Muskogee	Summer 2016	SqMV	43	31
Muskogee	Summer 2016	WMV	69	5
Muskogee	Summer 2016	ZYMV	70	4
Tulsa	Summer 2018	CABYV	11	4
Tulsa	Summer 2018	CGMMV	15	0
Tulsa	Summer 2018	CMV	15	0
Tulsa	Summer 2018	CYSDV	6	9
Tulsa	Summer 2018	MNSV	0	15
Tulsa	Summer 2018	PRSV	10	5
Tulsa	Summer 2018	SqMV	1	14
Tulsa	Summer 2018	WMV	15	0
Tulsa	Summer 2018	ZYMV	10	5

3.9. Validated samples before sequencing with qPCR.

ds-cDNA from the selected samples of summer 2018 tested positive as shown in Table 5.9 to qPCR. Table 5.10 compares the positive results using qPCR, EPI2ME, and EDNA-Cucurbits. The obtained results showed similar results at higher concentrations of the target virus, besides ZYMV in sample 10T and 12T were 7181 and 8513 reads respectively tested positive to the evaluated virus.

Table 5. 10 Comparison of virus detection among three virus detection methodologies: Multiplex RT-qPCR-HRM, EPI2ME, and EDNA-Cucurbits in this research and its positive results.

Sample	Host	Multiplex RT-qPCR-HRM	EPI2ME	EDNA-Cucurbits
2T	Pumpkin	CABYV-CGMMV-CMV-CYSDV-MNSV-PRSV-SqMV-WMV-ZYMV	PRSV-TRSV	PRSV-TRSV
4T	Pumpkin	CABYV-CGMMV-CMV-CYSDV-PRSV-SqMV-WMV-ZYMV	PRSV-TRSV-WMV	PRSV-TRSV-WMV

7T	Pumpkin	CABYV-CGMMV-CMV-CYSDV-PRSV-WMV-ZYMV	PRSV-TRSV-WMV-ZYMV	PRSV-TRSV-WMV
10T	Watermelon	CGMMV-CMV-CYSDV-PRSV-WMV-ZYMV	PRSV-ZYMV	PRSV-WMV
12T	Watermelon	CABYV-CGMMV-CMV-CYSDV-PRSV-WMV-ZYMV	CABYV-PRSV-TRSV-WMV-ZYMV	PRSV-WMV
15T	Watermelon	CGMMV-CYSDV-PRSV-SqMV-WMV-ZYMV	PRSV-TRSV-WMV-ZYMV	PRSV-WMV

Note: In orange virus consistently found in all detection methods tested.

The two selected hosts *Cucurbita pepo* (pumpkin) and *Citrullus lanatus* (watermelon), showed similar plant virus infection. In the three detection methods tested, there was a correlation in virus presence despite the genetic composition of the host. The common cucurbit viruses that were found by the detection methods were PRSV, TRSV, and WMV (Table 5.10).

4. Discussion

Specific E-probes were validated *in-vitro* for fifteen cucurbit infecting viruses. The studied cucurbit virome included 15 viruses infecting cucurbits was considet to build the EDNA-Cucurbits (MiFi™) E-probe set. Stobbe et.al. (2013), pioneered research on reliable plant pathogen detection using High Throughput Sequencing (HTS). EDNA research validation throughout the years found that randomness of sequencing technologies may play an important role in metagenomics analysis (Blagden et al., 2016; Espindola et al., 2015; Espindola et al., 2018). Therefore, it is important to determine the depth and coverage of each sequencing method to be able to detect plant viruses. The main objective of this study was to deliver accurate reliable virus diagnostic methods suitable for multiple pathogen detection. Using the EDNA bioinformatics pipeline allowed to develop E-probes for fifteen viruses infecting cucurbits *in-vitro* detection. The results of this first screening method clearly support the proof of concept that EDNA-Cucurbits (MiFi™) is a suitable and flexible database for plant pathogen diagnosis that can be adapted to detect any virome.

It is worth discussing the dissimilar results found in Table 5.10 were positive results found in Multiplex RTqPCR-HRM compared with EDNA-Cucurbits (MiFi™) and EPI2ME. Results showed that there are several variables to consider to avoid false negatives during diagnostics. First, the low quantity of pores was used for sequencing, even though Oxford Nanopore technologies delivers long reads, our samples have more reads from the host than the virus titer during infection, this can be a limitation for detection purposes. Second, there was cross-contamination with previous runs since there were identified reads belonging to *Aedes aegypti* (mosquito), previously sequenced in the same flow cell. One of the strategies to solve this issue may be treating the samples to deplete ribosomal RNA from the host and related bacteria. Similar studies using metagenomics reads removed environmental bacteria, as well as highly redundant RNA, reads from the host giving better performance to the plant virus-detection tool (Villamor et.al., 2019).

High Throughput Sequencing (HTS) outcome to be a powerful diagnostic tool used to establish and understand disease causality by a plant pathogen. A concern during research is not detecting viruses at low titers. By amplifying the dscDNA directly from samples increased significantly the presence of host and Poly A tail containing viruses while viruses lacking the Poly A tail have not been detected. A possible solution for this issue is amplifying the virus concentration before the library preparation using random hexamers or starting from small RNAs directly (Santala & Valkonen, 2018). Results of this research demonstrated that further research is needed, as well as increasing genome coverage and deepness in the selected sequencing methodology to detect lower titer of virus.

On the other hand, this study developed and validated a Multiplex RT-qPCR combined with HRM using positive controls and field samples, which can be used as a detection and discrimination method for nine viruses infecting cucurbits. The developed primer sets can be used individually or in multiplex RT-qPCR-HRM. In general, RT-qPCR detection is more sensitive, reliable and

accurate for detection and discrimination of the target if the primers are designed from conserved regions of a targeted gene (Arif & Ochoa-Corona, 2013).

For primer design purposes sequences from RdRp and coat protein genes were selected based on the presence of high conserved regions. Seeking the design of virus-specific primers to detect and discriminate against the selected targets. A conserved region of these genes allows reliable virus identification and detection (Tavazza, 2017). The selection of optimal thermodynamics parameters during primer design plays a significant role and are a key feature to achieve optimal PCR sensitivity. The designed multiplex method is able to detect femtogram per microliter of known targeted concentrations of cDNA and positive control. Efficient PCR amplification requires precise design and optimal oligonucleotide primers, high-quality PCR reagents, and optimal cycling conditions matching the thermodynamic features of the proposed oligos (Arif & Ochoa-Corona, (2013). Similarly, the development of oligos used for multiplex RT-qPCR-HRM, i.e. the design for this study, aimed to combine optimal thermodynamic requirements for high PCR primer efficiency, however, when designing primers for multiplex RT-qPCR-HRM special attention has to be given to product size (no larger than 350bp), GC% (lower than 50%) and differences in the melting temperature of the products. uMelt is a web application that allows to calculate and to predict melting curves *in-silico*, which are quite similar to the actual HRM obtained curves *in-vitro* (Arif, et. al., 2014). Therefore, we designed a multiplex RT-PCR amplifying different PCR products with +/- 2ºC of difference to facilitate discrimination (Farrar, 2017).

HRM has proved to be useful for accurate detection of somatic mutations in various genes (Wittwer, et.al., 2003). Recently, due to the HRM sensitivity, the method is also applied for diagnostics (Fadhil, et.al., 2010). Among the advantages of Multiplex RT-qPCR-HRM, obtained provides a robust analysis that allows accurate discrimination and quantification of viruses mix infecting reducing the number of reactions needed to screen a single sample against multiple virus targets. Simultaneous amplification facilitates virus detection from an asymptomatic host. The

detection and discrimination system described, demonstrated is possible to detect of up to five viruses in a single reaction, which is also useful when applied to microbial forensics (Figure 5.6).

The Tm of PCR products obtained during a serial dilution varied among concentrations of the target generating different high resolution melting profiles for each of the virus dilution (Figure 5.7) (Elkins et al., 2016). We hypothesized these differences are because although the primers bind to a conserved region, the sequence of the PCR product may contain single nucleotide polymorphism (SNPs) in the sample because viruses exist and are found in populations. This is supported by reports that shows that although primers are designed upon a conserved region as a template, viruses evolve rapidly creating variability in a population (Zhao, et. al., 2019). Also, since viruses exist in populations and are expected to be at different concentrations, PCR products harboring different SNPs may be plotting with different Tm along with the serial dilution (Figure 5.6). The most abundant SNPs within the population are expected to be the higher virus titer (Figure 5.7).

RNA isolation from infected plant tissue is often time-consuming even with the use of kits (MacKenzie, et.al., 1997). Direct virion capture on plastic was assessed and performed based on the similar isoelectric point of the capsid protein of all nine targeted viruses. The isoelectric point ranged between 4.62 and 4.78. The amino acid capsid protein phylogenetic Maximum Likelihood tree of the studied viruses allowed to regulate a broad buffer-pH range used during virion capture. The phylogenetic Maximum Likelihood tree and the theoretical isoelectric point of the studied viruses coat proteins are shown in Figure 5.3. The virus trapping was consistently successful using a neutral pH of 7.4. The virion coat protein increases its ionic strength by influence environmental pH (Vega et al., 2014). Direct virion trapping to plastic PCR tube is possible through a newly developed easy protocol, which is of low cost and practical.

Acute plant viral infection results in severe symptom development, in contrast, a persistent interaction of the virus with the host may result in mild symptoms. Symptoms are produced as a

reaction of plant-virus interaction when virus titers are high enough to become systemic in the host (Roossinck, 2010). Moreover, the symptoms can vary and have similarities when are combined as a result of multiple viral infections. Interestingly, during multiple infections plant viruses are able to recombine and reorganize leading to systemic movement through a host family (Islam et al., 2017). Our results showed field symptomatic samples with more than four infecting viruses from different genera. However, is difficult to predict by visual examination the causal agent and the specific virus-producing symptoms. This is why is important to rely on a detection system that can offer an accurate diagnosis to develop an integrated pest management plan (Ghoshal & Sanfaçon, 2015).

Plant disease epidemics in Cucurbit crops depends on an interaction of a susceptible host, a virus belonging to the Cucurbit virome, and an arid or semi-arid environment over the summer (Ram et al., 2019). The found higher prevalence of *Potyviruses* (PRSV, WMV, and ZYMV) in mixed infections showed in all tested detection tools point toward the existence of a synergistic interaction within this genus over the sampled time frame. Consequently, multiple infections lead to the establishment of cyclic epidemics over the years. Virus fitness boost during coinfection by influencing virus co-evolution with the host (Miralles et al., 2001). Infection cycle relays on plant virus dispersion by a vector, often an insect. Feeding preference of the insects in a host establishes the spread process of the virus on a specific host evolving to epidemics (Bak et al., 2017). Multiple infections depend on an adaptation of the viromes to a host. During the sampling, we observed abundant symptom development due to an increase of replication between viruses of the same genus leading to a high economic impact on the grower (Moreno & Lopez-Moya, 2020).

Advances in molecular and genomic technologies are being applied as a resource to ensure the sustainable production of cucurbits (Grumet et al., 2017). Integrated pest management, biosecurity, quarantine and routine diagnostics of viruses requires an accurate, reliable and sensitive discriminatory method (Dobhal et al., 2014). Therefore, there is a need for a rapid screening method

for cucurbit infecting viruses due to the increase in the incidence of disease during the past decade (Ali, 2012). The primer sets developed in this study can be combined in multiplex reactions that allow accurate screening of samples. The use of this method will contribute to a local and governmental plant diagnostic clinics as a routine diagnosis to the most common cucurbit viruses in Oklahoma and nationwide.

Cucurbit growers, Oklahoma County extension agents and diagnostic clinics with the National plant diagnostic network (NPDN) will directly benefit from this newly developed detection and diagnostic method. Similarly, cucurbit breeders will be able to track resistance to these viruses in new progenies with extended sensitivity, specificity, and broad range (nine viruses). Consistent and reliable multiple detection methods compatible with cucurbit virome are reported. This method can be applied as a routine diagnostic strategy for multiple viral detection methods. The method of application has to be simple, sensitive, reproducible and has application in detection and discrimination. With necessary adjustments to the detection method at *in-vitro* settings, EDNA-Cucurbits (MiFi™) may be applicable for epidemiological studies, biosecurity, microbial forensics, as well as a screening of the asymptomatic plants in breeding programs.

References

Aguilar, J. M., Franco, M., Marco, C. F., Berdiales, B., Rodriguez Cerezo, E., Truniger, V. & Aranda, M. A. (2003). Further variability within the genus Crinivirus, as revealed by determination of the complete RNA genome sequence of *Cucurbit yellows stunting disorder virus. J Gen Virol 84*, 2555–2564.

Ali, A., Mohammad, O., & Khattab. A. (2012). Distribution of viruses infection cucurbit crops and isolation of potential new virus-like sequences from weeds in Oklahoma. *Plant Disease, 96*, 243-248.

Ali, A., & Abdalla. O. (2012). The occurrence of viruses infecting watermelon, other cucurbits, and weeds in the parts of Southern United States. *Plant Health Progress, 10*, 284-290.

Arif, M., & Ochoa-Corona, F. M. (2013). Comparative assessment of 5′ A/T-rich overhang sequences with optimal and sub-optimal primers to increase PCR yields and sensitivity. *Molecular biotechnology*, *55*(1), 17-26.

Arif, M., Aguilar-Moreno, G. S., Wayadande, A., Fletcher, J., & Ochoa-Corona, F. M. (2014). Primer modification improves rapid and sensitive in vitro and field-deployable assays for the detection of high plains virus variants. *Applied and environmental microbiology*, *80*(1), 320-327.

Babu, B., Jeyaprakash, A., Jones, D., Schubert, T. S., Baker, C., Washburn, B. K., ... & Paret, M. L. (2016). Development of a rapid, sensitive TaqMan real-time RT-PCR assay for the detection of Rose rosette virus using multiple gene targets. *Journal of virological methods*, *235*, 41-50.

Bak, A., Cheung, A. L., Yang, C., Whitham, S. A., & Casteel, C. L. (2017). A viral protease relocalizes in the presence of the vector to promote vector performance. *Nature communications*, *8*, 14493.

Blanchard, C., Higgins, T., Anderson, B. (1997). RNAs 4A and 5 are present in tomato aspermy virus and both subgroups of cucumber mosaic virus. *Arch. Virol. 142*, 1273–1283.

Blua, M. J., & Perring, T. M. (1989). Effect of zucchini yellows mosaic virus on development and yield of cantaloupe (Cucumis melo). *Plant Dis*, *73*(4), 317-320.

Caasi, D. R. J., Arif, M., Payton, M., Melcher, U., Winder, L., & Ochoa-Corona, F. M. (2013). A multi-target, non-infectious and clonable artificial positive control for routine PCR-based assays. *Journal of microbiological methods*, *95*(2), 229-234.

Dwight, Z., Palais, R., & Wittwer, C. (2011) uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics. 27*(7), 1019 – 1020.

Dobhal, S., Arif, M., Olson, J., Mendoza-Yerbafría, A., Aguilar-Moreno, S., Pérez-García, M., & Ochoa-Corona, F. (2014). Sensitive detection and discrimination method for studying multiple infections of five major plant viruses infecting ornamental plants in nursery environments. *Annals of Applied Biology*, *20*, 286-296.

Doolittle. (1916). *Phytopathology, 6*, 145.

Dombrovsky, A., Tran-Nguyen, L., & Jones, R. (2017). Cucumber green mottle mosaic virus: Rapidly Increasing Global Distribution, Etiology, Epidemiology, and Management. *Annual review of phytopathology*.

Elnifro E.M., Ashshi A.M., Cooper R.J., & Klapper P.E. (2002). Multiplex PCR: optimization and application in diagnostic virology. *Clinical Microbiology Reviews*, *13*, 559–570.

Elkins, K., Perez, A., & Sweetin, K. (2016). Rapid and inexpensive species differentiation using a multiplex real-time polymerase chain reaction high-resolution melt assay. *Analytical Biochemistry*, *500*, 15–17.

Ensminger, A., Ensminger, M., Konlande, J., Robson, J. (1983) Foods and nutrition encyclopedia, vol. 1. Clovis: Pergus Press.

Espindola, A. S., Schneider, W., Cardwell, K. F., Carrillo, Y., Hoyt, P. R., Marek, S. M., Garzon, C. D. J. P. o. (2018). Inferring the presence of aflatoxin-producing Aspergillus flavus strains using RNA sequencing and electronic probes as a transcriptomic screening tool. *13*(10), e0198575.

Fadhil, W., Ibrahem, S., Seth, R., & Ilyas, M. (2010). Quick-multiplex-consensus (QMC)-PCR followed by high-resolution melting: a simple and robust method for mutation detection in formalin-fixed paraffin-embedded tissue. *Journal of clinical pathology*, *63*(2), 134-140.

Farrar, J. S., & Wittwer, C. T. (2017). High-resolution melting curve analysis for molecular diagnostics. In *Molecular Diagnostics (Third Edition)* (pp. 79-102).

Fletcher, J. D., Wallace, A. R., & Rogers, B. T. (2000). Potyviruses in New Zealand buttercup squash (Cucurbits maxima Duch.): Yield and quality effects of ZYMV and WMV 2 virus infections. *New Zealand Journal of Crop and Horticultural Science*, *28*(1), 17-26.

Finetti Sialer, M.M., Cillo, F., Paradies, F., Gallitelli, D. (1999). Biodiversity of populations of cucumber mosaic virus in Italy before and after virus outbreaks in 1988. Abstracts of papers presented at the VII the International Plant Virus Epidemiology Symposium, Aguadulce (Almeria), Spain, April 11–16, 1999, pp. 52–53.

Fraser, R. S. S. (1992). The genetics of plant-virus interactions: implications for plant breeding. *Euphytica*, *63*(1-2), 175-185.

Freitag, J. H. (1956). Beetle Transmission, Host Range, and Properties of Squash Mosaic Virus. *Phytopathology*, *46*(2).

Ghoshal, B., & Sanfaçon, H. (2015). Symptom recovery in virus-infected plants: revisiting the role of RNA silencing mechanisms. *Virology*, *479*, 167-179.

Gibbs, A. J., Hajizadeh, M., Ohshima, K., & Jones, R. A. (2020). The Potyviruses: An Evolutionary Synthesis Is Emerging. *Viruses*, *12*(2), 132.

Gonsalves, D. (1998). Control of papaya ringspot virus in papaya: a case study. *Annual review of phytopathology*, *36*(1), 415-437.

Gosalvez, B., Navarro, J. A., Lorca, A., Botella, F., Sanchez-Pina, M. A., & Pallas, V. (2003). Detection of Melon necrotic spot virus in water samples and melon plants by molecular methods. *Journal of virological methods*, *113*(2), 87-93.

Grieco, F., Cillo, F., Barbarossa, L., Gallitelli, D. (1992). Nucleotide sequence of a cucumber mosaic virus satellite RNA associated with a tomato top stunting. *Nucleic Acids Res. 24*, 6733.

Guilley, H., Wipf-Scheibel, C., Richards, K., Lecoq, H., & Jonard, G. (1994). Nucleotide sequence of cucurbit aphid-borne yellows luteovirus. *Virology*, *202*(2), 1012-1017.

Grumet, R., Katzir, N., & Garcia-Mas, J. (Eds.). (2017). *Genetics and genomics of Cucurbitaceae*. Springer International Publishing.

Hibi, T., & Furuki, I. (1985). Melon necrotic spot virus. *CMI/AAB descriptions of plant viruses*, *302*.

Islam, W., Zaynab, M., Qasim, M., & Wu, Z. (2017). Plant-virus interactions: disease resistance in focus. *Hosts and Viruses*, *4*(1), 5.

James, D. (1999). A simple and reliable protocol for the detection of apple stem grooving virus by RT–PCR and in a multiplex PCR assay. *Journal of Virological Methods*, *83*(1), 1-9.

Jelkmann, W., & Keim-Konrad, R. (1997). Immuno-capture Polymerase Chain Reaction and Plate-Trapped ELISA for the Detection of Apple Stem Pitting Virus. *Journal of Phytopathology*, *145*(11-12), 499-503.

Keinath, A. P., Wintermantel, W. M., & Zitter, T. A. (Eds.). (2017). *Compendium of Cucurbit Diseases and Pests*. St. Paul, MN: APS Press.

Kishi. (1966). *Ann. phytopath. Soc. Japan 32*: 138.

Koonin, E.V. (1991). The phylogeny of RNA-dependent RNA-polymerases of positive-strand RNA viruses. *J. Gen. Virol. 72*:2197–206.

Lecoq, H., Bourdin, D., Wipf-Scheibel, C., Bon, M., Lot, H., Lemaire, O., & Herrbach, E. (1992). A new yellowing disease of cucurbits caused by a luteovirus, cucurbit aphid- borne yellows virus. *Plant Pathology*, *41*(6), 749-761.

Lee, M., Kubota, C., Tsao, J., Bie, L., Echevarria, P., Morra, L., Oda, M. (2010). Current status of vegetable grafting: diffusion, grafting techniques, automation. *Scientia Hort*. *127*, 93–105.

Li, H., Lucy, A., Guo, H., Li, W., Ji, L., Wong, S., Ding, S., 1999. Strong resistance targeted against a viral suppressor of the plant gene silencing defense mechanism. *EMBO J. 18*, 2683–2691.

Li, Z., Zhang, Z., Yan, P., Huang, S., Fei, Z., & Lin, K. (2011). RNA-Seq improves the annotation of protein-coding genes in the cucumber genome. *BMC genomics*, *12*(1), 540.

Lindberg, G. D., Hall, D. H., & Walker, J. C. (1956). A study of Melon and Squash viruses. *Phytopathology*, *46*(9).

Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., & Wittwer, C. (2004). Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical chemistry*, *50*(7), 1156-1164.

Ng J., Perry K.L. (1999) Stability of the aphid transmission phenotype in Cucumber mosaic virus. *Plant Pathology*, 48, 388–394.

Nolasco, G., De Blas, C., Torres, V., & Ponz, F. (1993). A method combining immunocapture and PCR amplification in a microtiter plate for the detection of plant viruses and subviral pathogens. *Journal of Virological Methods*, *45*(2), 201-218.

MacKenzie, D. J., McLean, M. A., Mukerji, S., & Green, M. (1997). Improved RNA extraction from woody plants for the detection of viral pathogens by a reverse transcription-polymerase chain reaction. *Plant disease*, *81*(2), 222-226.

McCreight, J. D. (2016). Cultivation and uses of cucurbits. Plant Genetics and Genomics: Crops and Models pp 1-12.

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., Škorić, D. J. F. i. m. (2017). A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. *8*, 45.

Mascia, T., & Gallitelli, D. (2016). Synergies and antagonisms in virus interactions. *Plant Science*, *252*, 176-192.

Miralles, R., Ferrer, R., Sole, R. V., Moya, A., & Elena, S. F. (2001). Multiple infection dynamics have pronounced effects on the fitness of RNA viruses. *Journal of Evolutionary Biology*, *14*(4), 654-662.

Moreno, A. B., & López-Moya, J. J. (2020). When viruses play team sports: Mixed infections in plants. *Phytopathology*, *110*(1), 29-48.

Pasin, F., Menzel, W., & Daròs, J. A. (2019). Harnessed viruses in the age of metagenomics and synthetic biology: an update on infectious clone assembly and biotechnologies of plant viruses. *Plant biotechnology journal*, *17*(6), 1010-1026.

Okada, Y. (1986). Cucumber green mottle mosaic virus. In *The plant viruses*. pp. 267-281. Springer US.

Ram, C., Berwal, M. K., Gora, J. S., Kumar, R., Haldhar, S. M., Gurjar, K., & Singh, D. (2019). Genomic and biotechnological interventions for crop improvement in cucurbitaceous crops: A review.

Reingold V, Lachman O, Belausov E, Koren A, Mor N, Dombrovsky A. (2016). Epidemiological study of Cucumber green mottle mosaic virus in greenhouses enables the reduction of disease damage in cucurbit production. *Ann. Appl. Biol. 168*, 29–40

Roossinck, M. J. (2010). Lifestyles of plant viruses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1548), 1899-1905.

Rowhani, A., Maningas, M. A., Lile, L. S., Daubert, S. D., & Golino, D. A. (1995). Development of a detection system for viruses of woody plants based on PCR analysis of immobilized virions. *Phytopathology*, *85*(3), 347-352.

Rubio, L., Soong, J., Kao, J., & Falk, B. W. (1999). Geographic distribution and molecular variation of isolates of three whitefly-borne closteroviruses of cucurbits: Lettuce infectious yellows virus, Cucurbit yellows stunting disorder virus, and Beet pseudo-yellows virus. *Phytopathology*, *89*(8), 707-711.

Santala, J., & Valkonen, J. (2018). The sensitivity of Small RNA–Based Detection of Plant Viruses. *Frontiers in microbiology*, *9*, 939.

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., . . . Schneider, W. J. J. o. m. m. (2013). E-probe Diagnostic Nucleic acid Analysis (EDNA): a theoretical approach for handling of next-generation sequencing data for diagnostics. 94(3), 356-366.

Schaefer, H., & Renner, S. S. (2011). Phylogenetic relationships in the order Cucurbitales and a new classification of the gourd family (Cucurbitaceae). *Taxon*, *60*(1), 122-138.

Tavazza, M., Lucioli, A., & Ilardi, V. (2017). Gene Silencing Provides Efficient Protection against Plant Viruses. *Plant Gene Silencing: Mechanisms and Applications*, *5*, 193.

Tricoll, D. M., Carney, K. J., Russell, P. F., McMaster, J. R., Groff, D. W., Hadden, K. C., & Quemada, H. D. (1995). Field evaluation of transgenic squash containing single or multiple virus coat protein gene constructs for resistance to cucumber mosaic virus, watermelon mosaic virus 2, and zucchini yellows mosaic virus. *Nature Biotechnology*, *13*(12), 1458-1465.

United States Department of Agriculture. (2016). *Vegetables 2016 summary* [National Agricultural Statistics Service]. Retrieved from http://www.nass.usda.gov

United States Department of Agriculture. (2019). *National watermelon report* [National Agricultural Statistics Service]. Retrieved from https://usda.library.cornell.edu

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), 115.

Vafaei, H., & Mahmoodi, M. (2017). Presence of Recombinant Strain of Cucurbit Aphid Borne Yellows Virus (CABYV) in Iran. *Iranian Journal of Biotechnology*.

Vega-Acosta, J. R., Cadena-Nava, R. D., Gelbart, W. M., Knobler, C. M., & Ruiz-García, J. (2014). Electrophoretic mobilities of a viral capsid, its capsid protein, and its relation to viral assembly. *The Journal of Physical Chemistry B*, *118*(8), 1984-1989.

Verchot, J. (2011). Wrapping membranes around plant virus infection. *Current opinion in virology*, *1*(5), 388-395.

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., & Tzanetakis, I. E. (2019). High Throughput Sequencing For Plant Virus Detection and Discovery. *Phytopathology*, *109*(5), 716-725.

Villinger, J., Mbaya, M., Ouso, D., Kipanga, P., Lutomiah, L., & Masiga. D. (2016). Arbovirus and insect-specific virus discovery in Kenya by novel six genera multiplex high-resolution melting analysis. *Molecular Ecology Resources*, *10*, 1-15.

Wang, A. (2015). Dissecting the molecular network of virus-plant interactions: the complex roles of host factors. *Annual review of phytopathology*, *53*, 45-66.

Wintermantel, W.M., Gilbertson, R.L., McCreight, J.D., Natwick, E.T. 2016. Host-specific relationship between virus titer and whitefly transmission of Cucurbit yellows stunting disorder virus. *Plant Dis*. 1119.

Wittwer, C., Reed, G., Gundry, C., Vandersteen, J., & Pryor. R. (2003). High-resolution genotyping by amplicon melting analysis using LCgreen. *Clinical Chemistry, 49*(6), 853-860.

Zhao, L., Seth-Pasricha, M., Stemate, D., Crespo-Bellido, A., Gagnon, J., Draghi, J., & Duffy, S. (2019). Existing host range mutations constrain the further emergence of RNA viruses. *Journal of virology*, *93*(4), e01385-18.

CHAPTER VI

ASSESSING FILTRATION OF WATERBORNE PLANT VIRUSES in-vitro TO FIELD.

Abstract

Water-borne-viruses can spread long-distance through irrigation water sources. Agriculture and greenhouse production may be affected because of contaminated water supplies. Water-borne viruses have been reported to infect crop production worldwide and several of these plant viruses are reported in the United States. Early viral infections may not be diagnosed until symptoms develop. High throughput sequencing (HTS) can be used as a detection and diagnostic tool for water sources because of its ability to amplify multiple unique genomic signatures of plant viruses. Metagenomic analysis of HTS has been used as a microorganism discovery tool facilitating the characterization of novel microbial communities in diverse ecosystems. Pathogen diagnosis and discovery using HTS pipelines is a time-consuming process. Electronic probe Diagnostic Nucleic acid Analysis (EDNA) was used to detect water-borne viruses *in-silico* simulating metagenomic runs containing reference positive controls and the host genome. This approach was validated after experiments using *in-vitro* and field filtration systems. Although the type of filter used may bring low efficiency to virus filtration, EDNA-Water accurately detected plant viruses in water samples. EDNA Water MiFi^{TM} is a valuable screening method since allows rough metagenomes screening of three genera (*Potexvirus*, *Tombusvirus*, and *Tobamovirus*) of reported as water-borne plant viruses that may be detected in irrigation water.

1. Introduction

A relevant goal in aquatic ecology research focuses on plant-pathogen mitigation since virus-polluted irrigation water may carry primary virus inoculum which rapidly distributes toward agricultural, hydroponic, and greenhouse production systems. Research results demonstrated that two plant oomycetes (Bush, 2002: Wakeham, & Pettitt, 2017; Redekar et.al., 2019), 27 genera of fungi (Hong & Moorman, 2005; Dixon, 2015), eight species of bacteria (Lamichhane & Bartoli, 2015), 13 species of plant-parasitic nematodes (Brye, et.al., 2018) and seven genera of viruses are water-borne (Mehler & Ravnikar, 2012). The presence, movement, and spread of water-borne plant pathogens accelerate and augment the impact of an epidemic outbreak (Stewart-Wade, 2011).

This study selected three genera of plant water-borne viruses, *Potexvirus, Tombusvirus,* and *Tobamovirus,* like virus models. *Tobamovirus* is a stable and widespread RNA virus particle (Figure 6.1). Infectious virus particles have been reported even in abiotic reservoirs as clouds, fog, and in 14,000-year-old glacial ice subcores in Greenland (Castello et.al. 1995; Castello et.al. 1999). *Potexvirus* and *Tombusvirus* have been isolated previously from ditches, streams and recovered seawater, lakes, and rivers (Culley et.al., 2006; Djikeng et.al., 2009). These three genera were selected due to their biological and phytopathological significance as plant viruses as well as the stability of their viral particles.

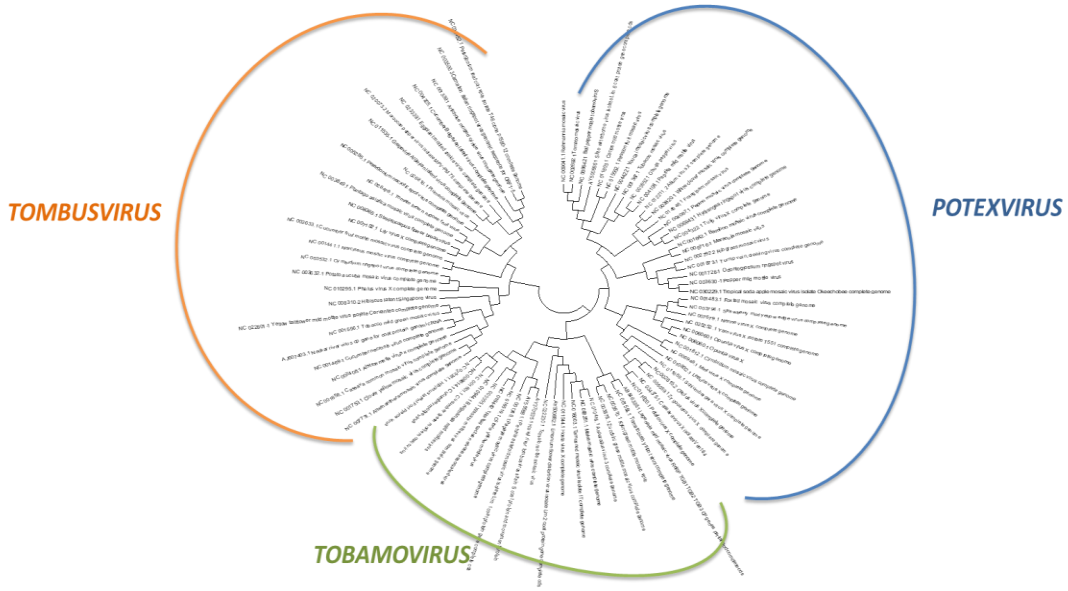


Figure 6. 1 Neighbor-joining phylogenetic tree (bootstrap 1000) of three genera of water-borne plant viruses, Potexvirus, Tobamovirus, and Tombusvirus selected for EDNA-Water MiFiTM database searches.

Pepino mosaic virus (PepMV) belongs to the genus *Potexvirus*. PepMV is an RNA positive sense single-strand viral particle. The virus particle is a flexible filament of 510 x 12.5 nm. PepMV has a narrow known host range. PepMV caused symptoms including mosaic and chlorosis in apical and basal leaves of tomato plants, also, fruit cosmetic damage and reduced the fruit quality and yield in *Solanaceous* plants. PepMV stability and high transmissibility allows it to be widespread mechanically. The virus is also seed-borne, insect-borne and transported by water systems (Hanssen & Thomma, 2010). PepMV was selected as a model virus for *in-vitro* water persistance studies.

Plant pathogen detection and their monitoring in water systems are very important when the presence of pathogens is suspected. Detection and biological determination of thresholds allow an unbiased assessment wheter if the pathogen is in a concentration considered a treat to the crop (Hong & Moorman, 2005).

Plant pathogen presence in irrigation water can be identified by symptom distribution in the crop and by the number of propagules infecting the system. The incidence and distribution patterns of infected plants differ from each crop, however, it is common to observe patches of symptomatic plants near to an irrigation outlet or all the way down running water lines. Followed by an increasing numbers of symptomatic plants will increase to the plants close to the infected plant since viruses can enter the plant through the roots (Sarra, 2005). Nucleic-acid based analysis may corroborate the association between an infected crop with the presence of the water-borne plant pathogen in water sources, particulates, and sediments (Hong & Moorman, 2005).

Plant virus presence and persistence in environmental and irrigation water may be the result of multiple contamination pathways including infected root/leaves or decaying organic material, infective virions released from animal sources as survivors of the alimentary tract, surface wash-off of decaying infected plant residues, virus transmitted by nematodes, seeds, and fungal resting structures (Mehler & Ravnikar, 2012). In greenhouse production systems, transmission of plant viruses occurs by vector (insect, fungi, nematodes) or without vectors through the roots. Plant virus infection rates and symptom severity depend on the viral concentration in water. Research demonstrated that water can become the primary inoculum with repeated inoculation in a single season (Hong & Moorman, 2005).

Rates of infective virion persistence in irrigation water depend on the structure of the viral particle and the environmental conditions. In recirculating nutrient solutions of greenhouse production systems, the persistence rate of infectious viral particles can be up to six months in *Tomato mosaic virus* (ToMV) and *Pepper mild mottle virus* (PMMoV) (Pares et.al., 1992). Non-traditional irrigation water sources are the most susceptible to viral contamination due to their chemical and microbiological complexity if compared with groundwater systems (Anderson-Coughlin & Kniel, 2019).

Risk assessment of plant virus potentially contaminated irrigation water is needed in order to develop an attainable and consistent detection method. However, plant virus detection in irrigation water is challenging. This research aims to determine differences in virus trapping when filtering *in-vitro* and in field settings. The virus adsorption-elution, VIRADEL, method proposed by the U.S. Environmental Protection Agency (EPA) set the basis for research focusing on viruses contaminating water (Cashdollar & Wymer, 2013). The VIRADEL method includes a water filtration system to get viruses get attached to the surface of the filter. After filtration, viruses are eluted from the filtering matrix. Virus recovery from the filtration systems takes into consideration multiple variables such as sample volume, physiochemical features of the water, virus concentration and environmental conditions of the system (Hong & Moorman, 2005). Under natural conditions, there is a stable equilibrium of adsorption, desorption, and re-adsorption of viruses depending on pH, salts presence, and the temperature of the environment (Koenig, 1986). Theoretically, water filtration system may achieve a 100% efficiency is expected when trapping plant waterborne viruses.

Viral metagenomics has been associated along the water cycle and plays an important role at the microbial diversity and distribution of the environment (Cassman et.al., 2012). Viral particles released from infected decaying plant material to irrigation water systems can cause disease. For instance, PepMV present in recirculating hydroponic systems is able to infect healthy plants, which develop symptoms at their shoots after 10 days. PepMV can cause up to 100% yield loss in tomato plants (Alfaro-Fernández, et.al. 2010). Consequently, there is a need for monitoring irrigation water to avoid or control plant virus infection.

Diagnostic assays for monitoring environmental waters must be highly specific, sensitive, rapid to develop, and low cost. In order to develop a suitable detection system, water samples must be concentrated. Even though there are widely used methods for plant virus detection in water systems using immunological as well as nucleic acid-based methods (Clark & Adams, 1977;

Candresse, et.al., 1998;), there is a lack of a sensitive and effective detection method that enables detection of multiple plant water-borne viruses, meaning single-step multiple detection systems of plant pathogens in environmental water.

Metagenomic plant virus diagnosis from environmental water can be achieved by the powerful use of High throughput sequencing (HTS). HTS has the ability to process millions of sequencing reads and can be applied as a universal detection tool in plant virology. Electronic Diagnosis Nucleic acid Analysis (EDNA) is a bioinformatics method developed to analyze high throughput sequence data rapidly, with minimal post-processing efforts to determine targeted plant virus presence in a sample. Electronic probes (E-probes) are genomic fingerprints designed, validated and uploaded to MiFi^{TM} server (Stobbe et al., 2013; Espindola et al., 2015; Blagden et al., 2016). This study seeks to develop EDNA-Water MiFi^{TM} databases that can detect the distribution of the plant viruses associated with irrigation water sources.

The specific objectives of this study are:

1. To assess the water-borne plant virus presence in a recirculation water system.

2. To develop and validate a water filtration system suitable for waterborne plant virus detection both *in*-and in the field.

3. To quantify waterborne virus titer of *Potexvirus*, *Tombusvirus,* and *Tobamovirus* using RT-qPCR and High throughput sequencing (MinION^{TM}).

2. Materials and methods

in-vitro assay

2.1 Plant material

Seeds of tomato (*Solanum lycopersicum* c.v. Rutgers) were germinated in seedling trays. Five days after germination, seedlings with visible roots were transferred from petri dishes to soil mixture in small pots and maintained in a growth chamber adjusted to daily cycles of 14 hours illumination of 250 μmol/m2/s at 25°C, and 10 hours of darkness at 20°C. Tomato plants were mechanically inoculated with a local isolate of PepMV, used as reference positive controls when they were two weeks old. Three weeks post-inoculation, symptomatic tissue was collected and stored at -80°C. Eight collections were done during eight months. Healthy controls were also planted under the same growth conditions in a separated growth chamber.

2.2 Methods used before water filtration

RNA extraction and quality

RNA extractions were done using three RNA extraction methods (for comparison): Trizol®, Modified Trizol plus precipitation of 8M LiCl (Wang et.al., 2009), and Qiagen RNeasy® plant mini kit (QIAGEN, Hilden, Germany) viral quantification was made from 100 mg of infected plant tissue and from healthy control. The RNA quality was measured using Nanodrop 1000 (ThermoFisher Scientific, Waltham, MA) and its integrity was visualized in 1% agarose electrophoresis gel. Samples showing both ribosomal units 18S and 28S were selected for further cDNA synthesis and RT-qPCR analysis. cDNA synthesis was performed using recombinant Moloney murine leukemia virus (M-MLV) (New England Biolabs, Ipswich, MA) reverse transcriptase following the manufacturer's protocol. cDNA was stored at –20°C until used for RT-qPCR analysis.

Reverse Transcription-quantitative PCR (RT-qPCR)

The RT-qPCR assays were performed in 10 µl reaction volumes consisting of 5 µl HotStart. Master Mix (New England Biolabs, Ipswich, MA), 1 µl of LCgreen (Biofire, Salt Lake City, UT), 0.5 µl of each *Potexvirus* primers (Olmedo-Velarde & Ochoa-Corona, 2016) PotexF3 forward (5'-CCT GAA ITC ICA RTG GGT IAA RAA-3') and Potex R3 reverse (5'-GCI ATR GTY TGI CCA GGI TT-3') primers (10 µM), 2 µl of cDNA template, and 1 µl nuclease-free water (Ambion, Austin, TX, USA). The RT-PCR combined with High-Resolution Melting was performed in a Rotor-Gene 6000 thermal cycler (QIAGEN, Hilden, Germany). The cycling parameters were: initial denaturation of 97°C for 2 min followed by 40 cycles of denaturation at 95°C for 20 s, annealing at 47°C for 30 s, extension at 72°C for 30 s, and a final extension at 72°C for 3 min. Finally, the 10 µl of the amplified PCR product continued toward HRM from 65 to 95 °C, temperature range. Positive, negative (non-template control) controls and healthy tissue were included.

The total RNA extraction, cDNA synthesis, and RT-qPCR conditions tested were performed during all the steps of the water filtration system experimentation.

Glass wool testing assays

Infected tissue sample weighted from 10 g down to 1g was diluted in 100 mL of autoclaved water separately. The PepMV spiked water was incubated in glass wool surfaces of 1g and 10g. The assessed time of contact was 10 minutes, 5 minutes, 30 seconds and 1-second samples were tested by triplicates. Autoclaved glass wool pads size was 4.8 cm of diameter and 1mm of density. The assay system was scaled for 18L *in-vitro* and 80L *field* water filtration (Figure 6.2).

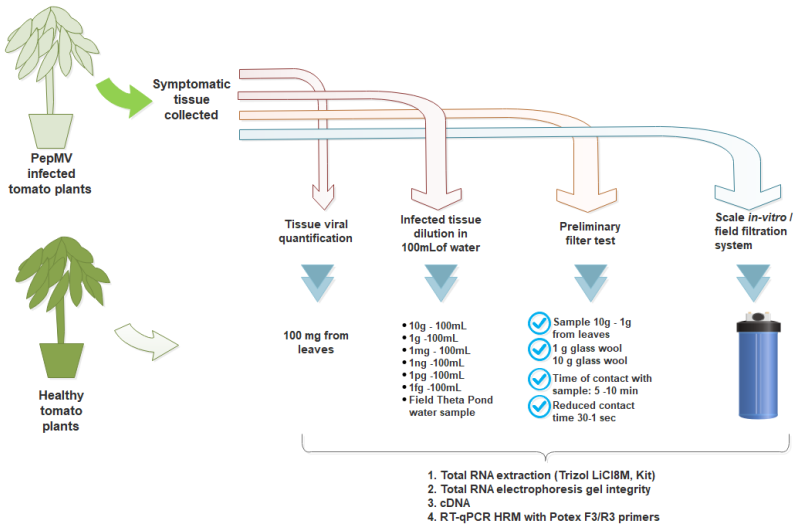


Figure 6. 2 Flow chart showing the steps of experimental design of the water filter.

2.3. Design of filtration system

Filters assembled using polyvinyl chloride (PVC) were used. The filter design used is a modification from Lambertini et.al. (2008). The modification of the structure of the filter was previously tested (Daniels 2016, Mazziero 2016). In this study the filter adaptations considered changing filter diameter and the grams of glass wool per cm^3(Figure 6.3). Glass wool is electro-positively charged after washes with 100mL of Reverse Osmosis (RO-18 ℧) autoclaved water, followed by one wash of 100mL 1M HCl, then 100mL 1M of NaOH, and a final wash with 100mL of RO autoclaved water per 100g of treated glass wool. Subsequently, the pH of the surface of the treated glass wool was adjusted to 6.5 and stored in PBS buffer pH 6.5 at 4 ºC until use (Lambertini et.al., 2008).

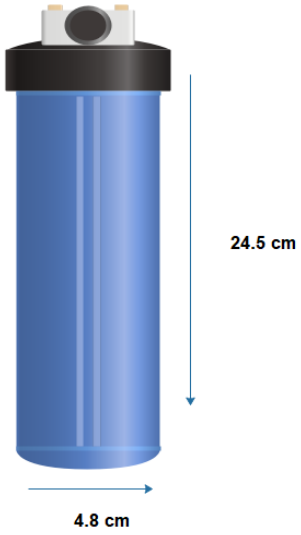


Figure 6. 3 Structural dimensions of the PVC water filter used in this study.

Positively charged glass wool, 300g (0.812g of glass wool per cm^3) was added to the filter. Filter prototypes were tested by triplicates at *in-vitro* and field.

2.4. Sample and *in-vitro* bioassay

Tomato leave tissue infected with PepMV (300g) was homogenized in 2L of autoclaved RO water. The homogenized tissue was filtered through autoclaved cheese cloth to avoid residual solid plant tissue. Aquariums were filled with RO water up to 54 L and spiked with the infected, freshly homogenized plant material. Aquariums were left 24h with an aeration system to avoid particle precipitation. After 24h filters were adapted with a submersible electric pump to filter 18L per filter (Figure 6.3). Temperature and pH were measured before filtering water.

2.5. Viral concentration, RNA extraction, RT-qPCR

After filtration, PepMV was eluted from the filter as described by Blanco et.al. (2019). The glass wool was immersed in elution buffer (3% beef extract with 0.05M of glycine, pH 11). The glass wool was incubated for 20 min in solution. The solution was filtered through autoclaved cheese

cloth and 20% PEG was added to precipitate virus particles. The solution was later incubated for ten minutes at room temperature. Then centrifuged at 10,000 rpm for 30min at 4 ºC. The supernatant was discarded. The pellet was re-suspended in 1mL of PBS pH 7.4. An aliquot of 450 µL was used for total RNA extraction with a modified Trizol 8M LiCl method (Wang et.al., 2009). The total RNA quality was measured by Nanodrop and its integrity was visualized by 1% agarose electrophoresis gel. The samples were prepared for RT-qPCR analysis, ds cDNA, and a library preparation at Oxford Nanopore Sequencing. cDNA synthesis and RT-qPCR were developed with the conditions described above (Section 2.6 Experimental design). dsDNA was synthesized using NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Kit (New England Biolabs, Ipswich, MA), as recommended by the manufacturer. dsDNA concentration, quality, and integrity were measured by Qubit 4™ (ThermoFisher Scientific, Waltham, MA) as recommended by the manufacturer. dsDNA was stored at –20°C until library preparation before Oxford Nanopore sequencing (Oxford Nanopore Technologies, Oxford, UK).

Field

2.6. Site description and sample collection

Field samples were collected at the Myriad Botanical Gardens, Oklahoma City, during August 2019. The sampling was composed of three filtrations on the same site. This study site was selected because the garden is made by a collection of tropical plants and has an artificial river to which plants contribute decaying foliage and the water source is recirculated simulating a tropical environment. Physical-chemical parameters, including pH, and temperature were measured at the collection site before filtration.

2.7. Filtration system

Three filters were prepared for each filtration site as described (section 2.6). The filters were modified by adding a pre-filter to avoid organic debris and soil accumulation in the glass wool

filter. Eighty liters were filtered. Filters were transported in ice and stored at 4 ^{o}C and processed the next day.

2.8. Viral concentration, RNA extraction, RT-qPCR

Viral concentration from field water was made with an addition of a 45µm filtration of the eluted pellet. This filtration was done to remove bacteria from the sample. Then, RNA extraction, RT-qPCR, and ds cDNA from the water filters were done as described in section 2.5, 2.8, and 2.10 respectively.

2.9. EDNA-Water (MiFi™) database

Six metagenomics databases generated by MinION sequencing were uploaded to EDNA-Water Mi/Fi^{TM}. The selected parameters were 100% percent identity and query coverage to assess. Hit frequencies between raw reads with E-probes were recorded for each *in-vitro* as well as field water samples. Hit frequencies data were analyzed with Tukey's HSD test and pairwise T-test at Pvalue= 0.05. Raw water metagenomic reads were parsed until the pipeline found a complete match with the E-probe set in the database. EDNA-Water correlates the number of matches in each sample as a semi-quantification of the plant water-borne virus in the host.

Sequencing

2.10. Library preparation and sequencing

Long fragments were amplified by a transposase enzyme (Oxford Nanopore) (Oxford Nanopore Technologies, Oxford, UK) as described in chapter IV and V. Library preparation was performed according to the Oxford Nanopore manufacturer's protocol. The library was kept on ice until ready to load on the MinION^{TM} device. The barcoding library preparation was according Oxford Nanopore protocol and the Ligation Sequencing kit (SQK-LSK108) (Oxford Nanopore Technologies, Oxford, UK). Barcoded libraries and dsDNA were quantified by Qubit 4™ (ThermoFisher Scientific, Waltham, MA), barcoded libraries were equimolarly-pooled before adding the adapters. The final library was stored at –20°C until MinION sequencing.

Sequencing was performed using the MinION flow cell (FLO-Min106 R9.4; Oxford Nanopore Technologies, Oxford, UK) Platform Quality Check (QC) was performed to determine the number of active pores available in the flow cell. There were 412 active pores after QC. The number of active pores is low if compared with a new flow cell that has ~1600 -1800 pores. After priming the flow cell following the per manufacturer's protocol, the pooled library mixed with Library Loading Beads (LLB; Oxford Nanopore Technologies, Oxford, UK) was loaded onto the SpotON port of the flow cell.

3. Results

***in-vitro* assay**

3.1. RNA extraction and viral quantification.

The selected RNA extraction method was a modified Trizol® method. The extraction method was selected based on the higher quantity and quality of the RNA obtained. High virus titer was detected 10^{-2} when the sample was in contact with 1 gram of glass wool per 30 seconds as expected. Lower detection was found with virus concentration at 10^{-6} (Table 6.1). Analysis of virus titer per cm^2 showed that the viral particles cluster in the glass wool. The efficiency of the virus recovery calculation determined that longer contact time of the sample with the glass wool up to 17% at 30 seconds.

Table 6. 1 Virus quantification testing in 1 gram of glass wool at 30, 10 and 1 seconds.

	Viral concentration [ng/uL]		
Repetition	**30 sec**	**10 sec**	**1 sec**
1	2.0E-06	0.0E+00	6.5E-03
2	0.0E+00	3.5E-05	0.0E+00
3	5.9E-02	1.7E-05	0.0E+00
Sample	0.112955832		
Mean Viral titter	0.019805306	1.70728E-05	0.002154543
Virus titter per cm2	0.003961061	3.41456E-06	0.000430909
% Efficiency	17.5336727	0.015114596	1.907420853

3.2. *In-vitro* filtration assay.

The filtration of 18L of water infected with PepMV showed low viral concentration in filters filled with 300 g of glass wool. Virus concentration was determined down to 10 fg per microliter. However, when filtering larger volumes a lower recovery efficiency of 2% was detected (Table 6.2).

Table 6. 2 PepMV recovery after filtration of 18L of virus spiked water.

Filter	Initial viral concentration[ng/uL]	RT- qPCR virus recovery[ng/uL]
1	6.35E-06	5.54E-08
2	1.14E-05	4.36E-08
3	5.18E-05	4.60E-07
Sample	0.112955832	
Mean Viral titter	2.31549E-05	1.86435E-07
% Efficiency	0.020499087	0.000165051

3.3 Field filtration assay

Field filtration assay of 80L of water with a detected low viral concentration of Potexvirus in filters filled with 300 g of glass wool. Virus concentration was detected down to 0.1 fg per microliter. The lower efficiency of 7.3% of virus recovery was detected (Table 6.3).

Table 6. 3 PepMV recovery after filtration of 80L of field virus filtration.

Filter	Viral [ng/uL]	qPCR from ds cDNA [ng/uL]
1	1.00E-07	0.00E+00
2	1.15E-07	0.00E+00
3	1.69E-07	0.00E+00
Sample	1.75E-07	
Mean	1.28092E-07	
%Efficiency	7.313E+01	

Before sequencing, a qPCR with Potex primers showed the presence of PepMV in samples filtered *in-vitro* at a low viral concentration (femtograms) (Table 6.2). PepMV presence (Table 6.3) was not detected in water samples from field filtration.

3.4 EPI2ME What is in my pot? (WIMP)

The MinION™ flow cell yielded 49,662 reads from 412 pores. Table 6.4 shows the taxonomically identified viruses found in each barcode. As expected, barcodes 07, 08 and 09, which correspond to *in-vitro* filtered samples detected PepMV and other related viruses described in Table 6.4.

Table 6. 4 EPI2ME Results of in-vitro and field filtration assays

Assay	Sample date	Sample ID	Liters filtered	Assigned Barcode	Total number of reads	Genus	Significant Viral Results	Number of matching reads
in-vitro	Summer 2019	Filter1	18	BC07	15965	Potexvirus	13 *Pepino mosaic virus* – 1 *Lily virus X*	14
						Flexovirus	5 *Erwinia phage phiEa104*	5
						Potyvirus	2 *Papaya ringspot virus* -1 *Zucchini yellow mosaic virus*	3
						Bracovirus	1*Cotesia congregata bracovirus*	1
						Alphabaculovirus	*1Orgyia pseudotsugata multiple nucleopolyhedroviruses*	1
						Triatovirus	1 *Homalodisca coagulata virus-1*	1
						Carlavirus	1 *Garlic common latent virus*	1
						Higrevirus	1 *Hibiscus green spot virus*	1
						Limmipivirus	1 *Carp picornavirus 1*	1
in-vitro	Summer 2019	Filter2	18	BC08	3829	Potexvirus	3 *Pepino mosaic virus*	3
						Potyvirus	1 *Zucchini yellow mosaic virus*	1
						Bracovirus	1 *Cotesia congregata bracovirus*	1
						Solendovirus	1 *Sweet potato vein clearing virus*	1
						Rabovirus	1 *RabovirusA*	1
in-vitro	Summer 2019	Filter3	18	BC09	14441	Flexovirus	*2 Erwinia phage phiEa104*	2
						Potyvirus	*1 Dasheen mosaic virus*	1
						Nepovirus	1 *Tobacco ringspotvirus*	1
						Potexvirus	1 *Pepino mosaic virus*	1
						Carlavirus	1 *Garlic common latent virus*	1
						Alphavirus	1 *Whataroa virus*	1

						Marafivirus	1 *Grapevine Syrah virus 1*	1
						Avastrovirus	1 *Turkey astrovirus 2*	1
						Flavivirus	1 *Tick-borne encephalitis virus*	1
						Alphacoronavirus	1 *Lucheng Rn rat coronavirus*	1
						Kobusvirus	1 *Kobuvirus cattle/Kagoshima-1-22-KoV/2014/JPN*	1
						Negevirus	1 *Piura virus*	1
						Bracovirus	*3 Cotesia congregata bracovirus*	3
						Alphabaculovirus	*2 Orgyia pseudotsugata multiple nucleopolyhedroviruses*	2
field	Summer 2019	Filter1	80	BC10	4432	Ichnovirus	*1 Glypta fumiferanae ichnovirus*	1
						Potyvirus	*2 Zucchini yellow mosaic virus-2 Papaya ringspot virus*	4
						Alphabaculovirus	*1 Orgyia pseudotsugata multiple nucleopolyhedrovirus-1 Agrotis ipsilon multiple nucleopolyhedroviruses*	2
						Bracovirus	*1 Cotesia congregata bracovirus*	1
						Muromegalovirus	*1 Murid betaherpesvirus 2*	1
field	Summer 2019	Filter2	80	BC11	10334	Sicinivirus	*1SicinivirusA*	1
field	Summer 2019	Filter 3	80	BC12	694	No viral reads found		

3.5 EDNA-Water rapid plant water-borne virus

Raw metagenomics databases from filters tested against the EDNA Plant water-borne viruses database detected the presence of *Potexvirus* and *Tobamovirus in-vitro* and field filtrated water samples. EDNA parameters were run at extremely sensitive E-value (eval 5) and a recommended minimum number of hits for semi-quantification (250 hits). The *Potexvirus* database reported the presence of PepMV, as expected (Table 6.5). Interestingly, it also revealed the presence of BaMV, CVX, CsCMV, CsVX, but there was not a correlation of the unexpected viruses in filters 1 and 2. PepMV was detected in the three repetitions of the *in-vitro* filtration. None of the filters used at the field showed the presence of PepMV but detected positive hits for CVX, CsVX, ClYMV, HVX, LaMMoV, MaMV, and PlAMV (Table 6.5).

The *Tobamovirus* database tested against raw metagenomes of *in-vitro* filtration detected positive hits for TMGMV in filter 2 only. None of the other filters detected other *Tobamovirus*. Field filters tested positive for BrMMV in filter 2 (Table 6.6).

None of the databases of *Tombusvirus* detected positive hits in any of the filters.

Table 6. 5 EDNA- Water positive hits for virus queried against *Potexvirus* database from filtered water *in-vitro* and field

	POTEXVIRUS																															
METAGENOME ID	**AlVX**	**AlsVX**	**AltMV**	**AV-3**	**BaMV**	**CVX**	**CsCMV**	**CsVX**	**ClYMV**	**CymMV**	**FoMV**	**HVX**	**HdRSV**	**LaMMoV**	**LeVX**	**LVX**	**MaMV**	**MVX**	**NMV**	**NVX**	**OpVX**	**PapMV**	**PepMV**	**PhVX**	**PlAMV**	**PVX**	**SchVX**	**SMYEV**	**TRMV**	**TVX**	**WClMV**	**YVX**	**ZyV**
Filter 1 in-vitro																							27										
Filter 2 in-vitro					34																		12										
Filter 3 in-vitro						11	12	10															12										
Filter 1 field						38		30				40		21											41								
Filter 2 field									1								13																
Filter 3 field																																	

Table 6.6. EDNA- Water positive hits for virus queried against *Tobamovirus* database from filtered water *in-vitro* and field

	TOBAMOVIRUS																									
METAGENOME ID	**BPeMV**	**BrMMV**	**CMMoV**	**ClitoriaYMoV**	**CFMMV**	**CuMoV**	**HLFPV**	**HLSV**	**MarMV**	**ORSV**	**PaMMV**	**PMMoV**	**PLUMERIAMOSAICVIRUS**	**RattailCNaV**	**RheMV**	**RMV**	**OpVX**	**SFBV**	**TMGMV**	**TMV**	**ToBrRuFV**	**ToMV**	**ToMoMV**	**TSAMV**	**TVCV**	**WMoV**	**ZGMMV**
Filter 1 in-vitro																											
Filter 2 in-vitro																											
Filter 3 in-vitro																			11								
Filter 1 field																											
Filter 2 field		14																									
Filter 3 field																											

4. Discussion

Plant waterborne viruses are particles that cluster in organic debris and travel through water sources. It is noteworthy that metagenomics research based on environmental samples is often separated by genome structure (Mehler & Ravnikar, 2012). This study seeks to detect plant water-borne viruses from HTS metagenomes. *Pepino mosaic virus* (PepMV) was selected as a model virus for this *in-vitro* study. A target-specific RT-qPCR combined to HRM using a genus-specific primer for PepMV were tested to allow to determine the PepMV concentration in the water sample. Results obtained showed RT-qPCR is able to detect virus particles down to femtograms. Quantification of PepMV determined the virus concentration in the sample during the validation process.

Environmental water research focuses heavily on virus detection. This is an area challenging and underdeveloped. To determine the size or volume of the sample is critical. Variables in play for consideration are virus capture and elution method, pH and temperature of the surrounding ecosystem, and environmental seasonal changes (Lambertini, et.al., 2008). A high efficiency of recovery was quantified when filtering PepMV *in-vitro* in spiked water samples. Low efficiency was observed, in large scale filtration at the field (80L). This may be caused by the speed of the water flow which interferes with the virus capture in the glass wool matrix. For many years the VIRADEL method has been in use for virus adsorption and elution. In this study, the VIRADEL method successfully allowed virus trapping using a slightly acid buffer (pH 6.5) (Hong & Moorman, 2005). The weight of the glass wool matrix gave a clear insight into the amount needed (0.825 treated glass wool per cm^3) to fill the filter to allow efficiency recovery. Blanco et.al. (2019), reported efficiencies of recovery up to 10 % after elution of glass wool with 3% beef extract 0.05M glycine at pH11 by agitation. In line with the observations in this study, Blanco et.al. (2019), also reported that virus recovery may be increased by recirculating the elution

buffer. Further studies engineering a recirculating elution system may improve efficiencies of virus recovery up to 18%.

The percentage of the recovery obtained in this study correlates with previous research indicating there is room for improvement. The evaluation of methods of concentration and recovery of plant viruses from irrigation water using HTS sequencing platforms is currently matter of study (Hjelmsø et.al., 2017). In this study, the RNA extraction method was standardized to ensure the high quality of RNA of the starting sample. The RNA integrity tested by gel electrophoresis was determined in a solution of 1g in 100 mL of water. in diluted samples tested during *in-vitro* filtration.

Pre-filtration step was incorporated to reduce organic matter debris accumulation in the glass wool filter of field samples. Similarly, a 45µm filter was used before the RNA extraction to reduce the bacterial input on HTS. However, this pre-filtration step may affect the virus titer (Hjelmsø et.al., 2017). The steps described above allowed to set a standardized method for RNA extraction, RT-qPCR, and filtration for HTS sequencing platforms. A solution to improve the efficiency of the filtration system may be adding a bag-mediated filtration system as described by van Zyl et.al., (2019) when collecting environmental samples in Kenya. Similarly, a large scale coagulation-sedimentation, followed by rapid sand filtration, ozonation, and biological activated carbon treatment was reported in Japan specifically for PMMoV (Kato et.al., 2018). Nonetheless, the proposed filtration system can be improved considering filter design and the efficiency of recovery.

RNA sequencing has been proposed for the detection of plant water-borne viruses (Hjelmsø et.al., 2017). The EDNA-Water-borne MiFi^{TM} database is a useful plant virus detection method for ss (+) RNA. This study aims to develop unique E-probes for three model genera known as plant-

waterborne viruses. Three genera, *Potexvirus, Tobamosvirus,* and *Tombusvirus* were selected as the model viruses (Gu et.al., 2018).

Oxford nanopore sequencing is a powerful HTS platform that may be field-deployable or for virus *in-situ* detection and diagnosis, therefore it was the selected sequencing method in this study. The EDNA-Water database was queried against *in-vitro* water samples (18L) and in the field (80L). A total of 49,692 reads divided into six barcodes were sequenced in 412 pores. After the demultiplexing process, the reads assigned per barcode indicated the expected presence of PepMV *in-vitro* samples including *Potexvirus* and *Tobamovirus* in field samples. The proposed EDNA-Water MiFi™ database successfully detected PepMV and other plant-waterborne viruses which is a contribution to aquatic environmental research.

EDNA-Water MiFi™ contains a larger database of E-probes able to screen and detect up to 80 viruses simultaneously in a single sample. Further research will include developing E-probes for the remaining four virus genus reported as plant-waterborne (Mehler & Ravnikar, 2012). EDNA-Water MiFi™ can be applied as a fast screening method that in the future for *in-situ* virus detection on irrigation water. This research shows the potential of EDNA-Water MiFi™ to be used in diagnostic clinics and a biosecurity protocol for ensuring clean recirculation water at greenhouse production areas as well as irrigation water for agricultural crops.

References

Alfaro-Fernández, A., Del Carmen Córdoba-Sellés, M., Herrera-Vásquez, J. Á., Cebrián, M. D. C., & Jordá, C. (2010). Transmission of Pepino mosaic virus by the fungal vector Olpidium virulentus. *Journal of phytopathology*, *158*(4), 217-226.

Anderson-Coughlin, B. L., & Kniel, K. E. (2019). Recovery and Detection of Enteric Viruses from Non-Traditional Irrigation Water Sources. *Methods and protocols*, *2*(3), 55.

Blagden, T., Schneider, W., Melcher, U., Daniels, J., & Fletcher, J. (2016). Adaptation and validation of E-probe diagnostic nucleic acid analysis for detection of Escherichia coli O157: H7 in metagenomic data from complex food matrices. *Journal of food protection*, *79*(4), 574-581.

Blanco, A., Abid, I., Al-Otaibi, N., Pérez-Rodríguez, F. J., Fuentes, C., Guix, S., ... & Bosch, A. (2019). Glass Wool Concentration Optimization for the Detection of Enveloped and Non-enveloped Waterborne Viruses. *Food and environmental virology*, *11*(2), 184-192.

Brye, K. R., Quarta, M., Morrison, C., & Rothrock, C. (2018). Long-term effects of residue and water management practices on plant-parasitic nematode abundance and soybean root infection. *Applied Soil Ecology*, *124*, 275-283.

Bush, E. A. (2002). *Characterization of Phytophthora species in recycled irrigation water at a container nursery in southwestern Virginia* (Doctoral dissertation, Virginia Tech).

Candresse, T., Hammond, R. W., & Hadidi, A. (1998). Detection and identification of plant viruses and viroids using polymerase chain reaction (PCR). *Plant virus disease control*, 399-416.

Cashdollar, J. L., & Wymer, L. (2013). Methods for the primary concentration of viruses from water samples: a review and meta-analysis of recent studies. *Journal of Applied Microbiology*, *115*(1), 1-11.

Cassman, N., Prieto-Davó, A., Walsh, K., Silva, G. G., Angly, F., Akhter, S., ... & Willner, D. (2012). Oxygen minimum zones harbor novel viral communities with low diversity. *Environmental Microbiology*, *14*(11), 3043-3065.

Castello, J. D., Lakshman, D. K., Tavantzis, S. M., Rogers, S. O., Bachand, G. D., Jagels, R., ... & Liu, Y. (1995). Detection of infectious tomato mosaic tobamovirus in fog and clouds. *Phytopathology*, *85*(11), 1409-1412.

Castello, J. D., Rogers, S. O., Starmer, W. T., Catranis, C. M., Ma, L., Bachand, G. D., ... & Smith, J. E. (1999). Detection of tomato mosaic tobamovirus RNA in ancient glacial ice. *Polar Biology*, *22*(3), 207-212.

Clark, M. F., & Adams, A. N. (1977). Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *Journal of general virology*, *34*(3), 475-483.

Decker, C. J., & Parker, R. (2014). Analysis of double-stranded RNA from microbial communities identifies double-stranded RNA virus-like elements. *Cell reports*, *7*(3), 898-906.

Dixon, G. R. (2015). Water, irrigation and plant diseases. *CAB Reviews*, *10*(009), 1-18.

Espindola, A., Schneider, W., Hoyt, P. R., Marek, S. M., & Garzon, C. (2015). A new approach for detecting fungal and oomycete plant pathogens in next-generation sequencing metagenome data utilizing electronic probes. *International journal of data mining and bioinformatics*, *12*(2), 115-128.

Gu, X., Tay, Q. X. M., Te, S. H., Saeidi, N., Goh, S. G., Kushmaro, A., ... & Gin, K. Y. H. (2018). Geospatial distribution of viromes in tropical freshwater ecosystems. *Water Research*, *137*, 220-232.

Hanssen, I. M., & Thomma, B. P. (2010). Pepino mosaic virus: a successful pathogen that rapidly evolved from emerging to endemic in tomato crops. *Molecular plant pathology*, *11*(2), 179-189.

Hjelmsø, M. H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., ... & Abril, J. F. (2017). Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. *PLoS One*, *12*(1), e0170199.

Hong, C. X., & Moorman, G. W. (2005). Plant pathogens in irrigation water: challenges and opportunities. *Critical Reviews in Plant Sciences*, *24*(3), 189-208. Hong, C. X., & Moorman, G. W. (2005). Plant pathogens in irrigation water: challenges and opportunities. *Critical Reviews in Plant Sciences*, *24*(3), 189-208.

Kato, R., Asami, T., Utagawa, E., Furumai, H., & Katayama, H. (2018). Pepper mild mottle virus as a process indicator at drinking water treatment plants employing coagulation-sedimentation, rapid sand filtration, ozonation, and biological activated carbon treatments in Japan. *Water Research*, *132*, 61-70.

Koenig, R. (1986). Plant viruses in rivers and lakes. In *Advances in virus research* (Vol. 31, pp. 321-333). Academic Press.

Lambertini, E., Spencer, S. K., Bertz, P. D., Loge, F. J., Kieke, B. A., & Borchardt, M. A. (2008). The concentration of enteroviruses, adenoviruses, and noroviruses from drinking water by use of glass wool filters. *Appl. Environ. Microbiol.*, *74*(10), 2990-2996.

Lamichhane, J. R., & Bartoli, C. (2015). Plant pathogenic bacteria in open irrigation systems: what risk for crop health?. *Plant Pathology*, *64*(4), 757-766.

Mehle, N., & Ravnikar, M. (2012). Plant viruses in an aqueous environment–survival, water-mediated transmission, and detection. *Water Research*, *46*(16), 4902-4917.

Olmedo-Velarde, F. Ochoa-Corona. 2016. Discriminating *Potexvirus* species by RT-PCR coupled to High Resolution Melting. (Abstr.) *Phytopathology 106: S4.118.*

Pares, R. D., Gunn, L. V., & Cresswell, G. C. (1992). Tomato mosaic virus infection in a recirculating nutrient solution. *Journal of Phytopathology*, *135*(3), 192-198.

Redekar, N. R., Eberhart, J. L., & Parke, J. L. (2019). Diversity of Phytophthora, Pythium, and Phytopythium Species in Recycled Irrigation Water in a Container Nursery. *Phytobiomes Journal*, *3*(1), 31-45.

Sarra, S. (2005). *Novel insights in the transmission of Rice yellow mottle virus in irrigated rice*.

Stewart-Wade, S. M. (2011). Plant pathogens in recycled irrigation water in commercial plant nurseries and greenhouses: their detection and management. *Irrigation Science*, *29*(4), 267-297.

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., ... & Schneider, W. (2013). E-probe Diagnostic Nucleic acid Analysis (EDNA): a theoretical approach for handling of next-generation sequencing data for diagnostics. *Journal of microbiological methods*, *94*(3), 356-366.

van Zyl, W. B., Zhou, N. A., Wolfaardt, M., Matsapola, P. N., Ngwana, F. B., Symonds, E. M., ... & Komen, E. (2019). Detection of potentially pathogenic enteric viruses in environmental samples from Kenya using the bag-mediated filtration system. *Water Supply*.

Wakeham, A. J., & Pettitt, T. R. (2017). Diagnostic tests and their application in the management of soil-and water-borne oomycete pathogen species. *Annals of Applied Biology*, *170*(1), 45-67.

Wang, H. M., Yin, W. C., Wang, C. K., & To, K. Y. (2009). Isolation of functional RNA from different tissues of tomato suitable for developmental profiling by microarray analysis. *Bot stud*, *50*, 115-125.

VITA

Lizbeth Daniela Pena-Zuniga

Candidate for the Degree of

Doctor of Philosophy in Plant Pathology

Thesis: EDNA-HOST: DETECTION OF GLOBAL PLANT VIROMES USING HIGH THROUGHPUT SEQUENCING

Major Field: Plant Pathology

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Plant Pathology at Oklahoma State University, Stillwater, Oklahoma in May, 2020.

Completed the requirements for the Bachelor of Science in Biotechnology Engeneer your major at Escuela Politecnica del Ejercito (ESPE), Sangolqui,/Ecuador in 2014.

Experience:

Research Scholar Oklahoma State University 2015

Professional Memberships:

American Phytological Society APS from 2016
International Horticultural Society since 2017