DETECTION OF ARTHROPOD TRANSMITTED VIRUSES OF CEREAL CROPS AND RNAI INDUCED RESISTANCE TO *WHEAT STREAK MOSAIC VIRUS*

By

PATRICK RYDZAK

Bachelor of Science in Marine Biology
Texas A&M University
Galveston, Texas
2012

Master of Science in Biology
University of Texas at Tyler
Tyler, Texas
2015

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
July, 2019

DETECTION OF ARTHROPOD TRANSMITTED VIRUSES OF CEREAL CROPS AND RNAI INDUCED RESISTANCE TO *WHEAT STREAK MOSAIC VIRUS*

Dissertation Approved:

Astri Wayadande, Ph.D.

_____

Dissertation Chair

Francisco Ochoa Corona, Ph.D.

_____

Co-Advisor

Robert Hunger, Ph.D.

_____

Committee Member

Robert Burnap Ph.D.

_____

Committee Member

ACKNOWLEDGEMENTS

On July 8, 2016, with the help of my parents I took the next step of my academic career by moving to Stillwater, Oklahoma to begin my doctorate at Oklahoma State University. Now in 2019, exactly three years to the day that I first moved to Stillwater, I will be defending my Ph.D. dissertation. This journey has been life changing for me and there are so many people that I would like to thank for their unwavering support. First, to my co-advisors Dr. Astri Wayadande and Dr. Franscisco Ochoa-Corona, I am honored to have had the privilege to be your student and witness you both be awarded the title of full professor. While working with you two over the years, I have never once been made to feel like 'just your student'. Instead, every day you have treated me like a peer and a professional, always showing me respect even when I felt it was unearned. This has meant more to me than you can know, thank you. To my other committee members, Dr. Robert Hunger and Dr. Robert Burnap, thank you for your support, your willingness to teach me and your constant positivity. I do not believe I could have selected a better group of scientists to be on my thesis committee and I am grateful for everything you have done for me. I would also like to acknowledge the entire department of Entomology and Plant Pathology. Our department head, Dr. Phil Mulder. In my mind, Dr. Mulder is the quintessential example of what a leader should be, confident, attentive and gregarious, traits I hope to emulate in the future. To the faculty, staff and students of the EPP department, thank you for letting me be a part of your family over these three years. It has been an absolute pleasure to be a part of such a supportive environment where everyone you are surrounded by is pushing you to improve and succeed. A special thanks to NIMFFAB and the Virus Chasers, I have learned so much from working with all of you closely in the laboratory every day. It is exciting and fulfilling to watch you succeed and I can't wait to see where you end up in your professional lives. Another special thanks to Dr. Peter Hoyt and everyone at the Core Facilities at Oklahoma State University for your assistance with my sequencing projects, I could not have done this work without you. Lastly, I would like to thank my parents Steve and Diane and my brother Andrew. Thank you for always believing in me. If any of you have ever had doubts about my ability to succeed you have never let them show. Once again, thank you to everyone mentioned above and to so many other who I do not have the space to mention by name, I am beyond grateful to you for being a part of this journey with me.

Name: PATRICK RYDZAK

Date of Degree: JULY, 2019

Title of Study: DETECTION OF ARTHROPOD TRANSMITTED VIRUSES OF CEREAL CROPS AND RNAI INDUCED RESISTANCE TO *WHEAT STREAK MOSAIC VIRUS*

Major Field: PLANT PATHOLOGY

Abstract: Bread wheat (*Triticum aestivum*) is one of the most important crops cultivated worldwide both in terms of nutrition and economic value. Some of the greatest challenges wheat growers face around the world are plant viruses, which may induce symptoms such as stunting or discoloration and can lead to yield losses, or in extreme cases, total crop failure. To identify potential solutions to the threat posed by plant viruses in wheat, one of the most important steps is to accurately and quickly detect and discriminate between viruses so the appropriate management strategy may be applied. Two of the most powerful technologies currently available for the detection and discrimination of plant viruses are PCR and massive parallel sequencing (MPS). The objectives of this study are to utilize multiplex PCR in combination with high-resolution melting (HRM), as well as the MPS based Electronic-Probe Diagnostic Nucleic Acid Analysis (EDNA), to develop new tools for the detection and discrimination of viruses of wheat such as *Wheat streak mosaic virus* (WSMV). Next, to test the efficacy of RNA interference (RNAi) as a potential treatment to induce resistance to WSMV in wheat. Primers were designed to simultaneously detect WSMV, *Maize mosaic virus* (MMV) and *Barley yellow dwarf virus* (BYDV) in multiplex, and to discriminate between the three species of BYDV using HRM. E-probes capable of detecting 21 different viruses of cereal in metagenomic data were designed and the theoretical limit of detection was assessed for WSMV using *in silico* predictive models. These models were validated by sequencing known ratios of WSMV to wheat nucleic acids in a series of dilutions, then analyzed using EDNA on the subsequent metagenomic data. Gallagher, a wheat variety susceptible to WSMV, was injected with an RNAi construct specific to WMSV, then inoculated with WSMV. Viral titer of WSMV was then monitored at 7, 14 and 21 days post inoculation with qPCR and results were compared to WSMV inoculated and uninoculated, susceptible and resistant wheat varieties. It was determined that susceptible wheat treated with RNAi specific to WSMV prior to inoculation with WSMV had significantly reduced viral expression.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER V

CHAPTER I

INTRODUCTION

Biological security, or 'biosecurity', was a term first introduced in the 1990's referring to the measures taken to reduce the transmission of infectious pathogens in agricultural systems (Meyerson & Reaser, 2002). Legislation centered around biosecurity includes, but is not limited to, policies that would prevent the spread of plant and animal pathogens, diagnose unknown or unidentified diseases and monitor the spread of pathogens and mitigate symptom severity in infected plants and animals. Agricultural biosecurity is considered important in many countries. Countries such as Australia, where the government implemented the *Quarantine Act of 1908,* or when the United States implemented new policies after 9/11 that emphasized biosecurity in an attempt to reduce the threat of bioterrorism. These policies are not only targeted against a deliberate and malicious release of pathogens into agriculture, but the unintentional release of pathogens or their respective vectors during trade and commerce as well.

The ever-expanding global agricultural market and the increasing threat of the release of invasive pest species and pathogens into new environments highlights the need for robust biosecurity tools. Following the establishment of defined borders, whether they are physical, political or geographical, the next step in the prevention of the spread of invasive pests and pathogens is the ability to detect and discriminate between them. Without the proper tools to quickly and accurately differentiate between a completely benign microorganism or

insect and potentially ecologically or economically crippling threat, global trade would be unfeasible due to the inherent risk taken with each transaction of goods. This risk to global biosecurity becomes even greater when considering the potential threat of agricultural bioterrorism. As with many discoveries in science, it is often noted that there is an ever-present potential for the misuse, or dual use, of the very same tools developed from them. This dual use potential of many scientific advancements poses significant challenges to biosecurity experts and law enforcement agencies as it is not always clear what form the next threat will take. Therefore, it is essential to expand the way potential threats are detected through the development of novel and evolving tools.

Currently, two of the most powerful diagnostic tools available for detection of nucleic acids from multiple viruses in a sample are multiplex polymerase chain reaction (PCR) and high-resolution melting (HRM). Unlike single target end-point PCR, multiplex assays detect multiple products in one sample by using multiple, non-overlapping pairs of primers designed to produce amplicons of varying lengths that can be visualized by gel electrophoresis (Deb & Anderson, 2008). HRM is a post-PCR analysis in which PCR product is denatured by heat over time. The rate of denaturization of individual amplicons is dependent on its putative nucleic acid composition, which is measured by the decline in florescence from a dye previously incorporated into the amplicon during PCR (Liew et al., 2008). Researchers can use a combination of multiplex PCR and HRM to accurately detect and discriminate between viruses or virus species within a single sample. Nucleic acid based diagnostic tools have been demonstrated to be sensitive and precise but are relatively expensive and limited in the number of targets that can be screened for in a single assay. Depending on the need, other diagnostic tools may also be required to detect the presence of a pathogen in a sample.

The advent of massive parallel sequencing (MPS) has had a significant impact on the field of plant pathogen diagnostics and has made it hypothetically possible to detect all the pathogens present within a single sample (Adams et al., 2009). Prior to the development of MPS, DNA sequencing was primarily performed using a chain-stopping method with dideoxy nucleic acids invented in 1977, known as Sanger sequencing (Sanger et al., 1977). The next sequencing method to emerge, shotgun sequencing, gained popularity during the human genome project and is the basis for MPS as it allows for the sequencing of relatively long DNA sequences when compared to Sanger Sequencing. MPS permits whole genomic DNA to be fragmented into shorter lengths, massively sequenced in parallel via the synthesis method with fluorescently labeled nucleotides, then assembled with no reference (*de novo),* or to a reference genome with the assistance of software (Shendure and Ji, 2008). This method allows high-throughput DNA sequencing, in which millions of sequences can be read in a single run.

The Electronic Diagnostic Nucleic Acid Analysis (EDNA) is an *in silico* technique where electronic probes (e-probes) are designed with high levels of specificity to detect the presence of target sequences in a metagenomic sample. EDNA is a bioinformatics tool designed to increase the speed of processing MPS outputs by searching for a select group of pathogens in a database and has the potential to detect every preselected pathogen target within a database. Until recently, the computational power required to analyze data, as well as the time needed to align and annotate sequences produced by MPS were limiting factors. A benefit to EDNA is that by using targeted computer-generated e-probes, it makes it possible to query a genomic database relatively quickly, without the need to assemble, align or annotate the sequences first. Between the ever-decreasing cost of sequencing and the use of *in silico* diagnostic techniques like EDNA, it is possible to take large pools of genomic data and analyze them in different ways. From a

biosecurity standpoint, EDNA is a powerful tool that can be used to monitor and detect the presence of any pathogen in an economically vital agricultural setting.

One flaw with diagnostics based on MPS is that there is currently no known measure of sensitivity available. For other diagnostic tools such as PCR and immuno-florescent assays, sensitivity is determined by isolating the target nucleic acids or protein, performing a series of dilutions and testing the limits of detection by performing the assay on each dilution. When compared to PCR, MPS and EDNA have several inherent differences in methodology that prohibit the use of a similar serial dilution strategy for determining sensitivity. The first limitation to determine sensitivity using MPS is the minimum amount of sample required for sequencing. For example, Illumina sequencing platforms have a much higher minimum required sample mass when compared to minimum mass of the target needed for PCR or enzyme-linked immunosorbent assay (ELISA). This fact alone restricts the side by side comparison of MPS sensitivity with that of PCR.

Even if this were not the case and MPS sequencing could be performed at lower concentrations, the traditional sensitivity assay may still not be applicable to EDNA due to the way it was designed to work. EDNA is based on the statistical probability that a sample contains the target nucleic acid sequence by the number of times the e-probes 'hit' on that target or targets. EDNA can be used to provide a qualitative answer to the presence or absence of target nucleic acids in a sample, not necessarily how many copies of the target are in the sample. In other words, despite how diluted and isolated a target may be, EDNA will always provide the same type of answer, for example, that 100% of the 'hits' in the sample are from target sequences and 0% from the host. Therefore, it is necessary to design sensitivity assays for EDNA

that work within the restrictions of MPS, as well as within the framework for which EDNA was designed.

Wheat is considered one of the most important crops in terms of human nutrition and economic value (Curtis & Rajaram, 2002). In November of 2016, the Food and Agricultural Organization (FAO) estimated that the total production of wheat worldwide was expected to reach 746.7 million tons, an increase of 9.24 million tons from what was produced in 2015 (FAO, 2016). Even with a steady growth in wheat production worldwide, there is an always increasing need for more food as the human population continues to rise. Since 1950, the global population has tripled, with a total of approximately 7.5 billion people reported by the Population Reference Bureau in 2016. It is predicted that the world population could reach 9.9 billion people by 2050 (PRB, 2016). With the ever-growing need for more food, agricultural biosecurity and food safety are paramount for economically important crops like wheat considering approximately half of all of the wheat worldwide is produced and consumed in developing countries (Aquino et al., 2000).

Among the threats to wheat biosecurity worldwide are pathogens including, but are not limited to the basidiomycete, *Puccinia graminis*, the causal agent of wheat stem rust, *Pseudomonas syringae*, the cause of a bacterial leaf blight in wheat, and *Wheat streak mosaic virus* (WSMV). WSMV, is one of approximately 47 viruses that are currently known to infect wheat (American Phytopathological Society, 2016), and one of ten economically important viruses of wheat prevalent in the high plains of North America (Seifers et al., 2008). Wheat streak mosaic (WSM) was first described in 1937 and is characterized for stunted growth, chlorosis, and a ‘streaking’ mosaic pattern (McKinney, 1937). In field trials conducted in 2011

and 2012, it was determined that cultivars of wheat that are susceptible to WSMV are also susceptible to *Triticum mosaic virus* (TriMV), and that losses in fields due to a combination of both viruses may reach as high as 100% (Byamukama et al., 2014).

Efforts to mitigate damage caused by WSMV include strategies aimed toward the management of its arthropod vector, *Aceria tosichella*, the wheat curl mite (WCM), breeding resistant varieties and use of transgene induced resistance (Seifers et al., 2009; Hunger et al., 2015; Fahim et al., 2012). All three of these strategies have been employed effectively to mitigate the severity of WSMV infections in the past, but no management strategy is perfect and there are deficiencies in each. Management of *A. tosichella* is primarily based on the elimination of volunteer wheat, which is part of the green bridge from season to season. If overlooked, volunteer wheat or volunteer grass species will act as a reservoir for populations of *A. tosichella*, which can spread across entire fields and can be carried by wind to neighboring fields (Oliveira-Hofman, 2015). Conventional breeding methods have been successful for developing pathogen resistant plant cultivars, but producing crosses is a time-consuming process and can take many years before a cultivar is commercially available. Transgenic or genetically modified (GM) crops are possibly the most rapid method available for the generation of resistant varieties, which can be tailored for resistance to many different pathogens and for a range of environments. A drawback to the production of GM wheat however, is that numerous countries have imposed regulations and banned the use or sale of GM crops, thereby reducing its access and value on the global market.

RNA interference (RNAi) occurs naturally as part of the post-transcriptional gene expression pathway, as well as the innate immune response against viruses in most eukaryotic

cells (Fire et al., 1998). This pathway is activated when double stranded RNA (dsRNA) enters the cytosol and is broken down by the RNAse III Dicer-2, into 21-25 base pair dsRNA sequences with a two-base overhang on the 3' end referred as short interfering RNA (siRNA) (Tabara et al., 1998). These siRNA are incorporated into an RNA-induced silencing complex (RISC) via the ribonuclease silencing protein Argonaute (Ago), in which the attached siRNA is used to guide the RISC to a messenger RNA (mRNA). Once a complimentary mRNA sequence is found and bound by the complementary guide siRNA, it is either cleaved by the ribonuclease Ago, or is blocked from translation (Zamore et al., 2000)

The RNAi pathway can be used to knock down host or viral gene expression in eukaryotic cells by designing, then introducing novel dsRNA sequences complementary to target mRNAs (Mello & Conte, 2004). For example, stable resistance in wheat against WSMV has been reported to occur during two generations of wheat by inserting a transgene expressing a micro RNA (miRNA) derived dsRNA hairpin complementary to conserved regions of the WSMV genome (Fahim et al., 2012). While transgene induced resistance in wheat to WSMV has been shown to be effective (Cruz et al., 2014), it poses the same problem of similar transgenic strategies, specifically the lack of acceptance worldwide. To resolve this problem, it may be possible to design a novel method that produces the effectiveness of transgenic derived resistance in wheat to WSMV without inserting new genes, or altering the genes naturally found in the plant. One possible solution is to use a naturally occurring endophyte species as a vector to transcribe and deliver targeted dsRNA transcripts to wheat cells as a method to induce resistance to WSMV via the RNAi pathway.

The objectives of this research will be to provide novel tools for the detection, discrimination and treatment of viral pathogens of wheat, with specific emphasis on one of the most economically important pathogens, WSMV. These objectives are:

1. Utilize a novel combination of multiplex end-point PCR with high-resolution melting for the detection of three important viruses of cereals, *Wheat streak mosaic virus* (WSMV), *Maize mosaic virus* (MMV) and *Barley yellow dwarf virus* (BYDV), as well as further discrimination of three species of BYDV, BYDV-PAV, BYDV-PAS and BYDV-MAV.
2. Develop e-probes for use in EDNA capable of detecting twenty-one different viral pathogens of wheat in unassembled metagenomic data, as well as determine the theoretical *in-silico* limits of detection for WSMV, BYDV and three species of BYDV, BYDV-PAV, BYDV-PAS and BYDV-MAV.
3. Validate EDNA as a tool capable of accurate quantification of viral titer within a mixed meta-transcriptomic sample via a serial dilution of a known quantity of target WSMV nucleic acids within a background of wheat nucleic acids.
4. Test the efficacy of RNAi as a potential protective agent against WSMV in a susceptible wheat variety.

CHAPTER II

REVIEW OF LITERATURE

Introduction to Wheat

Bread wheat (*Triticum aestivum*) is the most widely adapted crop plant and debatably the most important crop ever cultivated by humans both in terms of nutrition and economic value (Curtis et al., 2002). Wheat, maize (*Zea maize*) and rice (genus: *Oryzae*) together comprise approximately 75% of the carbohydrates and 50% of the protein consumed by humans (Curtis et al., 2002). Out of these three grains, wheat is considered the most nutritious as it contains relatively high amounts of carbohydrates, protein, fiber, vitamins and minerals (Klepper et al., 1982). Wheat is a staple food for almost half of the human population and provides 20% of all calories consumed, particularly in developing regions (Heyne et al., 1987). While wheat is predominantly considered a cool-season crop, it is commonly grown all over the world, accounting for approximately 20% of all cultivated land area (Curtis et al., 2002). One of the reasons wheat is capable of being grown across a wide range of environments is its ability to tolerate and grow in temperatures ranging from 3-4°C to 32°C, with an optimal growth temperature of 25°C (Klepper et al., 1982). For ideal growth rates, wheat should be planted in well-draining soils up to elevations of 3000m with annual rain fall between 37.5-87.5cm (Klepper et al., 1982), although it is known to grow in

regions with as little as 25cm of annual rainfall, to as much as 175cm (Klepper et al., 1982).

T. aestivum Life History

As a member of the family Poaceae, the tribe Triticeae and the genus *Triticum*, wheat is just one of the approximately 10,000 species of grass that are found on earth. Its evolutionary history is a relatively complicated one, including ancestors with a variety of ploidy levels such as diploid (2n = 14), tetraploid (2n = 28) and hexaploid (2n = 42) (Gill et al. 2006). This history begins with the hybridization of two wild diploid species, *Triticum uratu* and *Aegilops speltoides*, followed by a chromosome doubling event, leading to the fertile hybrid *Triticum dicoccoides*. Over the next 10,000 years, *T. dicoccoides* was cultivated into what is known as Emmer wheat and crossed with the wild diploid species *Aegilops tauschii*. Another chromosome doubling event occurred approximately 8,000 years ago and the resulting fertile hybrid hexaploid species became what is known as *T. aestivum*, or bread wheat. All of these wild wheat species are classified as winter wheats, characterized by small, red seeds (Klepper et al., 1982). Over time, cultivation of wheat with selection for soft, white seeds has led to what is now known as spring wheat species.

T. aestivum Anatomy and Growth Stages

Wheat, like other cereals, has a very specific and highly characterized growth scale known as Feeke's growth scale (Large, 1954). This scale is divided into 11 stages, with each stage falling under one of four general categories: tillering, stem extension, heading and ripening. The first five growth stages fall under tillering, with stage one beginning with the emergence of the first shoot, followed by stage two, when tillering first begins. Tillering, defined as the production of side shoots known as tillers, which occurs within members of Poaceae after the

initial parent shoot grows from the seed. Each tiller is segmented and emerges with its own leaf, ultimately forming dense tufts of grass characterized as stage three. The next stage, stage four, is when the leaf sheaths begin to lengthen, followed by stage five, where the sheaths become strongly erected. The next stages, six through ten, are under the category known as stem extension.

During stage six of stem extension, the first node, or the part of the plant stem from which the leaves arise, becomes visible. Next, in stage seven, the second node becomes visible, followed by stage eight, where the last leaf, known as the flag leaf emerges. In stage nine the ligule, or the junction of blade and sheath of the leaf, becomes visible. This is followed by stage ten, known as 'in boot'. The boot stage is the point in development where the seedhead is enclosed within the sheath of the flag leaf and is the stage at which the seed head is vulnerable to cold temperatures (Klepper et al., 1982). Still within stage ten, the wheat seedhead continues to emerge through a process known as heading. At the end of heading, wheat goes through flowering, when the monoecious wheat head self-fertilizes. Following fertilization, the seedhead goes through the final growth stage, growth stage eleven, known as ripening.

Management of Wheat Production Losses

Each year in the United States, approximately 25% to 30% of wheat crop is lost due to a combination of biotic and abiotic stresses such as weather, poor soil conditions, damage from insect pests and plant pathogens (Armbrust et al., 1974; Martin & McNamara, 1987; Cassman, 1999; Oerke, 2006; Anderson et al., 2004). Historically, the most effective method for combating yield losses over time in an environmentally responsible manner has been the adaptation of local wheat varieties through breeding programs and variety trials. However, recent advancements in

molecular biology and genetic engineering have made it possible to directly insert targeted, choice genes into the genome of plants producing what crops that are designated genetically modified (GM). Currently, no GM wheat varieties are available on the global market due to poor public perception of GM crops and strict government regulation on their import and export in most of the world (Davison, 2010).

Viruses of Wheat

Viruses are microscopic, virulent molecules that are composed of a DNA or RNA genome surrounded by a protein coat, which is some cases is enveloped with a membrane consisting of lipids and carbohydrates (Hunter et al., 1976). When not inside of a host, viruses are inert and require a vector or some other mechanism that facilitates entry into its host (Smith 1965). Once inside the appropriate host cell, the virus undergoes one of many possible replication strategies depending on the family to which it belongs. Most viral replication strategies follow a similar pattern: entry into the cell, release of genetic material, host-mediated translation of viral proteins, replication of virus genome, encapsulation of new virus particles and movement of virus particles into neighboring cells. In wheat, the most common route of infection is mediated by arthropod vectors, but other routes of infection exist including sap, seed, pollen, mechanical transmission and other vector species (Brakke 1987).

Once infected with a virus, a wheat host may respond with reactions ranging from a latent, or asymptomatic response, to plant death depending on the species. Common symptoms of virus infected wheat includes stunting and discoloration with some combination of mosaic, streaking, yellowing, rosetting and necrosis (Bawden 1950). In some cases, wheat is known to become infected with two or more viruses simultaneously, which is known as a multiple infection (Burrows et al., 2009). The effect of a multiple infection on the symptoms of an

infected wheat host can be additive, synergistic and even cross-protective in some cases (Adams et al., 2014; Burrows et al., 2009; Gonsalves & Garnsey 1989). Nearly all cells within wheat are susceptible to one virus species or another, but many viruses are limited to the phloem. Phloem-limited virus inoculation is limited to specific insect species such as aphids (Bennett 1940).

Taxonomy of wheat viruses is based on several factors including virus morphology, its physical, genomic and serological properties as well as its host range and mode of transmission (Murphy et al. 2012). The standard practice for naming these viruses is based on the combination of the host it infects and the symptoms it causes (Murphy et al. 2012). While classification of viruses down to the species is possible, any given individual virion is genetically distinct from any other virion and is considered a viral mutant or quasi-species (Lazarowitz & Shepherd 1992). The morphology of wheat virus particles varies and can be filamentous (rigid or flexuous), isometric, or spherical and bacilliform. Most viruses of wheat contain an RNA genome. A few exceptions do exist however, such as *Geminiviridae*, which have a DNA genome. Regardless of the type of nucleic acid present, individual virions range in diameter from 10 to 70 nm (Murphy et al. 2012).

Wheat streak mosaic virus

In 1922, a disease coined yellow mosaic was first reported in Nebraska and was soon found in much of the United States (Singh et al. 2018). Since then, yellow mosaic has been reported in Canada, Mexico, Australia and Eastern Europe. Now known as wheat streak mosaic (WSM), this virus disease has been found to infect entire fields covering hundreds of acres across (Burrows et al. 2009). In addition to infecting summer and winter wheat, WSMV is known to cause disease in rye, barley, oats and other grass species (Slykhuis 1961). Originally classified as a member of the genus *Rymovirus*, in 1999 *Wheat streak mosaic virus* (WSMV) was

placed into the genus *Tritimovirus*, within the family *Potyviridae* (Zagula et al. 1992). The morphology of WSMV is flexuous rods approximately 15 nm in width and 700 nm in length (Lee 1965). WSMV is a single-stranded, positive sense RNA virus that is 9,384 bp in length (Niblett et al. 1991).

Symptoms of WSM vary, and depend on the strain of the virus, host variety and environment conditions. Generally, symptoms of WSM include stunting, mottling and greenish-yellow, parallel and discontinuous streaking (Matthews 1993). In winter wheat, symptoms are usually expressed in the spring and as temperatures continue to rise, stunting and yellowing become more apparent (Hadi et al. 2011). When symptoms are severe, winter wheat yields can be lost due to partially formed or sterile heads, as well as extreme yellowing of leaves leading to necrosis and loss of photosynthetic ability (Gao and Nassuth 1993). Depending on a number of factors, the yield loss in a single season can vary from insignificant, to 100% loss of the entire field (Burrows et al. 2009).

WSMV is mechanically transmissible via infected sap and rarely transmitted via seed (~2%), but the primary route of infection of wheat fields from year to year is the semi-persistent transmission of the virus by *Aceria tosichella* (Acari: Eriophyidae), the wheat curl mite (Jones et al. 2005). During the spring and early summer, *A. tosichella* thrives on the lush green tissue of wheat and other grass species (Skoracka et al. 2013). During this time, these mites can develop from an egg to adult within eight to ten days, leading to significant increases in the population when the environment is favorable (Skoracka et al. 2013). Only 0.3 mm in length, these mites are difficult to detect with the unaided eye but are usually found feeding on the upper surface of leaves near the margins, causing the leaves to curl toward the midvein (Christian & Willis 1993).

This leaf curling is what gives the leaf curl mite its characteristic name and acts to protect it from the environment (Christian & Willis 1993). Wheat curl mites can migrate short distances across the leaf surface via its short legs, but for longer distances between plants and across fields, mites are carried by the wind (Gibson & Painter 1957).

During the spring and autumn, winds spread the mites across fields where they fall among their potential hosts (Gibson & Painter 1957). Besides wheat, the wheat curl mite is known to survive on a variety of commercially grown crops such as maize and millet, as well as other perennial wild and cultivated grass species (Thomas et al., 2004a). Transmission of WSMV from one season to another depends on the ability of the mite to find a viable host for both the virus and itself, as neither can survive on ripe grain or grass (Thomas et al., 2004b). This non-ripe material that the wheat curl mite and WSMV survive on over the summer is known as the 'green bridge'. Regions where there is a combination of annual wheat production and an abundance of green shoots and volunteer grasses for the wheat curl mite to over-summer on are where WSM is most severe (Jiang et al., 2005). For example, regions where both winter and spring wheat are planted, regions where wheat and maize overlap and regions where wheat is planted late for dual use as forage are where WSM is most prevalent (Velandia et al., 2010).

Management of WSM is primarily accomplished through cultural methods, such as destroying volunteer wheat at least two weeks before emergence of seedling wheat (Thomas et al., 2004a). This control of the green bridge is generally accomplished with tillage or the use of herbicides, which results in a decrease in the mite population that is carried over from the previous season (Thomas et al., 2004a). Another strategy to reduce the impact of WSM on yield is the use of resistant wheat cultivars (Thomas et al., 2004b). These cultivars can either convey

resistance against the mite vector, reducing their ability to transmit WSMV, or they could provide resistance against the virus directly (Tatineni et al., 2016; Thomas et al., 2004b). Transgenic lines of wheat have been produced that confer complete resistance against WSMV, however, these varieties are not commercially available (Fahim et al. 2010).

Barley yellow *dwarf virus*

Barley yellow dwarf (BYD), caused by members of the family *Luteoviridae*, is the most widespread virus disease of cereal crops worldwide and is found in every region where cereal grains are grown (D'Arcy & Burnett 1995). This family of viruses contains two genera, *Luteovirus* and *Polerovirus*, which are icosahedral and 25 mm in diameter (D'Arcy & Burnett 1995). Each virus particle contains two structural proteins and a single-stranded, positive sense RNA 6 kb genome that encodes between six to eight proteins (Miller et al., 1988). The number and organization of the corresponding genes for these proteins differ depending on which genera the virus is found in. Each species of BYD is taxonomically distinct from one another by variation in genomic sequences (Irwin & Thresh 1990). Another factor that differentiates one member of BYD from another is its insect vector. All BYD species are phloem limited and are acquired and transmitted by specific aphid vectors (Rochow 1970).

The taxonomy of the family *Luteoviridae* is a complex one and is often under debate. Originally, BYD viruses were divided into two groups, subgroup 1, composed of PAV, MAV, SGV and GAV, while subgroup two contained RPV, RMV and GPV (Miller et al., 1988). At the time, these distinctions in subgroup were based on cytology, while the more modern divisions are based on RNA genomics (Miller et al., 2002). There are currently eight described species of BYD: GPV, RMV, GAV, RPV, SGV, MAV, PAS and PAV (Pagan & Holmes 2010). Three of these species, PAV, PAS and MAV, are separately recognized as species of *Barley yellow dwarf*

virus (BYDV) and are categorized as belonging to the genus *Luteovirus*. RPV, within the genus *Polerovirus*, is unique among the BYD in that it's considered a species of *Cereal yellow dwarf virus* (CYDV-RPV) (Pagan and Holmes 2010). Also, like BYDV-PAV, CYDV-RPV has recently been divided into CYDV-RPV and CYDV-RPS (Malmstrom and Shu 2004). The remainder of the BYD viruses are currently unassigned to a genus and designated as BYD-RMV, BYD-GAV, BYD-SGV and BYD-GPV (Pagan and Holmes 2010).

Due to the wide variety of causal agents and the relatively large host range of BYD viruses, the symptoms caused by infection can vary. The most common symptoms are chlorosis of the leaves, particularly in the flag leaf, shortening of the internodes, stunted growth, leaf discoloration and distortion, leaf twisting and scorching and abnormal growth (McKirdy et al., 2002). In most hosts of BYD, infected plants will have stiff, brittle leaves that range in color across yellow, orange, red, purple and brown (McKirdy et al., 2002). With the exception of maize, where symptoms of BYD infection worsen when exposed to temperatures over 25°C, most plant hosts experience more severe symptoms when temperatures are cooler and light intensity is higher (Royer et al., 2005). Management of BYD viruses is generally attempted by the development of resistant varieties, of which there are few, as well as management of the aphid vectors.

The nomenclature for each of the eight species of BYD and CYD are based on the specific aphid vector that transmits them (Gildow 1987). RPV for example, is transmitted by the aphid *Rhopalosiphum padi* (L.), while SGV is transmitted by *Schizaphis graminum* (Rondani) (Sadeghi et al., 1997). Regardless of the aphid vector, the mode of transmission for BYD is circulative, with virus particles being acquired from the host during feeding (Gildow 1987).

After virus ingestion, ligands on its surface bind to receptors present on epithelial cells in the hindgut of the aphid, where virus particles then are actively transported through the epithelial cells and into the hemolymph (Gildow 1987). These virus particles then circulate through the hemolymph until they come into contact with the accessory salivary glands (Gildow 1987). Once there, virus particles are actively transported through the salivary glands, into the salivary duct and injected into the phloem of the host as the aphid feeds (James & Perry 2004). BYD viruses do not replicate within the body of the aphid vector (Gray et al., 1991). Starting from the point of feeding, the process of acquisition and circulation of BYD viruses takes between 12 and 48h before the insect is capable of inoculating a new host (Bath & Chapman 1967).

Polymerase Chain Reaction

Polymerase chain reaction (PCR) is one of the most fundamental tools of scientific research and a cornerstone of modern molecular biology. The original idea for PCR is attributed to Kary Mullis, who reportedly first developed the technique in 1983 (Mullis et al., 1986). Mullis would ultimately be awarded the Nobel Prize for chemistry in 1993. While the use of short, single stranded DNA sequences was common at the time, Mullis's contribution was the use of a second, complementary oligonucleotide to juxtapose the first. By combining two short DNA sequences in this way, it became possible to amplify DNA in a specific, targeted and repetitive manner so that each round of PCR doubles the available template for the next round. Around the same time, a colleague of Mullis, Henry Erlich, was performing work in the isolation of a thermostable DNA polymerase from the bacteria *Thermus aquaticus* (Saiki, 1988). Until this point, DNA polymerase used in PCR was isolated from *Escherichia coli*, which was sufficient for many applications but did not work for PCR due to the heat needed to denature DNA after

each round of PCR. This relatively high heat (~95°C) denatures the DNA polymerase isolated from *E. coli*, leading to the need to add new aliquots of polymerase after every round of PCR.

T. aquaticus is unique in that it was isolated from a hot spring, where it required a DNA polymerase enzyme that is stable at high temperatures to survive. This polymerase was cloned and subsequently utilized in PCR where it became possible to perform the entire reaction without the need to add enzyme between each cycle. A second benefit to using DNA polymerase from *T. aquaticus* is its ability functions at higher temperatures (72°C), leading to higher fidelity when copying template DNA strands due to more efficient primer binding (Innis et al., 1988). This also leads to a decrease in secondary products from non-specific amplification. Even with the improvement to PCR with the introduction of *T. aquaticus* DNA polymerase (*Taq* polymerase), the technique was slow and laborious, as the operator was required to manually transfer samples between water baths with varying temperatures. With the advent of the first thermal cyclers, PCR become a more widespread technique.

Over time, further enhancements were made to PCR such as the ability to perform the reaction on RNA by first reverse transcribing it in what became known as reverse transcriptase PCR (RT-PCR). Another powerful modification to PCR is the ability to quantify the reaction by measuring the exponential growth of double stranded DNA (dsDNA) in the reaction with the use of florescent dyes. There are currently several ways to perform what was formerly known as real-time PCR, now more accurately described as quantification PCR (qPCR), but the most common methods include intercalating dyes and fluorescently tagged DNA oligo probes. Intercalating dyes, such as EvaGreen or LCgreen, are proprietary molecules that fluoresce when incorporated into dsDNA (Eischeid 2011). As the PCR reaction progresses and more dsDNA

molecules are polymerized, the overall fluorescence of the reaction increases proportionally. A flaw of this method, however, is that the fluorescence measured in the reaction also includes any fluorescence produced from the incorporation of intercalating dyes into non-specific products or primer dimers (Mao, Leung, and Xin 2007).

One qPCR method that is used to solve this problem is the use of fluorescently tagged DNA oligo probes. During PCR, two short, DNA primers are used to flank both ends of the targeted region of DNA (Saiki, 1988). Some versions of PCR will incorporate a third short oligo sequence known as a hybridized probe, which has been modified with a fluorescent molecule and a quencher that suppresses its fluorescence (Johansson & Cook, 2003). During PCR, as DNA polymerase moves from 5´ to 3´, it dislodges the fluorescent molecule from the quencher, leading to an increase in fluorescence of the reaction. Similar to intercalating dyes, this fluorescence is measured, and the rate of PCR is quantified over time. Another advantage to this technique is that it can be used in multiplex to quantify the rate of reaction for multiple targets within a single reaction with the use of multiple probes modified with fluorescent tags that vary in the fluorescent wavelength emitted (Osman et al. 2015).

High Resolution Melting

Melting curve analysis was first introduced in combination with real-time PCR in 1997 to provide more detail regarding what products were amplified during PCR (Ririe et al., 1997). The strength of this technique is the ability to distinguish between two otherwise identical PCR products by the rate at which the two stands of DNA dissociate from one another when the temperature is slowly and incrementally increased over time. As the temperature increases and the strands of DNA separate, the intercalating dye incorporated into the dsDNA during PCR is released into the solution and ceases fluorescing. As more dye is released, fluorescence decreases

at a corresponding rate, which is measured and recorded. HRM was first introduced in 2002 through a collaborative effort between the University of Utah and Idaho Technology as a simple method for scanning for genotype mutations (Wittwer & Herrmann 2002). In combination with PCR, HRM is a powerful DNA diagnostic tool used in a variety of applications including differentiating between virus species, strains and subtypes (Hasiów-Jaroszewska & Komorowska 2013).

Massive Parallel Sequencing

Massive parallel sequencing (MPS), is a method for the simultaneous sequencing of relatively small DNA or complimentary DNA (cDNA) fragments. This technology is currently one of the fastest growing research and clinical tools in the world due to a combination of its decreasing cost, versatility and ability to generate relatively large data sets. Every year, new strategies emerge based on MPS technology including bioinformatic pipelines and laboratory benchtop procedures that continue to enhance its efficiency and the ways it can be used. MPS platforms were first made commercially available in 2005, having gained support from their successful use during the Human Genome Project, which was completed in 2003 (Collins 1998). Before the Human Genome Project, the predominate form of sequencing taking place was based on Sanger's dideoxynucleic acid chain stop method, which yielded maximum sequence reads of approximately 1000 nucleotides (Shendure et al., 2011). With the advent of MPS, the maximum sequence reads obtainable in a single reaction significantly increased due to a method known as shotgun sequencing. Shotgun sequencing is based on the random fragmentation of longer DNA sequences into shorter ones that would then be sequenced simultaneously in parallel. Once sequenced, these short reads would then be reassembled in a process known as *de novo* alignment, or an alignment in the absence of a reference genome.

The current methodology for MPS can be summarized in four general steps, DNA extraction, library preparation, target enrichment and sequencing. For DNA extractions, almost any method is considered adequate for the goal of sequencing so long as it meets the minimum mass requirements. Quantification of resulting DNA is most often performed using Picogreen or Qubit, and not standard spectrophotometry. The next step, library preparation, refers to the modifications made to DNA prior to MPS. Depending on the exact method, or company that manufactures the kit being used for library preparation, the exact protocol may vary, however, the result of these kits is the fragmentation of the DNA and the addition of adaptors to the end of each DNA fragment.

These adaptors function as targets for universal PCR primers and often also include a region known as a bar code, which allows the user to sequence mixed DNA samples while keeping the results separate. Next, the library undergoes what is known as target enrichment, a process where the adaptors attached during library preparation hybridize with DNA fragment sequences on a sequencing chip and undergo PCR. Over the course of PCR, one of several reactions will occur based on the sequencing platform being used. In Illumina sequencing for example, four fluorescently tagged nucleotides are released into the system and when incorporated into elongating DNA sequences during PCR, release florescent emissions that are measured and recorded. This occurs simultaneously for millions of DNA sequences at once, leading to the generation of a massive volume of sequencing data per run.

Metagenomics

Once generated, the data that arises from MPS can be analyzed in a wide range of applications and bioinformatic pipelines. One of the most common uses of this form of data is to analyze it as a metagenome, or the genetic material recovered directly from environmental

samples. A metagenomic sample taken from a plant for instance, would contain the genetic material of the plant, any latent genetic material on its surface from a recent interaction with another organism, every microorganism on its surface and every microorganism inside the plant. These sources of genetic material combined form what is known as the metagenome of the sample. There is also what is known as metatranscriptomics, which is the data generated from the cumulative gene expression of the organisms included when the sample was taken. To generate a metatranscriptome, the total RNA of the sample must be extracted instead of the DNA, then processed to remove unwanted repetitive sequences such as ribosomal RNA (rRNA) and transfer RNA (tRNA) before it is reverse transcribed to cDNA.

This ability to sequence all the genetic and transcriptomic information within a single sample represents an important opportunity within the field of plant pathology and more specifically, diagnostics. MPS technology used in this way has the theoretical potential to detect any number of pathogens within a single assay, a feat that is not possible using the most common detection methods such as PCR, ELISA and lateral flow. However, there are two attributes of performing an MPS based analysis that are currently prohibitive to its widespread adoption for diagnostic use, the price of sequencing and the time and expertise required to perform the bioinformatic analysis of the data.

E-probe Diagnostic Nucleic Acid Analysis

E-probe diagnostic nucleic acid analysis (EDNA) is a method developed for the detection of target specific nucleic acids within an unassembled metagenome using highly curated electronic probes (e-probes). The design process for e-probes is based off of a modified Tool for Oligonucleotide Fingerprint Identification (TOFI), where target genomic sequences are compared to a list of closely-related, non-target sequences known as ‘near neighbors’ (Umek et

al. 2001). E-probes between 30-120bp in length are then generated from isolated regions of the target genome that are unique and do not match to the 'near neighbors', or phylogenetically related species. The more 'near neighbors' used and the closer they are phylogenetically to the target, the more specific the e-probes will be. Once generated, these e-probes are then curated using the basic local alignment search tool (BLAST) to query the NCBI and Gene bank database for any matching sequences. Any e-probes that match non-target sequences after BLAST analysis, are removed from the pool of e-probes, known as the probe list. These now curated e-probes can be used to query unassembled metagenomes for the presence or absence of the desired target, while simultaneously ignoring non-target sequences.

There are several benefits to performing EDNA when compared to the traditional methods, such minimizing the bioinformatics steps needed by removing the necessity of read assembly, quality control and annotations. EDNA is theoretically capable of maximum sensitivity because of its ability to detect a target from a single matching e-probe and can be used to detect any number of targets for which e-probes have been designed. To date, EDNA has been validated for the detection of plant viruses (Strobbe et al., 2014), fungi, oomycetes and food borne pathogens of humans (Espindola et al., 2015; Blagden et al., 2015). For EDNA to become an established method for the detection of pathogens in metagenomics data however, more must be learned about its limits. Several important factors must be validated such as the accuracy of the *in silico* models it is based on and the limit of detection for the pathogens that it will be used to detect.

RNA Interference

In 1962, Francis Crick, James Watson and Maurice Wilkins were awarded the Nobel Prize in physiology and medicine for what has been described by some as the discovery that lead

to the birth of molecular biology. With the use of x-ray crystallography, it was possible to discern the structure of nucleic acids of the DNA double helix for the very first time, a discovery that had resounding implications for the future of the biological sciences. For the very first time, scientists were now able to observe the source for the inheritance of traits, even if the exact sequence of nucleic acids leading to these traits remained a mystery. This mystery was solved in 1977 however, when Fredrick Sanger and his team developed the first nucleic acid sequencing method known as Sanger sequencing (Sanger et al., 1977). Sanger sequencing works by utilizing PCR in combination with dideoxy nucleic acids to terminate the polymerization of DNA by DNA polymerase (Sanger et al., 1977). By terminating DNA polymerization in this way and then performing gel electrophoresis, it is possible to identify the number of times and locations that one of the four nucleic acids that DNA is comprised of appears in the sequence (Sanger et al., 1977). This process is then repeated with each other dideoxy nucleic acid until the entire sequence is determined.

While time consuming and limited to relatively short sequences, Sanger sequencing was the first step toward changing the landscape of how the biological sciences were investigated. For example, after the advent of sequencing, a new method for examining the interactions between genes and phenotypes emerged, a process known as forward genetics. Forward genetics is the experimental process to determine where in the genome a specific gene is located that is producing the phenotype of interest. The most commonly used forward genetic methods for determining the function of genes involve exposing the organism of interest to a mutagen, such as chemicals, radiation or transposons, then taking note of any phenotypic changes that occur. Once a change in phenotype of interest has been noted, the next step of the forward genetics approach is to determine where in the genome the mutation occurred that produced the new

phenotype. While forward genetics was instrumental in allowing scientists to determine the function of genes, this approach has two shortcomings that makes its use restrictive. First, because of the random nature of where mutagens alter the genome, targeting the genetic root of specific phenotypes is difficult, especially considering that many of these resulting phenotypes may be the product of more than one genetic mutation. Secondly, forward genetics can be costly as identifying where in the genome a specific mutation occurred is a long and difficult process that required the use of DNA sequencing, which was cost prohibitive.

As the years passed, and DNA sequencing technology improved while its cost decreased, the scientific community began pushing to sequence entire genomes. This goal was met in 1995 with the complete sequencing of *Haemophilus influenza*, the bacterium that was once believed to be the cause of the flu (Fleischmann et al., 1995). This feat was followed six years later with the completion of what is considered by some to be one of the greatest accomplishments in the history of science, the complete sequencing of the human genome (Lander et al., 2001).

With the breakthrough of full genome sequencing came the ability to conduct what is now known as reverse genetics. Reverse genetics is the approach used by scientists to determine gene function by knocking out a gene, or reducing its expression, then examining the resulting phenotype. Unlike forward genetics, reverse genetics is a targeted approach that allows the researcher to use sequencing data to examine the expression of genes one at a time in a more controlled manner. To take advantage of known DNA sequences however, researchers could not use the same techniques found in a forward genetics approach as the mutagens used in forward genetics produced random gene mutations. Therefore, it was necessary for researcher to identify new ways to target specific genes.

Like many other great scientific advancements before, the first recorded use of reverse genetics was performed by accident. In 1986, a group of researchers wanted to make a more purple petunia by inserting a plasmid into the plant expressing the gene for chalcone synthase, an enzyme that is associated with the formation of many of the organic pigments found in plants (Veenstra et al., 1988). Not only did these modified petunias not express more color, but instead were shown to lose almost all color and became white (Napoli et al., 1990). It was not until 1998 that the mechanism causing this downregulation of gene expression was identified, when it subsequently became known as RNA interference (RNAi) (Fire et al., 1998). RNAi is a component of the innate immune response found in almost all eukaryotes that is activated by the introduction of double stranded RNA (dsRNA) into the cell (Judge et al., 2005). Once present, dsRNA triggers an enzyme known as dicer, which cleaves it into what is known as short interfering RNA (siRNA) between 21-25bp in length (Judge et al., 2005). These siRNA are then incorporated into what is known as an RNA induced silencing complex (RISC) by the endonuclease argonaute (Ago) that then uses the now single stranded siRNA as a guide strand to bind to a homologous messenger RNA (mRNA) (Judge, 2005). Once bound, RISC cleaves the mRNA, or inhibits translation by blocking ribosomal activity, thereby reducing gene expression.

Research Justification

Biosecurity can be roughly summarized in two general categories, detection and mitigation. Detection refers to the ability of biosecurity experts, port of entry officials and diagnostic laboratories to rapidly identify and differentiate between economically important pests and pathogens. The scope of maintaining robust biosecurity measures across an entire nation within laboratories, farms, hospitals, restaurants, food processing plants, government facilities, schools and many other economically and socially critical targets is enormous. Considering this,

biosecurity agents have a wide variety of resources at their disposal such as borders, drones, personnel, rapid communication and a wide array of powerful diagnostic tools. These tools include, but are not limited to, microscopes, ELISA, lateral flow assays, PCR, DNA sequencing, genomic databases and bioinformatics. Using these resources, biosecurity officials attempt to manage and screen the massive influx and efflux of goods and people moving into and out of the country each day. However, due to increasing global trade and increased pressure on agriculture from the spread of invasive pests and pathogens, the challenge of maintaining the integrity of national biosecurity is also increasing. To combat theses ever increasing and fluctuating challenges, experts in the field of biosecurity need more diverse, targeted, rapid and cost-effective diagnostic tools.

Mitigation, defined as the actions taken in preparation, response to the entry of, or identification of a pathogen or pest that has passed through a border. The responses that can be taken by biosecurity officials vary greatly depending on the threat posed by the pathogen, the local resources available and the environment in which the pathogen or pest is found. Some examples of responses to the emergence or identification of a pest or pathogen include, but are not limited to, quarantine, the release resistant hosts, chemical or biological treatments, genetically modified organisms and in extreme cases, host eradication. Once a biosecurity threat has been detected, biosecurity officials may utilize one or more of these responses to resolve the occurrence or outbreak as rapidly as possible. As with detection, due to the enormous pressure placed on borders and ports of entry because of increasing trade and international travel, the risk of pathogen and pest introduction increases. Therefore, novel, targeted and efficient procedures and responses must be available to handle the introduction of a wide range of biosecurity threats.

It is for these reasons that the focus of this dissertation will encompass elements of both halves of biosecurity as described above. The first objective will consist of the development of a novel diagnostic tool capable of detecting three economically important viruses of cereal crops, MMV, WSMV and BYDV and discriminating between three species of BYDV, namely PAV, PAS and MAV. This tool utilizes a novel combination of multiplex PCR with high resolution melting to first amplify target regions of three viruses of cereals including BYDV. The primers designed to amplify BYDV will be degenerate and capable of amplifying one of three distinct species. Post amplification melting will then be performed to differentiate between species of BYDV via unique nucleotide polymorphisms between the otherwise identical BYDV species amplicons. This diagnostic multiplex PCR will not only provide diagnosticians a new tool to better distinguish between virus outbreaks in regions where cereal crops are grown, but also provides justification for the coupling of HRM to increase the number of pathogens that can be screened for in a single PCR assay without adding more primer pairs to the reaction.

The second objective will be to utilize a powerful and rapidly expanding diagnostic technology, MPS, and to validate novel uses for it. One emerging technique is electronic probe diagnostic nucleic acid analysis, or EDNA, which is theoretically capable of detecting any number of *in silico* pathogen signatures in unassembled metagenomic and metatranscriptomic databases. The specific goal of this project will be to develop e-probes capable of detecting 21 viruses of wheat. To validate the sensitivity of EDNA, WSMV will be used as a model by isolating pure virus nucleic acids and spiking known concentrations of virus negative host nucleic acids with them. It is hypothesized that by performing a series of dilutions with known ratios of WSMV nucleic acids to host nucleic acids, then sequencing, it will be possible to determine the *in vitro* sensitivity of EDNA to detect virus reads. Up until now, only the

theoretical sensitivity of EDNA has been assessed using mock metagenomic databases generated with the sequencing simulator MetaSim. The justification for this project is derived from the first ever demonstration of the limit of detection for EDNA and its potential use as a quantitative tool by correlating the number of detected virus reads to known reads.

The third and final objective will address the mitigation component of biosecurity by testing the efficacy of RNAi as a treatment to reduce the viral titer of a susceptible wheat variety inoculated with WSMV relative to a WSMV resistant wheat variety. This will be done by designing and synthesizing dsRNA homologous to a WSMV gene necessary to virus replication and injecting a WSMV susceptible wheat variety with it prior to inoculation with WSMV. It is hypothesized that that by triggering the RNAi pathway of a susceptible wheat variety prior to exposure of WSMV, virus replication will be reduced and may lead to a temporary or permanent down-regulation of viral expression in the host. A second component of this project will be to quantify the efficacy of RNAi in virus susceptible wheat varieties when compared to known resistant varieties. This will be accomplished by inoculating dsRNA treated susceptible wheat and non-dsRNA treated resistant varieties with WSMV, then assessing viral titer of each at three time points with RT-qPCR. The justification for this project is two-fold. First, quantifying the effect of a single dose of dsRNA on the viral titer of WSMV-infected wheat is valuable to understand the nature of the RNAi mechanism and if there is the potential for dsRNA as a treatment for virus infected wheat. Secondly, when resistance to WSM is assessed in the field, it is based on qualitative measurement of symptoms, not necessarily viral titer. Utilizing RT-qPCR, it is possible to quantitatively measure virus infection in wheat over time and provide more evidence for the efficacy of virus resistant genes in wheat

REFERENCES

Adams, I. P., Harju, V. A., Hodges, T., Hany, U., Skelton, A., Rai, S. & Ngaboyisonga, C. (2014). First report of maize lethal necrosis disease in Rwanda. *New Disease Reports*, *29*, 22-22.

Bawden, F. C. (1950). *Plant viruses and virus diseases* (Vol. 70, No. 4, p. 332). LWW.

Bennett, C. W. (1940). The relation of viruses to plant tissues. *The Botanical Review*, *6*(9), 427.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H. & Kuehn, M. S. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799-816.

Brakke, M. K. (1987). Virus diseases of wheat. *Wheat and Wheat Improvement*, (wheatandwheatim), 585-624.

Burrows, M., Franc, G., Rush, C., Blunt, T., Ito, D., Kinzer, K. & Ziems, A. (2009). Occurrence of viruses in wheat in the Great Plains region, 2008. *Plant Health Progress*, *10*(1), 14.

Christian, M. L., & Willis, W. G. (1993). Survival of wheat streak mosaic virus in grass hosts in Kansas for wheat harvest to fall wheat emergence. *Plant Disease*, *77*(3), 239-242.

Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., & Walters, L. (1998). New goals for the US human genome project: 1998-2003. *Science*, *282*(5389), 682-689

Curtis, Byrd C., Sanjaya Rajaram, and Macpherson Gómez. (2002). Bread wheat: improvement and production. Food and Agriculture Organization of the United Nations (FAO).

D'Arcy, C. J., & Burnett, P. A. (1995). *Barley yellow dwarf: 40 years of progress*.

Davison, J. (2010). GM plants: science, politics and EC regulations. *Plant Science*, *178*(2), 94-98.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A. & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602-607.

Fahim, M., Ayala-Navarrete, L., Millar, A. A., & Larkin, P. J. (2010). Hairpin RNA derived from viral NIa gene confers immunity to wheat streak mosaic virus infection in transgenic wheat plants. *Plant* B*iotechnology Journal*, *8*(7), 821-834.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(6669), 806-811.

Fleischmann, R. D., Adams, M. D., White, O., & Clayton, R. A. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, *269*(5223), 496.

Gao, J. G., & Nassuth, A. (1993). Alteration of major cellular organelles in wheat leaf tissue infected with wheat streak mosaic rymovirus (Potyviridae). *Phytopathology*, *83*(2), 206-213.

Gibson, W. W., & Painter, R. H. (1957). Transportation by aphids of the wheat curl mite, Aceria tulipae (K.), a vector of the wheat streak mosaic virus. *Journal of the Kansas Entomological Society*, *30*(4), 147-153.

Gildow, F. E. (1987). Virus—membrane interactions involved in circulative transmission of luteoviruses by aphids. In *Current Topics in Vector Research* (pp. 93-120). Springer, New York, NY.

Gill, B. S., Friebe, B., Raupp, W. J., Wilson, D. L., Cox, T. S., Sears, R. G. & Fritz, A. K. (2006). Wheat genetics resource center: the first 25 years. *Advances in Agronomy*, *89*, 73-136.

Gonsalves, D., & Garnsey, S. M. (1989). Cross-protection techniques for control of plant virus diseases in the tropics. *Plant Disease*, *73*(7), 592-597.Hadi, B. A. R., M. A. C. Langham, L. Osborne, and K. J. Tilmon. (2011). Wheat Streak Mosaic Virus on Wheat: Biology and Management. *Journal of Integrated Pest Management* 2(1): J1–J5.

Hasiów-Jaroszewska, B., & Komorowska, B. (2013). A new method for detection and discrimination of Pepino mosaic virus isolates using high resolution melting analysis of the triple gene block 3. *Journal of Virological Methods*, *193*(1), 1-5.

Heyne, E. G. (1987). *Wheat and wheat improvement* (Vol. 2). S. H. Mickelson, W. R. Luellen, & D. R. Buxton (Eds.). Madison, WI: American Society of Agronomy.

Hunter, T. R., Hunt, T., Knowland, J., & Zimmern, D. (1976). Messenger RNA for the coat protein of tobacco mosaic virus. *Nature*, *260*(5554), 759.

Innis, M. A., Myambo, K. B., Gelfand, D. H., & Brow, M. A. (1988). DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proceedings of the National Academy of Sciences*, *85*(24), 9436-9440.

Jiang, W., Garrett, K. A., Peterson, D. E., Harvey, T. L., Bowden, R. L., & Fang, L. (2005). The window of risk for emigration of Wheat streak mosaic virus varies with host eradication method. *Plant disease*, *89*(8), 853-858.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *Elife*, *2*, e00471.

Johansson, M. K., & Cook, R. M. (2003). Intramolecular dimers: a new design strategy for fluorescence-quenched probes. *Chemistry–A European Journal*, *9*(15), 3466-3471.

Jones, R. A., Coutts, B. A., Mackie, A. E., & Dwyer, G. I. (2005). Seed transmission of Wheat streak mosaic virus shown unequivocally in wheat. *Plant Disease*, *89*(10), 1048-1050.

Judge, A. D., Sood, V., Shaw, J. R., Fang, D., McClintock, K., & MacLachlan, I. (2005). Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nature Biotechnology*, *23*(4), 457-462.

Klepper, B., Rickman, R. W., & Peterson, C. M. (1982). Quantitative Characterization of Vegetative Development in Small Cereal Grains 1. *Agronomy Journal*, *74*(5), 789-792.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J. & Funke, R. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860-921.

Large, E. C. (1954). Growth stages in cereals illustration of the Feekes scale. *Plant Pathology*, *3*(4), 128-129.

Lazarowitz, S. G., & Shepherd, R. J. (1992). Geminiviruses: genome structure and gene function. *Critical Reviews in Plant Sciences*, *11*(4), 327-349.

Lee, P. E. (1965). Electron microscopy of inclusions associated with wheat streak mosaic virus. *Journal of Ultrastructure Research*, *13*(3-4), 359-366.

Malmstrom, C. M., & Shu, R. (2004). Multiplexed RT-PCR for streamlined detection and separation of barley and cereal yellow dwarf viruses. *Journal of Virological Methods*, *120*(1), 69-78.

Mao, F., Leung, W. Y., & Xin, X. (2007). Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC biotechnology*, *7*(1), 76.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, *1*(1), 7.

Matthews, R. E. F. (1993). *Diagnosis of plant virus diseases*. Florida (EUA).

McKirdy, S. J., Jones, R. A. C., & Nutter Jr, F. W. (2002). Quantification of yield losses caused by Barley yellow dwarf virus in wheat and oats. *Plant Disease*, *86*(7), 769-773.

Miller, W. A., Waterhouse, P. M., & Gerlach, W. L. (1988). Sequence and organization of barley yellow dwarf virus genomic RNA.

Mullis, K., Faloona, F., Scharf, S., Saiki, R. K., Horn, G. T., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In *Cold Spring Harbor Symposia on Quantitative Biology* (Vol. 51, pp. 263-273). Cold Spring Harbor Laboratory Press.

Murphy, F. A., Fauquet, C. M., Bishop, D. H., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P. & Summers, M. D. (Eds.). (2012). *Virus Taxonomy: Classification and Nomenclature of Viruses* (Vol. 10). Springer Science & Business Media.

Napoli, C., Lemieux, C., & Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *The Plant Cell*, *2*(4), 279-289.

Niblett, C. L., Zagula, K. R., Calvert, L. A., Kendall, T. L., Stark, D. M., Smith, C. E. & Lommel, S. A. (1991). cDNA cloning and nucleotide sequence of the wheat streak mosaic virus capsid protein gene. *Journal of General Virology*, *72*(3), 499-504.

Osman, F., Hodzic, E., Kwon, S. J., Wang, J., & Vidalakis, G. (2015). Development and validation of a multiplex reverse transcription quantitative PCR (RT-qPCR) assay for the rapid detection of Citrus tristeza virus, Citrus psorosis virus, and Citrus leaf blotch virus. *Journal of Virological Methods*, *220*, 64-75.

Pagán, I., & Holmes, E. C. (2010). Long-term evolution of the Luteoviridae: time scale and mode of virus speciation. *Journal of Virology*, *84*(12), 6177-6187.

Royer, T. A., Giles, K. L., Nyamanzi, T., Hunger, R. M., Krenzer, E. G., Elliott, N. C. & Payton, M. (2005). Economic evaluation of the effects of planting date and application rate of imidacloprid for management of cereal aphids and barley yellow dwarf in winter wheat. *Journal of Economic Entomology*, *98*(1), 95-102.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T. & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, *239*(4839), 487-491.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, *74*(12), 5463-5467.

Sadeghi, E., Dedryver, C. A., Riault, G., & Gauthier, J. P. (1997). Variation in transmission of two BYDV-MAV isolates by multiple clones of Rhopalosiphum padi L. *European Journal of Plant Pathology*, *103*(6), 515-519.

Shendure, J. A., Porreca, G. J., Church, G. M., Gardner, A. F., Hendrickson, C. L., Kieleczawa, J., & Slatko, B. E. (2011). Overview of DNA sequencing strategies. *Current Protocols in Molecular Biology*, *96*(1), 7-1.

Singh, K., Wegulo, S. N., Skoracka, A., & Kundu, J. K. (2018). Wheat streak mosaic virus: a century old virus with rising importance worldwide. *Molecular Plant Pathology*, *19*(9), 2193-2206.

S Skoracka, A., Kuczyński, L., Szydło, W., & Rector, B. (2013). The wheat curl mite Aceria tosichella (Acari: Eriophyoidea) is a complex of cryptic lineages with divergent host ranges: evidence from molecular and plant bioassay data. *Biological Journal of the Linnean Society*, *109*(1), 165-180.

Slykhuis, J. T. (1961). Differential hosts and new host records for five grass viruses. *Canadian Journal of Plant Science*, *41*(1), 211-216.

Smith, K. M. (1965). Plant virus-vector relationships. In *Advances in virus research* (Vol. 11, pp. 61-96). Academic Press.

Tatineni, S., Wosula, E. N., Bartels, M., Hein, G. L., & Graybosch, R. A. (2016). Temperature-dependent Wsm1 and Wsm2 gene-specific blockage of viral long-distance transport provides resistance to Wheat streak mosaic virus and Triticum mosaic virus in wheat. *Molecular Plant-Microbe Interactions*, *29*(9), 724-738.

Thomas, J. A., Hein, G. L., & Lyon, D. J. (2004). Spread of wheat curl mite and Wheat streak mosaic virus is influenced by volunteer wheat control methods. *Plant Health Progress*, *5*(1), 2.

Thomas, J. B., Conner, R. L., & Graf, R. J. (2004). Comparison of different sources of vector resistance for controlling wheat streak mosaic in winter wheat. *Crop Science*, *44*(1), 125-130.

Umek, R. M., Lin, S. W., Vielmetter, J., Terbrueggen, R. H., Irvine, B., Yu, C. J. & Chen, Y. P. (2001). Electronic detection of nucleic acids: a versatile platform for molecular diagnostics. *The Journal of Molecular Diagnostics*, *3*(2), 74-84.

Veenstra, J., van der Meer, I. M., & Koes, R. E. (1988). An anti-sense chalcone synthase gene in transgenic plants inhibits flower pigmentation. *Nature*, *333*, 3.

Velandia, M., Rejesus, R. M., Jones, D. C., Price, J. A., Workneh, F., & Rush, C. M. (2010). Economic impact of Wheat streak mosaic virus in the Texas High Plains. *Crop Protection*, *29*(7), 699-703.

Wiedenheft, B., Sternberg, S. H., & Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, *482*(7385), 331-338.

Wittwer, Carl T., and Mark G. Herrmann. (2002). Homogeneous Multiplex Hybridization Analysis by Color and Tm. https://patents.google.com/patent/US6472156B1/en (June 12, 2019).

Zagula, K. R., Niblett, C. L., Robertson, N. L., French, R., & Lommel, S. A. (1992). Potyviridae: genus rymovirus. In *Potyvirus Taxonomy* (pp. 269-276). Springer, Vienna.

CHAPTER III

COMBINING MULTIPLEX PCR AND HIGH-RESOLUTION MELTING FOR THE DETECTION AND DISCRIMINATION OF ARTHROPOD-TRANSMITTED VIRUSES OF CEREALS

Abstract

The Great Plains of the United States is a region comprised of approximately 45 million hectares of grasslands where several economically important cereal crops are grown. Arthropod-transmitted, cereal-infecting viruses vary in incidence from year-to-year and are often difficult to detect in large acreages. To facilitate the detection of economically important viruses of cereals that often exist in co-infections, a multiplex reverse transcriptase PCR (RT-PCR) platform assay was developed. This method can be used in combination with high resolution melting (HRM) to detect and allow for discrimination between three arthropod-transmitted plant viruses; *Wheat streak mosaic virus* (WSMV), *Maize mosaic virus* (MMV) and *Barley yellow dwarf virus* (BYDV). Multiplex PCR in combination with HRM allowed for successful detection of WSMV, MMV, and BYDV, as well as discrimination between three BYDV species, BYDV-PAS, BYDV-PAV and BYDV-MAV. All designed primer pairs amplified products of the predicted size. The BYDV-RT-PCR product amplified an identical product for all three species of BYDV. HRM was then used to discriminate between these products by determining significant differences between the melting rate for each ($p < 0.05$).

1. Introduction

Cereal crops such as maize (*Zea mays* L.), wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) are planted in large acreages and represent a significant economic investment each year in the United States. In 2016, 90 million acres of maize were planted and in 2017 the reported economic value of wheat and barley combined was approximately $10 billion dollars (U.S.) (USDA, 2017). Most of these economically important crops are grown in the Great Plains of the United States, a region that includes parts of Montana, North Dakota, South Dakota, Wyoming, Minnesota, Nebraska, Colorado, Kansas, New Mexico, Texas and Oklahoma. In Oklahoma, wheat is one of the state's most important exports, with approximately five million acres valued at $471 million (U.S.) dollars in 2016 alone (USDA, 2017). Every year, several factors contribute to significant yield losses including, but not limited to, abiotic damage caused by wind and ice, competition for nutrients with weeds and native plant species, under fertilized or dry soil, insect damage and plant diseases (Armbrust et al., 1974; Martin & McNamara, 1987; Cassman, 1999; Oerke, 2006; Anderson et al., 2004).

Arthropod-transmitted viruses are among the most common pathogens of cereals and contribute to significant crop yield losses each year, with *Wheat streak mosaic virus* (WSMV) alone causing losses up to $464.5 (U.S.) per hectare (Byamukama et al., 2014). *Maize mosaic virus* (MMV), WSMV and *Barley yellow dwarf virus* (BYDV) are three viruses of commonly grown cereal crops found in the United States (Burrows et al., 2009; McDaniel et al., 1985). Detection and discrimination of these viruses in the field is difficult due to the diversity, levels of severity and expression of symptoms, especially considering that many viral infections of cereals present similar symptoms such as chlorosis, streaking and stunting (Bos, 1970). Diagnostics can be further complicated when different species or strains of the same virus are found within the

same region and exist in co-infections, such as in the case of BYDV in the Great Plains (Hall and Little, 2013). Currently, there are three known species of BYDV, BYDV-PAV, BYDV-MAV and BYDV-PAS, each of which is transmitted by different aphid vectors (Ali & Hameed, 2017; Li et al., 2001; Gildow, 1987).

One of the challenges faced by plant diagnosticians is the lack of specific, sensitive and flexible diagnostic tools capable of detecting and discriminating multiple viruses at once. Reverse transcriptase polymerase chain reaction (RT-PCR) is currently considered the gold standard of viral diagnostics both in terms of sensitivity and target specificity and has the capability of being utilized in multiplex assays. However, RT-PCR is relatively expensive when compared to other single-target diagnostic tools such as enzyme-linked immunosorbent assay (ELISA) and has limits to the number of targets that can be screened for in a single assay due to the potential for off target amplification and primer-primer interactions. Therefore, when developing a RT-PCR based diagnostic assay, it is important to maximize the number of targets that can be screened for in a single reaction while minimizing the number of primer pairs included in each. One possible solution for increasing the number of targets in a RT-PCR based assay without increasing the number of primers used, is to supplement the RT-PCR assay with another nucleic acid based diagnostic tool.

High resolution melting (HRM) is a post PCR analysis in which the PCR product is denatured by heat over time, where the rate of denaturation of an individual amplicon is dependent upon its putative nucleic acid composition (Liew et al., 2008). The rate of denaturization is measured by a decline in fluorescence from a double-stranded DNA (dsDNA) specific intercalating dye added to the PCR solution (Wittwer et al., 2003). By using a

combination of multiplex RT-PCR and HRM with specifically designed primers that simultaneously target multiple virus species, it may be possible to expand the detection capability of the number of viruses that can be screened for in a single RT-PCR assay. This application of RT-PCR is achievable in combination with HRM, providing diagnosticians with a more flexible tool that can be used to detect viruses at different taxonomic levels. The purpose of this study was to develop a RT-PCR based diagnostic tool to detect the economically important viruses MMV, WSMV and BYDV, then, if necessary, further discriminating between the three-known species of BYDV, BYDV-PAV, BYDV-MAV and BYDV-PAS with HRM analysis.

2. Materials and Methods

2.1. Primer Design and HRM Melting Predictions

Nucleic acid sequences used to design primers for MMV (NC_005975), WSMV (NC_001886), BYDV-PAV (NC_002160.2), BYDV-MAV (NC_002160.2) and BYDV-PAS (NC_002160.2) were retrieved from NCBI. For BYDV, species sequences for PAV, PAS and MAV were aligned using Geneious (v.9.0.4) (Kearse et al., 2012) and degenerate primers were designed using the consensus gene region of RNA dependent RNA polymerase (RDRP) (NC_002160.2). Primers specific to MMV, WSMV and all three species of BYDV were generated from using the sequences listed above as a template in Primer 3 (v.4.0.0) (Untergasser et al., 2012). The HRM melting temperature (HRM Tm) of each expected PCR product was predicted with uMELT ^{SM} (Dwight et al., 2011). To serve as an internal control, a primer pair described by Jarošová and Kundu (2010), which amplifies an 84bp region of the exon for beta-tubulin (TUBB) in the family Poaceae was used. Primers generated for each of the targeted virus species were then uploaded to the NCBI Primer BLAST database to predict *in silico* primer specificity.

2.2. RNA Extraction and Reverse Transcription

Reference positive and negative controls for WSMV, MMV and all three species of BYDV were obtained from Agdia (Elkhart, IN). Total RNA was extracted from each Agdia positive control using TRIzol (Thermofisher, Waltham, MA) as per the manufacturer's protocol. The extracted RNA was reverse transcribed using random hexamer primers and MMLV reverse transcriptase (Promega, Madison, WI). Complimentary DNA (cDNA) synthesized this way was used for PCR and stored in nuclease free water at -20°C.

2.3. Gradient and Sensitivity PCR Assays

To determine the optimal annealing temperature for each primer pair, 20µL PCR reactions were performed as follows: 10µL GoTaq (Promega, Madison, WI), 3µL nuclease free water, 2µL of each forward and reverse primer (5µM) and 3µL of cDNA. The thermocycler conditions for each reaction were as follows: four min at 94°C, 40 cycles of 20 sec at 94°C, 20 sec at a 56°C-63°C gradient and 20 sec at 72°C, followed by a final extension for five min at 72°C. PCR was performed in a gradient capable Biometra Thermocycler (Biosciences, Dublin, Ireland). PCR products were then visualized using gel electrophoresis. To test the sensitivity of each primer pair, cDNA generated from Agdia positive controls for MMV, WSMV, BYDV-PAV, BYDV-MAV and BYDV-PAS were quantified with nanodrop (Thermo Fisher, Waltham, MA) and diluted in a series of 10-fold dilutions ranging from 10ng/µL to 1fg/µL. Each dilution was then used as template for a PCR of each primer pair using the protocol described above with an annealing temperature of 58°C. Each PCR experiment was performed three times. Gel electrophoresis with SYBR™ Safe stain (Invitrogen, Carlsbad, CA) was then used to vitalize PCR products in a Gel Doc™ XR+ (Biorad, Hercules, CA).

2.4. Assessment of Primers in Multiplex

To test the capability of the primers for MMV, WSMV and three species of BYDV to react in multiplex, a 100μM stock primer solution was prepared by mixing 1μL of each stock primer in 294μL of water. This stock primer solution was then used in a multiplex PCR master mix containing 12.5μL of GoTaq, 4.75μL nuclease free water, 3.75μL primer solution, and 4μL template cDNA per 25μL reaction. The cycling conditions used for multiplex PCR were 94°C for 5 min, 40 cycles of 94°C for 20 sec, 57°C for 20 sec, 72°C for 20 sec, followed by a final extension of 72°C for 5 min. PCR was performed as described above with template from each target pathogen individually, as well as together in a mock co-infection. Each single target and multiplex PCR was performed three times each.

2.5 Primer Specificity

The specificity of each primer pair was tested in multiplex using cDNA template from 12 other commonly occurring pathogens of cereal crops including: *High Plains wheat mosaic emaravirus* (WMoV), *Maize chlorotic mottle virus* (MCMV), *Wheat spindle streak mosaic virus* (WSSMV), *Triticum mosaic virus* (TriMV), *Barley stripe mosaic virus* (BSMV), *Maize dwarf mosaic virus* (MDMV), *Maize stripe virus* (MSpV), *Johnsongrass mosaic virus* (JGMV), *Cereal yellow dwarf virus* (CYDV-RPV), *Maize white line mosaic virus* (MWLMV), *Soil borne wheat mosaic virus* (SBWMV) and Corn stunt spiroplasma (Css). Healthy wheat tissue was used as a negative control. All positive and negative controls were obtained from Agdia and cDNA for each was prepared as described in section 2.2. The resulting cDNA of each sample was then used as template in multiplex PCR using the protocol and thermocycler profile described in section 2.4. Gel electrophoresis was used to assess if any non-specific, off-target PCR products were amplified. All amplified PCR products were excised from the gel and purified using a QIAquick

(QIAgen, Hilden, Germany) gel extraction kit and sequenced at the core facility at Oklahoma State University using a 3730XL genetic analyzer (Thermo Fisher, Waltham, MA).

2.6. HRM Discrimination of Three Species of BYDV

RNA extracted from reference positive controls of BYDV-PAV, BYDV-MAV and BYDV-PAS were reverse transcribed as described in section 2.2 and used as the template in quantitative PCR (qPCR). qPCR reactions of 20µL were comprised of: 10µL of Hot Start master mix (Thermofisher, Waltham, MA), 5µL nuclease free water, 1µL of each 5µM forward and reverse primer, 2µL LC Green (BioFire, Salt Lake City, UT) and 1µL of cDNA. PCR was performed in a Corbett Research Rotorgene (QIAGEN, Hilden, Germany) and the thermocycler profile used was 94°C for 5 min, 40 cycles of 94°C for 20 sec, 57°C for 20 sec, 72°C for 20 sec, followed by a final extension of 72°C for 5 min, followed by a HRM ramping from 50°C to 90°C, increasing by 0.25°C every second. The HRM species discrimination of BYDV-PAV, BYDV-PAS and BYDV-MAV was assessed by melting points using allelic discrimination analysis and by normalizing the melting curves of each BYDV species.

3. Results

3.1. Primer Design and HRM Melting Predictions Results

Three primer pairs were designed to be specific for the cereal infecting viruses WSMV, MMV and three species of BYDV. Table one compiles thermodynamic features of designed primer sequences such as: annealing temperature (Tm), predicted amplicon melting point (HRM Tm), oligo self-complimentary score (Any) and oligo 3′ self-complementarity at the 3´ termini (3′). Primers for MMV amplify a segment of 307bp (GenBank: AY618418.1) annealing from positions 2123 to 2429 of the MMV genome. Primers for WSMV amplify a 198bp fragment of the WSMV genome (GenBank: AF285169.1) annealing from nucleotide positions 5444 to 5641.

Degenerate primers designed based on the consensus of BYDV-PAV (GenBank: EF043235.1), BYDV-PAS (GenBank: AF218798.2) and BYDV-MAV (GenBank: D11028.1) amplify a 150bp region of BYDV annealing from nucleotide positions 2625 to 2774.

Table 1. Table of primer sequences and corresponding thermodynamic values calculated by Primer 3 (v.0.4.0).

Target	Primer	Primer Sequence (5′ - 3′)	Length	Tm(°C)	Size (bp)	HRM Tm(°C)	Any	3′
MMV	MMV F	CACTTCACACGACCTTTGCA	20	58.99	307	84.0	0	0
	MMV R	CTCGTCTTAAATTGCGCCGA	20	59.0			1.68	0.13
WSMV	WSMV F	CGACAATCAGCAAGAGACCA	20	57.92	198	82.0	0	0
	WSMV R	TGAGGATCGCTGTGTTTCAG	20	57.92			0	0
BYDV	BYDV F	GAGMGGTACTWCGACRRTCT*	20	53.46-59.46	150	83.0 (PAV) 82.5 (PAS) 81.5 (MAV)	4-5	0-3
	BYDV R	CCTATYCCAAACCCRGCTAA*	20	56.88-63.19			4	1
Internal Control	TUBB F	CAAGGAGGTGGACGAGCAGATG	22	62.93	84	83.5	0	0
	TUBB R	GACTTGACGTTGTTGGGGATCCA	23	62.93			0	0

*Sequence may contain the degeneracy M (A or C), W (A or T), R (A or G) and Y (C or T).

3.3. Multiplex PCR Temperature Gradient and Sensitivity Results

Primers designed for MMV, WSMV and the degenerate primers for the three species of BYDV, successfully amplified their respective expected products, which matched predicted lengths. Gradient PCR was used to assess the optimal annealing temperature of each designed primer pair (Fig. 1). Each of the primer pairs performed according to optimal thermodynamic values in table 1. The annealing temperatures studied ranged from 56-63°C, producing a clear, bright PCR product of the predicted length. When used to amplify cDNA from BYDV-PAV, the primer pair for BYDV produced lower DNA yields when used at annealing temperatures above 61°C as determined by visual assessment of DNA yields. Gradient PCR experiments were also performed in multiplex PCR using all three primer pairs and cDNA template of all three target

viruses, MMV, WSMV and BYDV-PAV. The amplification of all three targets was successful, with products of expected lengths present across annealing temperatures ranging from 56-63°C (Fig. 2). Bright, clear bands were seen across all annealing temperatures, with products for MMV (300bp) and WSMV (200bp) appearing slightly brighter across all annealing temperatures compared to amplicon for BYDV-PAV (150bp).

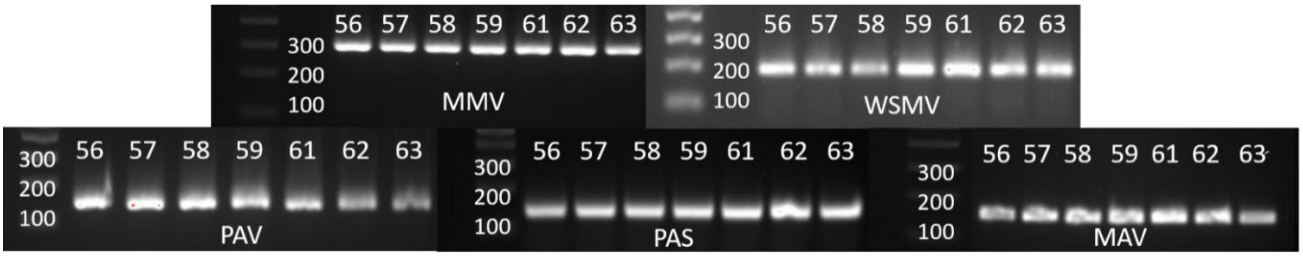


Figure 1. Gradient PCR of each primer pair and target template cDNA from 56-63°C. Gel electrophoresis was performed on a 2% agarose gel in 1% TAE at 95V for approximately 45 min, along with a 100bp ladder (New England Biolabs, Ipswich, MA).

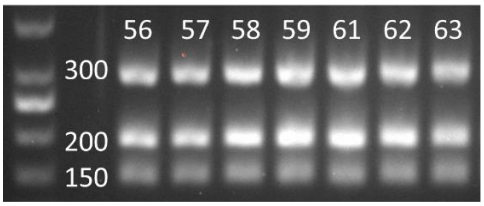


Figure 2. Gradient PCR results for multiplex PCR of cDNA template for MMV (300bp), WSMV (200bp) and BYDV-PAV (150bp) from 56-63°C. Gel electrophoresis was performed on a 2% agarose gel in 1% TAE at 95V for approximately 45 min, alongside a 50bp ladder (New England Biolabs, Ipswich, MA).

The detection sensitivity was determined for all sets of primers with serial dilutions made from 10ng/µL to 1fg/µL. Template used for each sensitivity assay was from known concentrations of gel purified PCR product of each target virus. Each primer pair successfully amplified the expected product down to a template concentration of 1fg/µL (Fig. 3). As expected, all primer pairs amplified the predicted corresponding product and decreasing product

fluorescence intensity that correlated with the corresponding concentrations of target template. It was possible to visually confirm each band at all template concentrations tested. Except for the template for BYDV-MAV (Fig. 3, MAV), no PCR artifacts occurred as target concentrations decreased. At the lowest template concentration of BYDV-MAV tested, 1fg/µL, a second, faint, 100bp product was present. This band was excised using an Illustra™ GFX™ PCR DNA and gel band purification kit (GE Healthcare, Chicago, IL) and sequenced. BLAST analysis results of sequencing revealed that this 100bp product most closely aligned with the complete genome sequence of BYDV-GAV (NCBI accession: KF523381.1), a known strain of BYDV closely related to BYDV-MAV (Jin et al., 2004).

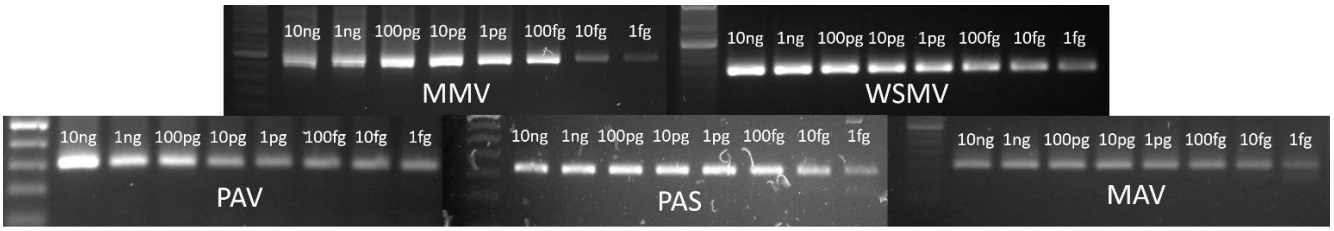


Figure 3. Results for PCR testing the sensitivity of primers for MMV, WSMV, BYDV-PAV, BYDV-PAS, and BYDV-MAV from concentrations of 10ng/µL to 1fg/µL. Gel electrophoresis was performed for each on a 2% agarose gel in 1% TAE at 95V for 45 min.

3.4. Primer Specificity Results

To determine the potential for off-target amplification of non-target pathogens, the target virus primers were used in PCR with positive control cDNA template for MMV, WSMV, all three species of BYDV, 12 commonly occurring pathogens of cereals, a virus negative wheat control and a no template control (Fig. 4). The results of this specificity assay demonstrated that the designed primers produced products of predicted sizes for MMV, WSMV and BYDV-PAV together (lane 1), MMV alone (lane 2), WSMV alone (lane 3), BYDV-PAS alone (lane 4) and BYDV-MAV alone (lane 5). Non-predicted products were detected in lanes 12 and 16

corresponding to positive control template for MSpV and SBWMV respectively. Sequencing results indicated that the bands at 300bp, 200bp and 150bp found in MSpV, corresponded to BLAST results for MMV, WSMV and BYDV-PAV respectively. In SBWMV, sequencing results for the amplicon detected at 150bp had BLAST results corresponding to BYDV-PAV. The sequencing results indicated that the reference positive control for MSpV and SBWMV contained tissue with mixed infections.

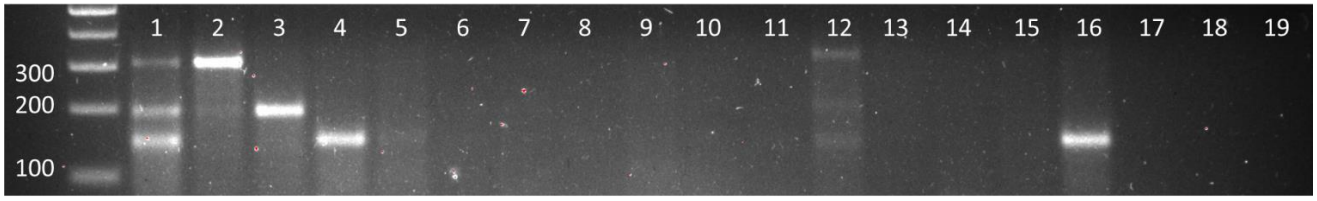


Figure 4. Results of multiplex PCR specificity assay testing primers for MMV, WSMV and three species of BYDV against cDNA template for 12 commonly occurring pathogens of cereal crops. From left to right, lanes contain; 100bp ladder (New England Biolabs, Ipswich, MA), template from positive controls for a combination of MMV, WSMV and BYDV-PAV (1), MMV only (2), WSMV only (3), BYDV-PAS (4), BYDV-MAV (5), WMoV (6), MCMV (7), WSSMV (8), TriMV (9), BSMV (10), MDMV (11), MSpV (12), JGMV (13), CYDV-RPV (14), MWLMV (15), SBWMV (16), CSS (17), negative control (18) and no template control (19). Gel electrophoresis was performed on a 2% agarose gel in 1% TAE at 95V for 45 min.

3.6. HRM of BYDV Amplicon and Species Differentiation

Following RT-PCR amplification of each species of BYDV, BYDV-PAV, BYDV-PAS and BYDV-MAV, each amplicon underwent HRM and allelic discrimination. HRM and allelic discrimination revealed clear differentiation in the rates at which PCR products of each species of BYDV degraded as temperature increased (Fig. 5, A). Allelic discrimination analysis of the three species of BYDV revealed significant differences in the rates of denaturation due to increasing temperature (Fig.5, B, C and D). PCR amplicon of BYDV-PAV differentiated from BYDV-PAS and BYDV-MAV with a confidence between 98.59% and 98.95% ($p < 0.05$), while

amplicon for BYDV-PAS was reported to differentiate from BYDV-PAV and BYDV-MAV with confidence between 90.87% and 99.80% ($p < 0.05$). Finally, amplicon of BYDV-MAV was reported to differentiate from -PAV and -PAS with a confidence between 95.19% and 99.10% ($p < 0.05$).

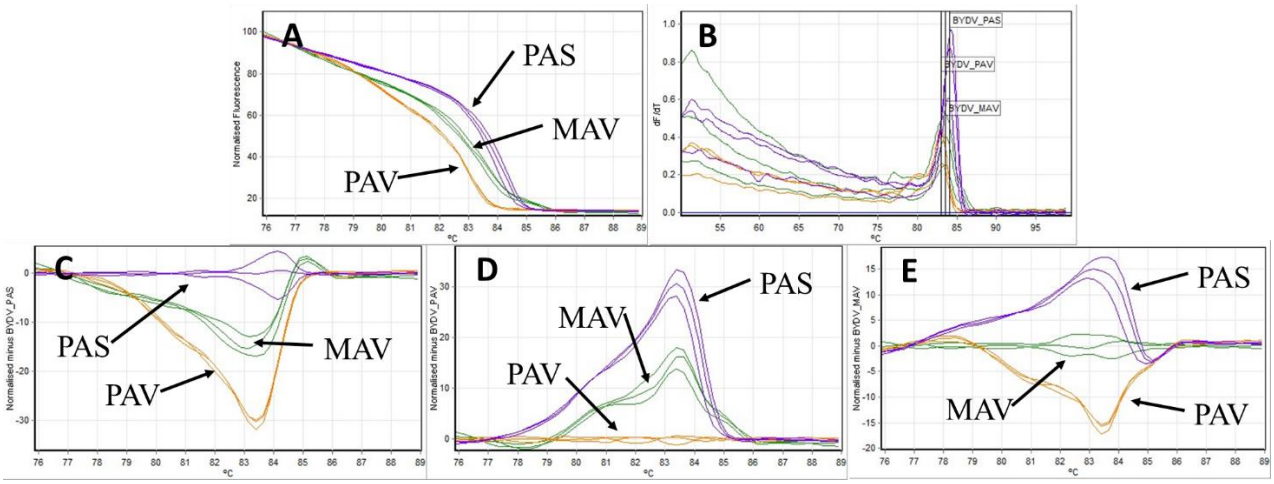


Figure 5**.** Results of HRM discrimination (A) of BYDV-PAV (Yellow), BYDV-PAS (Purple) and BYDV-MAV (Green), HRM df/dT melting curve analysis (B) and allelic discrimination of three species normalized to BYDV-PAS (C), BYDV-PAV (D) and BYDV-MAV (E). Each graph plots the change in normalized fluorescence to the change in temperature from 76°C to 89°C.

4. Discussion

The purpose of this study was to develop a robust, accurate and flexible diagnostic tool capable of detecting and discriminating between three economically important viral pathogens of cereal crops, MMV, WSMV and BYDV. These pathogens can be found in multiple co-infections within their hosts (maize, wheat and barley) and present with similar symptoms making visual diagnosis difficult. The primers designed and tested in this study were developed as a tool to provide diagnosticians the ability to detect MMV, WSMV and BYDV in multiplex and the flexibility of differentiating between three species of BYDV using HRM. The results of gradient PCR assays demonstrated that the designed primers for MMV, WSMV and BYDV amplified the

expected and produced products of the predicted size in individual reactions and in multiplex across a range of annealing temperatures. This broad range of annealing temperatures at which the assay functions allows the multiplex to be performed across a variety of thermocycling equipment with confidence as minor differences in measured temperatures would be unlikely to significantly alter results.

Sensitivity was assessed for each primer pair and each target virus by performing a 10-fold serial dilution. Using the template for BYDV-MAV at a concentration of 1fg/µL, a second band of 100bp in length was noted. After gel excision and sequencing, NCBI nucleotide BLAST results showed that this unexpected band aligned most closely with BYDV-GAV, a strain of BYDV closely related to BYDV-MAV. This BLAST result is possibly due to a combination of the similarity in nucleotide sequences between BYDV-MAV and BYDV-GAV, as well as the higher rate of sequencing error that occurs when sequencing a product less than 100bp (Schuster 2007). This combination of error and high sequence similarity with BYDV-MAV may have resulted in a false positive result for BYDV-GAV. The effect this unexpected band has on the multiplex assay is limited however, as the product was only present at the lowest concentration tested, is smaller than any expected bands and therefore would not produce a false positive result.

A specificity assay was performed using positive control template for commonly occurring pathogens of cereals to determine if any non-specific products may be produced within a host with multiple infections. Among the pathogens that the assay was tested against, unexpected bands were noted in two, MSpV and SBWMV. Sequencing results indicated that these unexpected bands were products of the target pathogens (MMV, WSMV and BYDV),

within the reference control for MSpV and SBWMV. The positive controls used for this study were lyophilized, certified positive controls from Agdia, inc. While Agdia certified that their controls were positive for the advertised pathogen, they did not guarantee that it was the only pathogen they are positive for. This means it was possible for the template for MSpV and SBWMV to also contain sequences for MMV, WSMV and BYDV, as these are other viruses that Agdia generates positive controls for. The controls used in this study are designed by Agdia for use in ELISA based assays and not necessarily for PCR.

One of the primary goals of this study was to develop a method for increasing the range of targets that can be screened for in a single multiplex PCR assay without increasing the number of primer pairs used during the reaction. To accomplish this, multiplex PCR was combined with HRM to distinguish between three otherwise identical products generated by the multi-target, degenerate BYDV primers designed in this study. These primers were designed to provide the operator of the assay the flexibility to detect the presence of BYDV in a sample or identify the species of BYDV if necessary. For many diagnostic laboratories, identifying the genus of a virus is sufficient, in which case the use of the BYDV primers in multiplex is all that would be required. If more information is needed on the species of BYDV present however, as may be the case in an epidemiological study, these same BYDV primers can be used in qPCR and HRM to discriminate between the three species of BYDV

REFERENCES

Ali, M., Tahir, M., & Hameed, S. (2017). Phylogenetic and genome-wide pairwise distance analysis of the genus Luteovirus. *Pakistan Journal of Agricultural Sciences*, *54*(2).

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, *19*(10), 535-544.

Armbrust, D. V., Paulsen, G. M., & Ellis, R. (1974). Physiological responses to wind-and sandblast-damaged winter wheat plants. *Agronomy Journal*, *66*(3), 421-423.

Bos, L. (1970). Symptoms of virus diseases in plants. *Symptoms of Virus Diseases in Plants.*, (Ed. 2).

Burrows, M., Franc, G., Rush, C., Blunt, T., Ito, D., Kinzer, K., Olson, J., O'Mara, J., Price, J., Tande, C., Ziems, A., Stack, J. (2009). Occurrence of viruses in wheat in the Great Plains region, 2008. *Plant Health Progress doi*, *10*.

Byamukama, E., Wegulo, S. N., Tatineni, S., Hein, G. L., Graybosch, R. A., Baenziger, P. S., & French, R. (2014). Quantification of yield loss caused by Triticum mosaic virus and Wheat streak mosaic virus in winter wheat under field conditions. *Plant Disease*, *98*(1), 127-133.

Cassman, K. G. (1999). Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. *Proceedings of the National Academy of Sciences*, *96*(11), 5952-5959.

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.

Gildow, F. E. (1987). Virus—membrane interactions involved in circulative transmission of luteoviruses by aphids. In *Current Topics in Vector Research* (pp. 93-120). Springer, New York, NY.

Hall, G. S., & Little, D. P. (2013). Within-host competition between barley yellow dwarf-PAV and-PAS. *Virus research*, *174*(1), 148-151.

Jarošová, J., & Kundu, J. K. (2010). Validation of reference genes as internal control for studying viral infections in cereals by quantitative real-time RT-PCR. *BMC Plant Biology*, *10*(1), 146.

Jin, Z., Wang, X., Chang, S., & Zhou, G. (2004). The complete nucleotide sequence and its organization of the genome of Barley yellow dwarf virus-GAV. *Science in China Series C: Life Sciences*, *47*(2), 175-182.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. & Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3 Bioinformatics 23(10):1289-91.

Li, C., Cox-Foster, D., Gray, S. M., & Gildow, F. (2001). Vector specificity of barley yellow dwarf virus (BYDV) transmission: identification of potential cellular receptors binding BYDV-MAV in the aphid, Sitobion avenae. *Virology*, *286*(1), 125-133.

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.

Martin, R. J., Cullis, B. R., & McNamara, D. W. (1987). Prediction of wheat yield loss due to competition by wild oats (Avena spp.). *Australian Journal of Agricultural Research*, *38*(3), 487-499.

McDaniel, L. L., Ammar, E. D., & Gordon, D. T. (1985). Assembly, morphology, and accumulation of a Hawaiian isolate of maize mosaic virus in maize. *Phytopathology*, *75*(10), 1167-1172.

Oerke, E. C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, *144*(1), 31-43.

Pearce, R. S. (2001). Plant freezing and damage. *Annals of Botany*, *87*(4), 417-424.

Scheets, K. (1998). Maize chlorotic mottle machlomovirus and wheat streak mosaic rymovirus concentrations increase in the synergistic disease corn lethal necrosis. *Virology*, *242*(1), 28-38.

Schuster, S. C. (2007). Next-generation sequencing transforms today's biology. *Nature Methods*, *5*(1), 16.

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.

Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G., & Pryor, R. J. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. Clinical chemistry, 49(6), 853-860.

CHAPTER IV

THE VALIDATION OF *IN SILICO* PREDICTIVE MODELS FOR THE LIMIT OF DETECTION OF *WHEAT STREAK MOSAIC VIRUS* WITH ELECTRONIC PROBE DIAGNOSTICS

Abstract

Electronic probe diagnostic nucleic acid analysis (EDNA) is a massive parallel sequencing (MPS) based diagnostic tool that uses highly curated, target-specific electronic probes (e-probes), 30-120 nucleotides in length to query unassembled metagenomic databases for the presence of target DNA sequences. E-probes specific to 21 viral pathogens of wheat were designed, curated and validated with mock metagenomic databases generated using MetaSim and transcriptomic data generated from wheat tissue. *In silico* detection of all 21 viral pathogens of wheat and discrimination between three species of *Barley yellow dwarf virus* (BYDV) was possible in both single and multiple infections. The *in silico* limit of detection was generated using *Wheat streak mosaic virus* (WSMV) by comparing the number of target e-probe 'hits' to the percentage of WSMV present in a series of mock metagenomes containing varying ratios of WSMV to host nucleic acids. Isolated WSMV DNA was mixed in known ratios with wheat DNA in a serial dilution, sequenced with Illumina and screened with EDNA to determine the ratio of target e-probe 'hits' to percentage of WSMV nucleic acids in solution. *In vitro* models for the limit of detection of WSMV with EDNA produced similar results to those produced with *in silico* models. Overall, *in vitro* models were shown to produce significantly more e-probe 'hits' at every WSMV read abundance tested than *in silico* models ($p < 0.05$).

1. Introduction

Wheat (*Triticum aestivum*) is one of the most important plants ever to be cultivated in terms of human nutrition and economic value (Curtis et al., 2002). In November of 2016, it was estimated that the total production of wheat worldwide was expected to reach 746.7 million tons, an increase of 9.24 million tons from what was produced in 2015 (Food and Agricultural Organization, 2016). Even with a steady growth of wheat production worldwide, there is also an increasing need for more as the population of humans on the planet continues to rise. Since 1950 the global population has tripled, leading to a population of approximately 7.5 billion people reported in 2016, a number that is predicted to reach 9.9 billion by 2050 (United Nations, 2016). With this increasing need for food, the fields of agricultural biosecurity and diagnostics become more important to help manage crop yield losses through the rapid and accurate detection of agricultural pathogens. This is particularly important in the case of wheat because approximately half of all of the wheat worldwide is produced and consumed in developing countries (Aquino et al., 2001).

Some of the greatest threats to agricultural biosecurity worldwide are plant diseases. In wheat for instance, many of the most devastating losses occur each year due to diseases caused by plant viruses (Byamukama et al., 2014). As of 2016, there were a total 47 viruses that are currently known to infect wheat (American Phytopathological Society, 2016). In field trials conducted in 2011 and 2012, wheat losses due to plant viruses were estimated at approximately 4% annually (Appel et al, 2015). When taking into account the average value of wheat for these two years and the acreage that was planted, it is estimated that U.S. farmers lose approximately $35 million as a direct result of plant viruses each year (USDA-NASS, 2014). Early and accurate

diagnosis is an essential first step to determine the appropriate action experts must take to mitigate the effects of plant viruses.

Traditional methods for the detection and discrimination of viruses of wheat include ELISA, PCR and lateral flow assays. While techniques like PCR are capable of being used for the simultaneous detection of multiple pathogens, there is a finite number of targets that can be included in a single reaction (Souaze et al., 1996). An ideal detection method would be capable of detecting any number of pathogens simultaneously from a single sample. For this reason, efforts have been made to shift the focus of diagnostics to high-throughput molecular methods such as massive parallel sequencing (MPS). Due to the relatively large volume of data generated from MPS, it is often coupled with bioinformatics and computational technology to perform a wide range and variety of data analysis (Prabha et al., 2013). One such data analysis method is the adapted bioinformatic pipeline strategy known as e-probe diagnostic nucleic acid analysis (EDNA), used to detect pathogens in a database composed of unassembled nucleic acid nucleic acid sequence reads (Umek et al., 2001). These pathogens are detected by querying the sample database with highly curated, electronic pathogen-specific sequences know as electronic probes (e-probes). EDNA has been successfully employed in the detection of pathogens from several taxonomic kingdoms within a variety of sample substrates (Strobbe et al., 2014; Espindola et al., 2015; Blagden et al., 2015).

EDNA is theoretically capable of perfect sensitivity, as it can be used to detect the presence of a pathogen from a single MPS read. However, to properly validate EDNA as an accurate, rapid and reliable tool for the detection of pathogens, its diagnostic limits such as the minimum threshold for detection must be understood. Currently, EDNA limits of detection are

established through *in silico* sequencing simulator modeling programs such as MetaSim (Richter et al., 2008). One benefit of MetaSim is that it can be used to simulate metagenomic databases at no cost, facilitating the generation, curation and testing of e-probes at a significantly reduced cost compared to *in vitro* MPS. What is not currently known is how accurately the *in silico* predictive models generated using MetaSim match *in vitro* sequencing data for use with EDNA.

To date, EDNA has only been used to assess presence or absence data of a pathogen within a metagenomic sample and not quantitative data to assess relative pathogen concentration. If it can be shown that the total number of e-probe matches correlates predictively to a known concentration of target reads within a metagenomic sample, it would be evidence to support the use of EDNA as a quantitative tool for the detection of pathogens. If it is determined that EDNA can be used to produce consistent quantitative results, it may be possible to establish a predictive model to determine the minimum amount of target that must be present in the sample before it can be detected. A potential flaw of using simulated metagenomes to test the potential of EDNA as a quantitative tool is the fact that viruses are known to have differentially expressed genes during varying stages of infection (Whitham et al., 2003). While simulated metagenomes can produce consistent virus to host nucleic acid ratios, they do no account for variation in differential viral gene expression within the host prior to sequencing. To establish EDNA as a quantitative diagnostic tool, it must first be established if EDNA results accurately reflect variations that may occur within a metatranscriptome due to differentially expressed virus genes.

This information would provide diagnosticians the knowledge needed to make more informed decisions regarding the appropriate number of reads, number and length of e-probes or sequencing coverage and depth needed to accurately detect a pathogen within a sample. The first

objective of this study is to develop e-probes specific to 21 of the most commonly occurring viral pathogens of wheat and to determine if EDNA can be used to detect them within simulated metagenomes in single infections and multiple infections. The second objective will be to use WSMV as a model to validate the predicted sensitivity of simulated metagenomes. The basis for this analysis will be to determine if the total number e-probes that match with, or 'hit', their specific target within a metagenomic database correlates to the known percentage of WSMV nucleic acids in the sample.

It is hypothesized that by sequencing known percentages of target WSMV nucleic acid mixes, then performing EDNA on the resulting metagenomic databases, it will be possible to determine the accuracy of *in silico* sequencing simulator models for use with EDNA. The third and final objective of this study will be to determine if EDNA can be used to distinguish between differentially expressed virus genes. This will be done by amplifying the genome of WSMV in two approximately equal segments with PCR, then mixing them in a ratio of two to one. This mixture will then be combined with background host nucleic acids, sequenced and tested with EDNA. If EDNA can be used to distinguish between differentially expressed virus genes within a metagenome, the number of e-probe 'hits' for one of the two segments will be greater than 'hits' from probes specific to the other.

2.0 Materials and Methods

2.1 E-Probe Design

In silico, pathogen-specific sequences of 30 base pairs in length were generated for each of the twenty-one wheat infecting viruses used for this study. The sequences used to generate the e-probes for each of these viruses were retrieved from the National Center for Biotechnology Information (NCBI; Bethesda, MD). A list of the viruses used to generate e-probes and their

corresponding NCBI accession number can be found below in Table 1. One virus, *Barley yellow dwarf virus* (BYDV), and the three species that it describes, BYDV-PAV, BYDV-PAS and BYDV-MAV, were chosen as models to determine if EDNA could distinguish between closely related species of a virus within single and mixed infections. Therefore, e-probes for BYDV were designed for each of the three species individually, as well as in combination.

Table 1. Names, acronyms and NCBI accessions of viruses used to generate e-probes.

Virus Name	Acronym	NCBI Accession Number(s)
Agropyron mosaic virus	AgMV	NC_005903.1
Brome Mosaic Virus	BMV	NC_002026.1; NC_002027.1; NC_002028.2
Barley Stripe Mosaic Virus	BSMV	NC_003469.1; NC_003478.1; NC_003481.1
Barley yellow dwarf virus MAV	BYDV-MAV	NC_003680.1
Barley yellow dwarf virus PAS	BYDV-PAS	NC_002160.2
Barley yellow dwarf virus PAV	BYDV-PAV	NC_004750.1
Barley yellow striate mosaic virus	BYSMV	NC_028244.1
Cocksfoot mottle virus	CoMV	NC_002618.2
Cereal yellow dwarf virus-RPV	CYDV-RPV	NC_004751.1
Digitaria striate mosaic virus	DiSMV	NC_014547.1
Maize dwarf mosaic virus	MDMV	NC_003377.1
Maize streak virus	MSV	NC_001346.2
Northern cereal mosaic virus	NCMV	NC_002251.1
Paspalum striate mosaic virus	PSMV	NC_018530.1
Rice stripe virus	RSV	NC_003753.1; NC_003754.1; NC_003755.1; NC_003776.1
Soil-borne wheat mosaic virus	SBWMV	NC_002041.1; NC_002042.1
Triticum mosaic virus	TriMV	NC_012799.1
Wheat dwarf virus	WDV	NC_003326.1
Wheat streak mosaic virus	WSMV	NC_001886.1
Wheat yellow mosaic virus	WYMV	NC_002350.1 NC_002349.1

To generate unique sequences to serve as e-probes for each virus a modified version of the Tool for Oligonucleotide Fingerprint Identification (TOFI) was used (Satya et al., 2008). This method generates candidate e-probes by comparing the genome of the target virus to the genomes of 'near neighbors'. For this study, the 'near neighbors' of each of the 21 target viruses

consisted of all of the other 20 viruses used in the study. Each virus genome and the genomes of its 'near neighbors' were uploaded to the EDNA2 server (http://bioinfo.okstate.edu) at Oklahoma State University and e-probes for each were designed using MiProbe function. The EDNA2 server is an online system designed to combine e-probe design with pathogen detection. E-probes for each virus were designed to be 30bp in length to maximize the number of probes generated for each virus.

Due to the inability of some MPS platforms to accurately call a string of identical nucleotides, those e-probes that were generated containing homo-oligomers (a consecutive string of five or more nucleotides) were removed. These resulting 'first-draft' e-probes that were then queried against the NCBI nucleotide database and any that did not fall within an E-value of 1 x 10^{-9} were removed to increase specificity to their corresponding target virus. Once curation was complete, e-probe sets for each virus were then uploaded to the EDNA2 server. To serve as a negative control, 'decoy' e-probes were generated to control for the possibility of false positives 'hits' or random matches of e-probes to non-related sequences. These decoy e-probes were designed by reversing the sequence of each virus-specific e-probe within the set of e-probes and would be used to query the dataset alongside the target-specific e-probes.

To serve as a negative transcriptomic control, virus-negative wheat transcriptomes produced from another study were used. This wheat transcriptomic data was generated during a study by Espindola et al. (unpublished) in which wheat was grown in sterile soil within greenhouse conditions. Root samples were taken, flash frozen in liquid nitrogen, ground with a sterile mortar and pestle, then had total RNA extracted with RNeasy PowerSoil Total RNA Kit (Qiagen, Hilden Germany) following the manufacture's protocol. This process was repeated for a

total of six wheat samples in duplicate. Transcriptomic data generated from sequencing wheat root tissue RNA was then uploaded to the EDNA2 sever and a detection run was performed for each sample and duplicate transcriptome using e-probes for all target viruses of wheat.

2.2 Mock Sample Database Generation and Query

To determine to accuracy and theoretical sensitivity of virus-specific e-probes generated above, *in silico* mock MPS datasets were generated using MetaSim (Richter et al., 2008). Mock MPS datasets generated using MetaSim were constructed by mixing known percentages of target virus nucleic acids with host wheat background sequences. To do this, each virus genome was loaded onto MetaSim along with 22 sequences corresponding to the chromosomes of *T. aestivum* that were retrieved from NCBI. The accession numbers for each of the chromosome sequences of *T. aestivum* can be found in Appendix 2.2.7.

Once each virus and wheat background genome were loaded onto MetaSim, simulated metagenomes of ten million reads containing sequences from both the target virus genome and the wheat background DNA. Each simulation was performed using an algorithm designed to mimic the nucleotide substitution error rate (~0.1%) of Illumina sequencing platforms (San Diego, CA) (Bolger et al., 2014). For BYDV, mock databases were generated for each of the three species as well as a combined pool of all three BYDV species within a background of wheat reads. For each of the 21 target virus, metagenomes containing approximately 1,000, 100, and 10 target virus reads out of a total of ten million reads were generated. Each mock metagenome simulation was performed in triplicate and uploaded to the EDNA2 server.

Mock databases were then analyzed using the EDNA2 server. A pairwise sequence alignment was performed between the mock databases and curated e-probes with the BLASTn

function (Camacho et al., 2009). A mock metagenome was considered positive for the targeted pathogen when one or more e-probes successfully match, or 'hit'. To limit the chances of a false positives, e-probes that had 'hits' with an e-value greater than $1x10^{-5}$ or a percent identity score of less than 95% were not counted toward a positive match. All e-probe 'hits' that had e-values lower than $1x10^{-5}$ and a percent identity score $\geq$ 95% were considered quality 'hits'. If the number of quality 'hits' exceeded the minimum threshold for detection, four quality 'hits' or more, the metagenome was considered positive for the target virus.

This standard for establishing a positive hit has been previously validated on sequencing platforms with substitution error rates up to 0.5% (Espindola et al., 2015). Even with higher error rates, consensus accuracy scores of 99.99% have been reported (Margulies et al., 2005). To serve as a negative control, a mock database containing only wheat genome reads was generated with MetaSim, uploaded to the EDNA2 server, and was tested with each set of 21 virus e-probes. Statistical analysis of EDNA analysis was performed using decoy e-probes to assess the 'background' level of BLASTn 'hits' that randomly occur between two sequence databases as described in Strobbe et al., 2013.

2.3 MetaSim Simulated EDNA Detection Model for WSMV

A predictive model for limit of detection for WSMV using EDNA was developed by simulating a series of metagenomes with known percentages of WSMV reads to background host reads. A total of six metagenomes were generated containing WSMV reads in the following percentages: 0%, 0.000001%, 0.00001%, 0.0001%, 0.001% and 0.01%. Each of these metagenomes were simulated six times, uploaded to the EDNA2 server and queried with WSMV e-probes. The BLASTn summary output was then retrieved for each run and the number of

quality ‘hits’ for each WSMV e-probe was correlated to the percentage of WSMV reads in each queried metagenome and replicate. A linear regression analysis was performed using R (V.3.5.3) (Team, 2013) to determine if there was a significant correlation between the total number of e-probe ‘hits’ and the percentage of WSMV reads in each metagenome. A one-way ANOVA was performed with Tukey’s HSD in Excel (Redmond, WA) to determine any significant difference between the average total e-probe ‘hits’ among simulated WSMV abundances with an alpha of 0.05.

2.4 Validation of Simulated WSMV Detection Model

A total of four primers pairs were designed to amplify a combined 9044bp of the 9384bp WSMV genome from positions 229 to position 9273. These primers were designed using consensus sequences of the entire WSMV genome retrieved from NCBI and aligned with Clustal X (V.20) (Larken, 2007). The consensus genome for WSMV was used as the basis for primer design using Primer3 (Untergasser et al., 2007). A total of four primer pairs, WSG1, WSG2, WSG3 and WSG4 were generated that amplified regions of the WSMV genome from bp 229-2477, bp 2445-2464, bp 5045-7433 and bp 7418-9273 respectively. Table 2 contains all designed primer pairs, their thermodynamic properties and regions of the WSMV genome that they amplify.

Table 2. Table of designed primers for the combined amplification of the WSMV genome with corresponding thermodynamic values calculated by Primer 3 (v.0.4.0)

Primer	Primer Sequence (5′ - 3′)	Length	Tm(°C)	Size(bp)	Any	3′	Start	Stop
WSG1F	AAGCACTGAGGAGGAGGTTG	20	59.31		3.00	0.00	229	248
				2249				
WSG1R	TTAAGCCTCCCAACACGAAG	20	57.82		4.00	0.00	2477	2458
WSG2F	ATCCTTCAACGCTCTTCGTG	20	58.29		3.00	0.00	2445	2464
				2370				
WSG2R	CCAAATGGTGCTTTTCGTCT	20	56.91		4.00	0.00	4814	4795
WSG3F	TTCCAGCAGCAACAATCAAC	20	57.20		4.00	0.00	5045	5064
				2389				
WSG3R	CATCAGCGTCAATGAACCAC	20	57.47		5.00	0.00	7433	7414
WSG4F	TTCATTGACGCTGATGGTTC	20	56.43		5.00	0.00	7418	7437
				1856				
WSG4R	TCCTGGTACTCGTGGATTTGT	21	58.75		4.00	0.00	9273	9253

Field samples of wheat received by the Plant Disease and Insect Diagnostic Laboratory at Oklahoma State University (PDIDL) were screened for the presence of WSMV with ELISA. An ELISA confirmed, WSMV-positive wheat samples, as well as ELISA confirmed, WSMV-negative wheat samples were selected for this study. Total RNA was extracted from the ELISA-confirmed, WSMV infected wheat sample using a RNeasy RNA extraction kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol. The resulting RNA was then used as the template for an MMLV reverse transcription reaction (Thermo Fisher, Waltham, MA), which was performed as follows: 9.5µL nuclease-free water, 1µL of 10mM dNTP, 0.2µL of 1µg/µL WSG4R primer and 4µL of template RNA. Each reaction was denatured at 70°C for five minutes then incubated on ice for one minute. Following incubation, samples were mixed with a reaction mixture containing: 0.5µL of 40U/µL RNAasin (Promega, Madison, WI), 4µL of 5x M-MLV RT buffer and 0.8µL of 200U/µL M-MLV reverse transcriptase (RT). Once combined, each 20µL reaction was incubated for 90 min at 37°C. ELISA-confirmed, WSMV negative wheat samples had DNA extracted using a DNeasy DNA extraction kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol.

The cDNA produced from the reverse transcription of WSMV positive wheat RNA was then used as the template for PCR with designed primers. First, to confirm the presence of the entire, unbroken length of WSMV cDNA, PCR was performed to amplify the predicted 9,044bp region of WSMV in four segments. Four separate PCRs were performed with primer pairs WSG1, WSG2, WSG3 and WSG4 respectively. Each 50µL PCR reaction was as follows: 25µL of Platinum™ SuperFi™ Green PCR Master Mix (Invitrogen, Carlsbad, CA), 2.5µL of each 10µM forward and reverse primer, 50ng of template cDNA and molecular grade water to bring the total volume up to 50µL. PCR was performed using the following thermal conditions: 98°C for 30 sec, followed by 30 cycles of 98°C for 7 sec, 58°C for 10 sec, 72°C for 4 min, followed by a final extension of 72°C for 5 min. Following confirmation with gel electrophoresis, bands of expected length of ~2,000bp for each reaction were excised and purified using the PureLink ™ Quick Gel Extraction and PCR Purification Kit (Invitrogen, Carlsbad, CA) following the manufacture's protocol, then sequenced. Sequencing was performed at the Oklahoma State University Core Facilities using a 3730XL gene analyzer (Thermo Fisher, Waltham, MA).

To isolate pure WSMV dsDNA, the WSMV genome was amplified in two segments, then mixed. WSMV segment 1 consisted of a 4,547bp region of the WSMV genome from positions 248 and 4,795 and WSMV segment two was a 4,189bp region of the WSMV genome from positions 5,064-9,253. These sequences were generated by performing two PCR reactions with primer pairs WSG1F and WSG2R, and WSG3F and WSG4R respectively. These PCR reactions were performed as follows: 25µL of Platinum™ SuperFi™ Green PCR Master Mix (Invitrogen, Carlsbad, CA), 2.5µL of each 10µM forward and reverse primer, 50ng of template WSMV cDNA and molecular grade water to bring the total volume up to 50µL. The thermal

conditions for this reaction were: 98°C for 30 sec, followed by 30 cycles of 98°C for 7 sec, 58°C for 10 sec, 72°C for 4 min, followed by a final extension of 72°C for 5 min.

The two WSMV genome products generated from these PCR reactions were isolated and gel purified using the PureLink ™ Quick Gel Extraction and PCR Purification Kit (Invitrogen, Carlsbad, CA). Each gel purified amplicon was then used as the template for a second enrichment PCR following the same concentrations and thermal conditions described above. This enriched PCR product was then gel purified using the PureLink ™ Quick Gel Extraction and PCR Purification Kit (Invitrogen, Carlsbad, CA) and each enriched, isolated WSMV product had its concentration measured using the Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher, Waltham, MA). Once concentrations of each WSMV dsDNA fragment were determined, they were mixed in a ratio of two to one in favor of segment one (WSMV positions 248 to 4,795) to simulate differences in gene expression between the two WSMV segments.

Samples were prepared for Illumina sequencing by pooling DNA of WSMV segment 1 and 2 and aliquoting the pooled nucleic acids into 15 equal volumes. Each aliquot was then mixed with the appropriate volume of WSMV negative wheat DNA to generate 3 replicates of five ratios containing WSMV DNA and WSMV negative wheat DNA. These five solutions included mixtures containing the following percentages of WSMV DNA: 100%, 0.001%, 0.0001%, 0.00001% and 0.000001%. Three replicates of a sixth solution containing only WSMV negative wheat DNA were also generated to serve as a negative control. Library preparation was then performed for each of the six mixtures and their two other replicates using the KAPA HyperPlus library prep kit (Roche, Basel, Switzerland) and KAPA Single Index Adaptors (Roche, Basel, Switzerland) as per the 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 triplicate.

Following sequencing, each treatment and triplicate metagenome was uploaded to the EDNA2 server and queried using WSMV e-probes. The BLAST results for each detection run was then used to determine the number of total number of 'hits' for each of the WSMV e-probes individually and the total number of 'hits' for all e-probes per run. This data was then correlated to the known percentage of WSMV nucleic acids to WSMV negative host nucleic acids used for each treatment. A regression analysis was then calculated using R (V.3.5.3) to determine if there was a significant correlation between the concentration of target within a sample and the number of e-probe 'hits'. A one-way ANOVA was performed with Tukey's HSD to determine significance between the mean total e-probe 'hits' among levels of WSMV abundance. To determine if a significant difference existed between the number of 'hits' per any given e-probe within simulated metagenomes containing WSMV reads, a one-way ANOVA was performed on BLASTn summary output data for the number of 'hits' per probe for every mock metagenome.

3.0 Results

3.1 E-Probe Design

E-probe generation was successful for all 21 viruses used for this study (Table 3). The number of e-probes generated for each virus varied, and a linear regression analysis revealed a strong correlation between the size of the virus genome and the number of e-probes generated ($R^2 = 0.89$) ($p < 0.05$) (Fig. 1). The average number of e-probes generated for all 21 viruses was 24.65, with a median of 29.50. The viruses with the greatest and fewest number of e-probes were RSV and BYDV-PAV, with 55 and 2 e-probes respectively. While BYDV did have 61 e-probes

generated, it was not included in this direct comparison between all other viruses tested because it is a virus comprised of the genomes from three species of BYDV, BYDV-MAV, BYDV-PAS and BYDV-PAV. Across the viruses tested in this study, an average of 2.843 e-probes were generated per 1000bp of genome present.

Table 3. List of target viruses, their respective genome size, the number of curated e-probes generated for each. The genome size for *Barley yellow dwarf virus* (All) is not listed because it is a virus comprised of three species with varying genome sizes (BYDV-MAV, BYDV-PAS and BYDV-PAV).

Virus Name	Genome Size	Number of E-Probes
Agropyron mosaic virus	9,540	35
Brome Mosaic Virus	8,210	32
Barley Stripe Mosaic Virus	10,096	41
Barley yellow dwarf virus MAV	5,273	6
Barley yellow dwarf virus PAS	5,695	11
Barley yellow dwarf virus PAV	5,677	2
Barley yellow dwarf virus (All)	-	61
Barley yellow striate mosaic virus	12,706	30
Cocksfoot mottle virus	4,082	14
Cereal yellow dwarf virus-RPV	5,662	17
Digitaria striate mosaic virus	5,578	19
Maize dwarf mosaic virus	9,515	33
Maize streak virus	2,689	6
Northern cereal mosaic virus	13,222	45
Paspalum striate mosaic virus	2,816	6
Rice stripe virus	17,144	55
Soil-borne wheat mosaic virus	10,692	31
Triticum mosaic virus	10,282	40
Wheat dwarf virus	2,750	6
Wheat streak mosaic virus	9,384	29
Wheat yellow mosaic virus	11,295	35

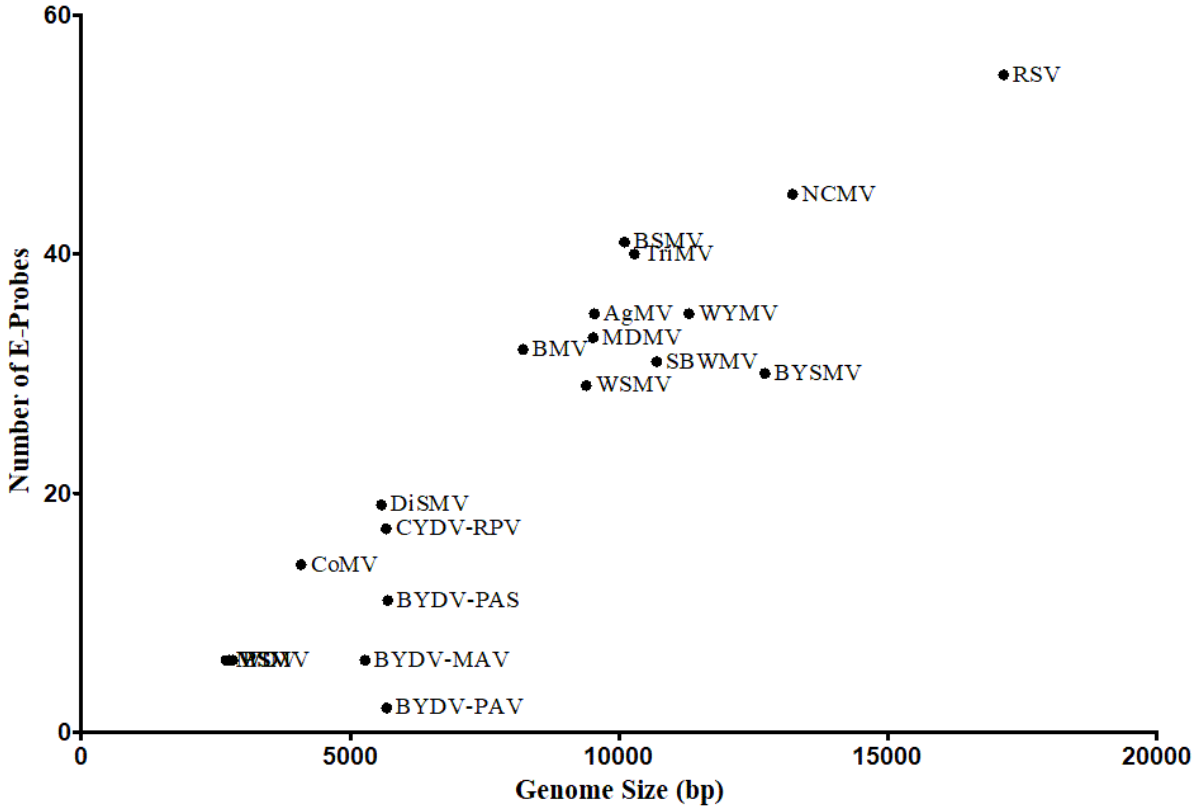


Figure 1. Graph displaying a direct correlation (R^2=0.89) between the size of the virus genome and number of e-probes that were generated. Compared to all other viruses, BYDV-PAV, BYDV-PAS and BYDV-MAV had fewer e-probes generated due to the use of more closely related 'near neighbors' when compared to all other viruses.

3.2 Mock Sample Database Generation and Query

For each virus, a total of nine metagenomes containing ten million, 100bp long reads each were simulated with the majority of reads belonging to a wheat genome background. These nine metagenomes were divided into three even groups ranging in 1-10, 100-200 and 1000-2000 virus reads respectively. None of the 21virus e-probe sets were capable of detecting between their target pathogen when only 1-10 reads out of the ten million total belonged to the target. At the 1000-2000 range, every e-probe set could detect its respective virus ($p < 0.002$). Within the

100-200 read range, only one e-probe set, BYDV-PAV, was unable to detect its respective pathogen ($p = 0.450$). However, to serve as a model, a total of 30 metagenomes were generated containing WSMV reads with a wheat genome background. These simulations included a read range of 30-99 reads, which was not tested for in the other 20 viruses in this study. WSMV e-probes were shown to be able to detect WSMV at a total of 31 reads out of ten million reads total ($p = 0.043$).

3.3 MetaSim Simulated EDNA Detection Model for WSMV

To determine the relationship between the total number of WSMV e-probe 'hits' and the ratio of 'hits' per probe to the abundance of WSMV reads in a metagenome, an *in silico* predictive model was simulated. This model contained a total of 36 mock metagenomes with relative abundances of WSMV reads to wheat genome reads under the following approximate percentages: 0%, 0.00001%, 0.00001%, 0.0001%, 0.001% and 0.01%. Each of these metagenomes were composed of ten million, 100bp long reads, with six metagenomes simulated for each level of relative WSMV abundance. The average total number of e-probe 'hits' for each WSMV read abundance and the average number of 'hits' per probe at each abundance can be found in Appendix 2.2.8 and 2.2.9.

Tukey's HSD revealed no significant difference in the average total 'hits' between simulated WSMV read abundances of 0%, 0.000001% and 0.00001% ($p = 0.899$). There was determined to be a significantly increased number of average total e-probe 'hits' for each subsequent simulated WSMV read abundance above 0.00001% and all other abundance ($p = 0.001$). For example, at WSMV read abundance of 0.0001%, there was a significantly higher number of average e-probe 'hits' than at any abundance level below it, but a significantly lower

number of e-probe ‘hits’ than any abundance level above it (p = 0.001). This trend was the same for simulated WSMV read abundances of 0.001% and 0.01% respectively (p = 0.001). A graph of the average number of e-probe ‘hits’ to the relative abundance of WSMV reads within simulated metagenomes can be found below (Fig. 4). It was also determined that within any of the tested simulated metagenomes, there was no significant difference between the number of ‘hits’ per any single e-probe (p = 0.944) (Fig. 2).

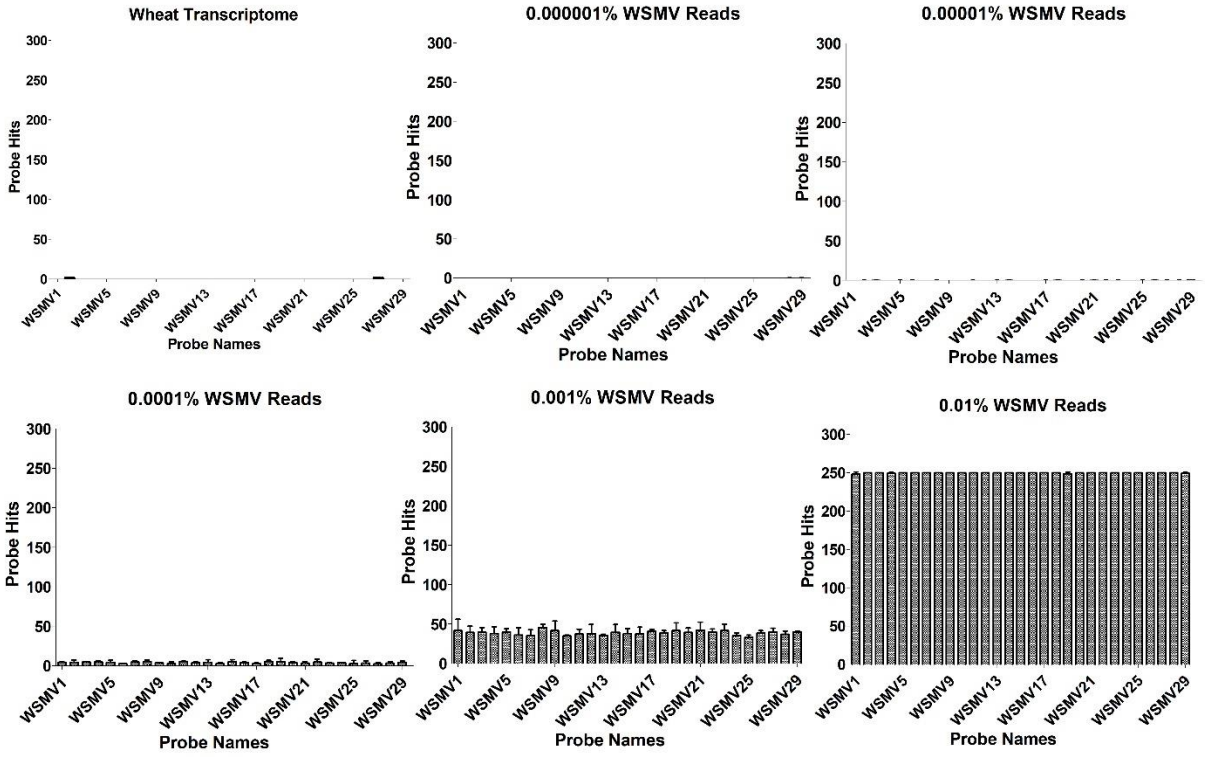


Figure 2. Graphs displaying the *in silico* predictions for the mean number of WSMV e-probe ‘hits’ for a WSMV-negative wheat transcriptome and WSMV abundances of 0.000001%, 0.00001%, 0.0001%, 0.001% and 0.01%. It was determined that no significant difference exists between the mean number of probe ‘hits’ for any of the 29 WSMV probes at any treatment (p > 0.05) (n = 6). There was determined to be a significant difference between the number of e-probe ‘hits’ for each probe and the abundance of WSMV reads in the metagenome. At WSMV abundance 0.001%, there were significantly more ‘hits’ than at lower abundances and significantly fewer ‘hits’ than at 0.01% WSMV abundance (p < 0.05) (n = 6). Error bars present are representative of SEM.

3.4 Validation of Simulated WSMV Detection Model

Out of the 29 designed e-probes specific to WSMV, probe set one (WSMV1 - WSMV19) matched PCR amplified segment one of the WSMV, while probe set two (WSMV20 - WSMV28) matched amplified segment two of the PCR amplified WSMV genome (Fig. 3). WSMV29, designed to act as an internal negative control, does match with the WSMV genome, but does not match with either of the PCR amplified WSMV genome segments used to generate the metagenomes in this study. WSMV29 was found to have zero 'hits' across all non-simulated metagenomes tested.

In total, 18 metagenomes of approximately 33 million 100bp long reads were generated with Illumina sequencing. Each of these metagenomes had both the forward and the reverse strands sequenced leading to three true replicates and three pseudo-replicates for the following levels of WSMV abundance: 0% (wheat metagenome), 0.000001%, 0.00001%, 0.0001%, 0.001% and 100%. A graph displaying the average number of e-probe 'hits' in relation to each of these WSMV abundances can be found below alongside the average total e-probe 'hits' for WSMV-negative wheat transcriptomes (Fig. 4). A table containing the mean total number of e-probe 'hits' and the mean number of 'hits' per probe within each abundance of WSMV for e-probe sets one and two can be found in Appendix 2.2.9.

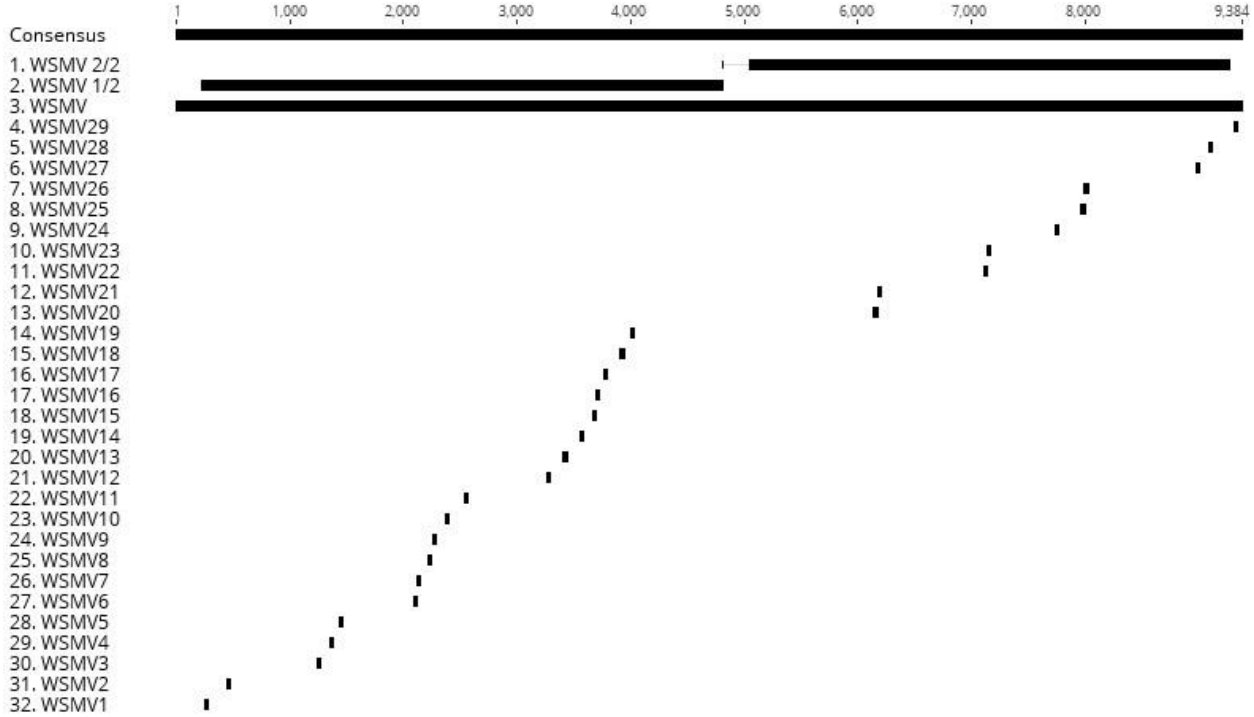


Figure 3. Alignment of the WSMV genome, curated WSMV e-probes and the two amplified segments of the WSMV genome (WSMV 1/2 and WSMV 2/2) generated with Geneious (v.9.0.4). WSMV specific e-probes WSMV1-WSMV28 align with both the WSMV genome and only one of the two amplified WSMV segments. The negative control WSMV e-probe, WSMV29, aligns with the WSMV genome but does not align to either of the amplified WSMV segments.

Tukey's HSD revealed that no significant difference existed between the average number of WSMV e-probe 'hits' for *in vitro* abundances of WSMV within the WSMV-negative wheat metatranscriptome, WSMV-negative wheat metagenome, or metagenome containing 0.000001% WSMV nucleic acids ($p = 0.580$) ($n = 6$). For subsequent nucleic acid abundances (0.00001%, 0.0001%, 0.001% and 100%) it was determined that significant differences in the average e-probe 'hits' existed between each ($p = 0.001$). For example, the average total number of e-probe 'hits' for WSMV nucleic acid abundance of 0.00001% was significantly higher than lower abundances, but significantly lower than higher abundances ($p = 0.001$). This trend was the same for all subsequently higher abundances of WSMV nucleic acids ($p = 0.001$). Non-linear regression analysis revealed a direct correlation between the concentration of WSMV nucleic

acids within the metagenome and the average total number of e-probe 'hits' ($R^2 = 0.590$, p = 0.0001).

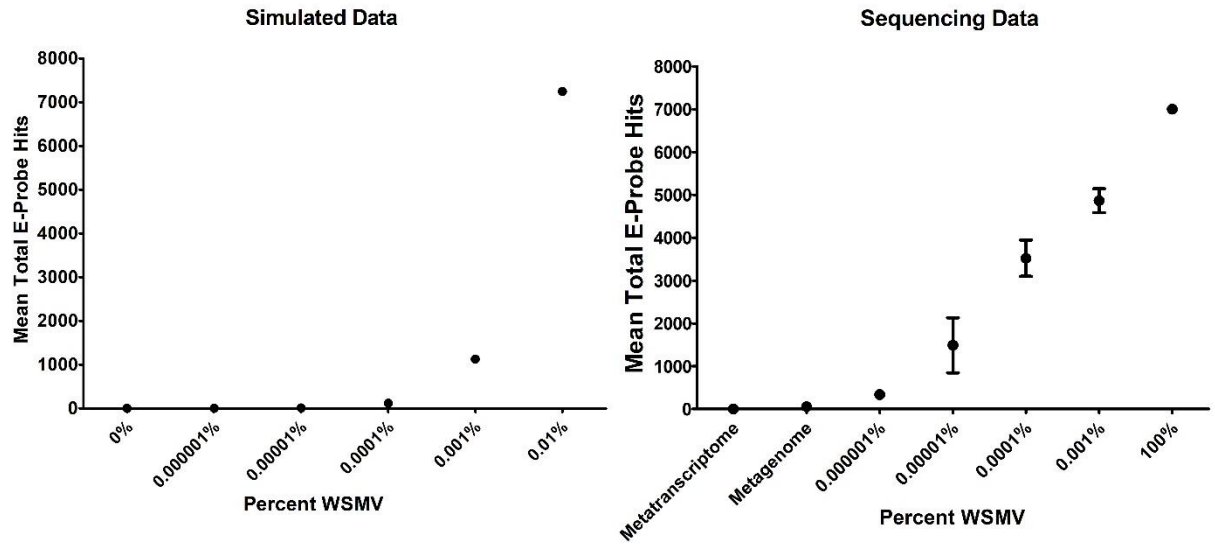


Figure 4. Comparison of *in silico* simulated data for the ratio of the average total number of WSMV e-probe 'hits' in relation to the abundance of WSMV within the metagenome vs the *in vitro* sequencing data for the ratio of the average total WSMV e-probe 'hits' to the abundance of WSMV within the metagenome. The average total number of e-probe 'hits' for WSMV-negative wheat metatranscriptomic data and WSMV-negative wheat metagenomic data can also be found on graph containing sequencing data. Non-linear regression analysis of simulated data showed an $R^2 = 0.997$ ($p < 0.0001$) (n = 6) and non-linear regression analysis of sequencing data revealed an $R^2 = 0.590$ ($p < 0.0001$) (n = 6). For simulated data there was a significant increase in the average total number of e-probe 'hits' at 0.001% WSMV abundance and 0.01% WSMV abundance when compared to all other WSMV abundances tested. All bars represent standard error of the mean.

A significant difference was observed between the total number of e-probe 'hits' between WSMV probe set one (probes 1-19) and WSMV probe set two (probes 20-28) at all abundances of WSMV nucleic acids tested except for 0.00001% and 100% (p = 0.36, p = 0.89). There was also a significant difference between the number of probe 'hits' for probe set one and probe set two at WSMV abundances of 0% (p = 0.003), 0.000001% ($p = 1.19 \times 10^{-15}$), 0.0001% ($p = 1.87 \times 10^{-26}$), and 0.001% ($p = 4.59 \times 10^{-57}$). At WSMV abundance of 100%, the mean number of probe 'hits' between probe sets one and two, as well as the number of 'hits' per probe became

identical for all WSMV e-probes. Within probe sets, one-way ANOVA analysis revealed no significant difference between the number of 'hits' for any individual e-probe at any abundance of WSMV tested. A graph displaying the differences in the number of e-probe 'hits' for all WSMV e-probes across all abundances of WSMV tested can be found below (Fig. 5).

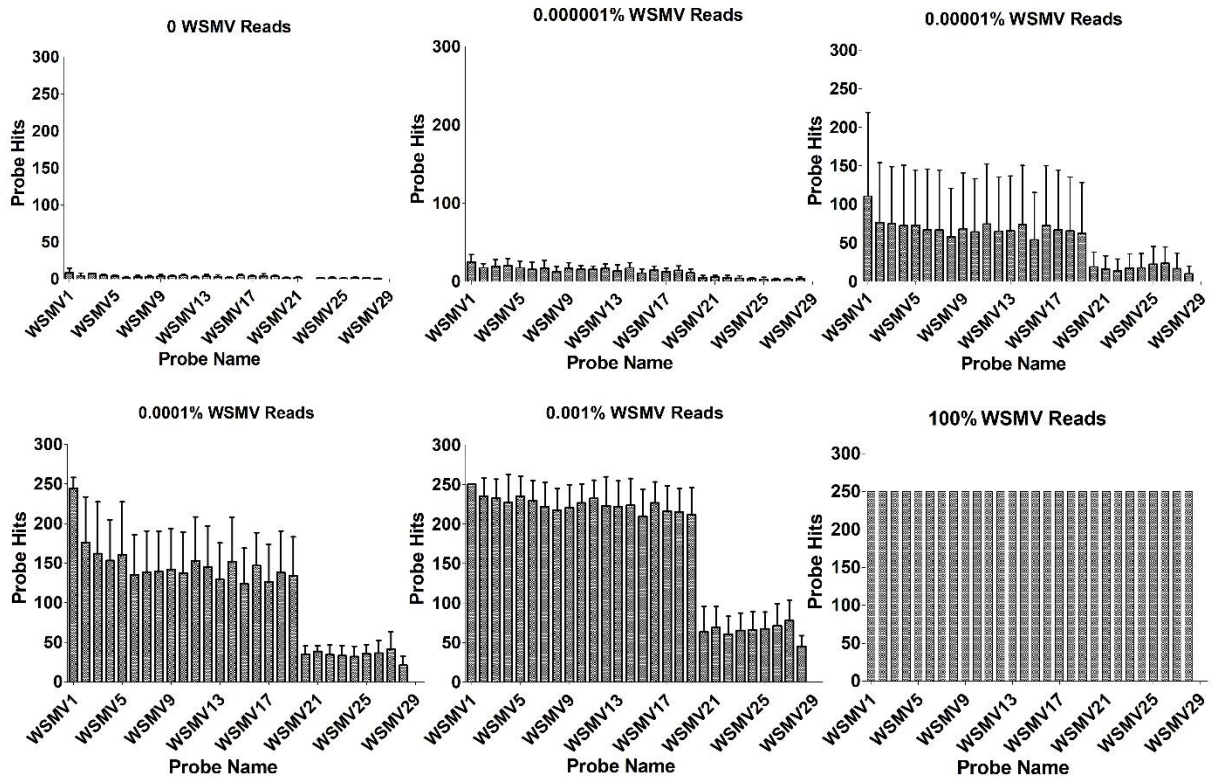


Figure 5. Graphs displaying sequencing data results for the number of WSMV e-probe 'hits' at WSMV read abundance at 0%, 0.000001%, 0.00001%, 0.0001%, 0.001% and 100%. To simulate differential virus gene expression, WSMV nucleic acids corresponding to WSMV e-probes 20-29 had half of the starting concentration than did WSMV nucleic acids corresponding to WSMV e-probes 1-19. WSMV e-probe 29 was used as a negative control as the region of the WSMV genome that it corresponds to was not included in the nucleic acids found in any of the tested metagenomes. For WSMV e-probes 1-19 and 20-28, there were no significant differences between the mean number of 'hits' within any treatment group respectively. A significant difference was observed between the mean total 'hits' of e-probes 1-19 vs 20-28 within treatment groups across all probes ($p < 0.05$). At two WSMV read abundances, 0% and 0.000001%, there were significantly fewer mean 'hits' for all e-probes when compared to all other treatment groups ($p < 0.05$) ($n = 6$). All bars represent standard error of the mean.

A final one-way ANOVA analysis was performed to assess the frequency of e-probe 'hits' per number of e-probe across all abundances of WSMV tested for both simulated and *in vitro* sequencing data. This analysis was repeated for both WSMV e-probe sets one and two. It was determined that in simulated metagenomic data where the starting concentration of the entire genome of WSMV was present revealed no significant difference between the hit per e-probe frequency ($p = 0.89$). For sequencing data, the starting concentration of WSMV nucleic acids differed for corresponding e-probe sets one and two. In all tested abundances, the hit to e-probe ratio was significantly lower for probe set two ($p < 0.05$) ($n = 6$), where the ratio of corresponding WSMV nucleic acids was half that of the corresponding WSMV nucleic acids for probe set one. This was true for all abundances of WSMV nucleic acids except for 100%, where all probes tested for both probe sets had the maximum number of 'hits' possible (data not shown). Below in figure six is a side by side comparison for the hit frequency between probe set one and probe set two for both simulated and *in vitro* sequencing data.

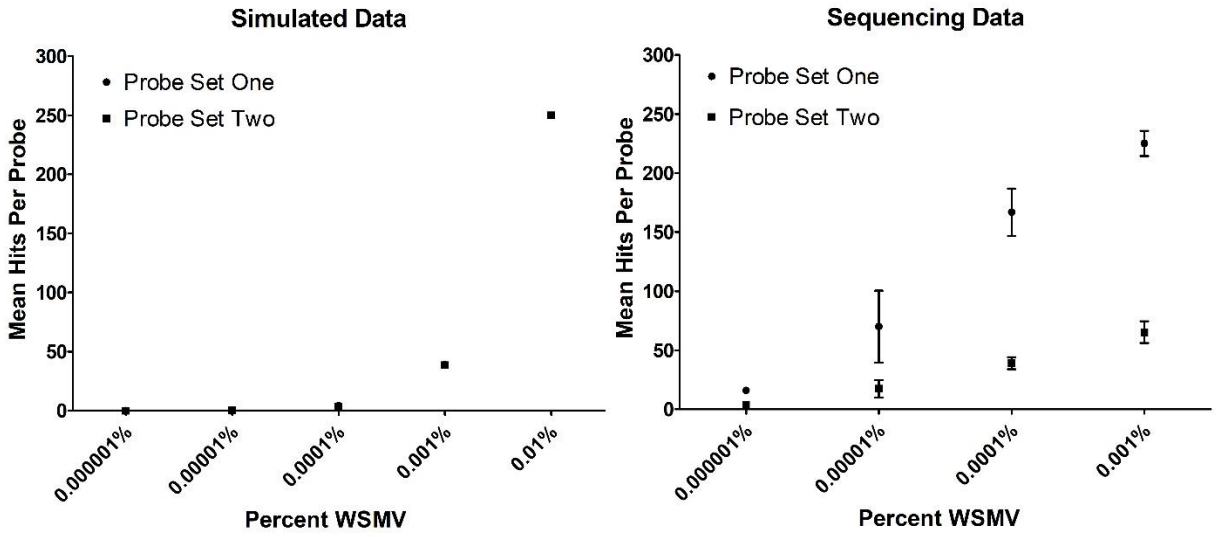


Figure 6. Graphs comparing total mean 'hits' to probe ratio vs the percentage of WSMV reads for probe set one (WSMV1-WSMV19) and probe set two (WSMV20-WSMV28). For simulated data, where e-probe coverage of the WSMV genome is equal for all probes, there was no difference between the mean hit to probe ratio across any treatment group ($p > 0.05$) ($n = 6$). For sequencing data, where the concentration of WSMV nucleic acids was higher for probe set one compared to probe set two, the mean hit to probe ratio was significantly higher for all treatment groups ($p < 0.05$) ($n = 6$). The rate of change for probe set one was determined to be greater than the rate of change for the frequency of probe set two, suggesting a non-linear relationship between the number of e-probe 'hits' and the concentration of target within the metagenome. Error bars shown represent standard error measurement.

Discussion

The objective of this study was to develop a MPS-based diagnostic tool for the detection of some of the most common viruses of wheat and then use WSMV to validate *in silico* predictive models. Using EDNA, it was possible to develop target specific e-probes for 21 different viruses of wheat and predict the limit of detection for each using simulated metagenomic data. Based on the results of e-probe generation, it was determined that the length of the virus genome was a significant variable for predicting the ultimate number of e-probes generated for it. RSV for example, a virus with a quad-partite genome had the largest genome in

terms of total length (17kb) and had the highest number of e-probes (55) for any single virus used in this study. Another virus, BYDV-PAV, only had two e-probes generated for it after curation even though it did not have the smallest genome present. The most likely explanation was the method of curation.

Unlike every other virus in this study, BYDV-MAV, BYDV-PAS and BYDV-PAV are all different species of the same virus, BYDV. To develop e-probes capable of discriminating one from another, many prospective e-probes would have been removed due to significant sequence homology to the other BYD viruses. Thus, even though BYDV-PAV has a genome size of ~5kb, it produced the fewest probes out of all the viruses tested. The number of e-probes within a probe set is an important factor for determining the limit of detection. This was apparent for BYDV-PAV as it was the only virus for which the generated e-probe set was unable to detected it between 100-200 target reads within a metagenome. All other virus e-probe sets were able to detect at this level.

To establish a more defined model for the limit of detection of viruses of wheat using EDNA, a series of mock metagenomes were created for WSMV. These mock genomes were created using MetaSim to simulate Illumia sequence runs containing reads from both WSMV and a host wheat background. These mock metagenomes ranged in concentration or WSMV reads from 0% to 0.01%. After performing EDNA on each using curated WSMV probes, it was noted that the total number of e-probe ‘hits’ increased exponentially as the concentration of WSMV reads increased. This relationship is important to note as it is a critical component for the development of a predictive model for the limit of detection for viruses of wheat. Another important factor is how well any single e-probe produces quality ‘hits’ when compared to the

other e-probes within its respective probe set. It was determined that there was no significant difference between the frequency at which any individual probe produces 'hits' if the target nucleic acids are present in equal proportions.

Another objective of this study was to validate the predictive, *in silico* model for the limit of detection of WSMV by mimicking the method by which MetaSim generates mock metagenomes in *in vitro* settings. This was done by amplifying, then isolating the WSMV genome with reverse transcription, PCR and gel purification before mixing it into known ratios of WSMV nucleic acids to host wheat nucleic acids. To further test the limits of EDNA, the WSMV genome was amplified in two approximately equal lengths, then mixed together in un-even proportions. The purpose of this unequal mixing of the two halves of the WSMV genome was to determine if EDNA would be capable of distinguishing between variations in virus gene expression. The final difference between the *in vitro* metagenomes and the *in silico* ones was the use of one e-probe, WSMV29, as a negative control. While WSMV29 does match with the WSMV genome, it does not match with either half of the amplified regions of the WSMV genome. This was done to increase confidence in the method and that the BLASTn analysis was performing as expected.

Once mixed, the WSMV nucleic acids were then mixed with WSMV-negative wheat DNA to produce a series of WSMV concentrations like those produced in the MetaSim mock metagenomes. After EDNA was performed, it was determined that the trend for the number of e-probe 'hits' to WSMV concentration was similar to that produced through simulated genomes. However, while the relationship between 'hits' and concentration for both *in silico* and *in vitro* was exponential, there were significantly more 'hits' seen in the *in vitro* data when compared to

the simulated data. One likely explanation is the number of reads that were produced by both methods. Metagenomes produced by MetaSim contained ten million reads each, while the metagenomes produced by *in vitro* Illumina sequencing had approximately 33 million reads each. This difference alone could account for much of the variation. Therefore, the most important variable influencing the number e-probe 'hits' is the number of WSMV reads within the metagenome, and not necessarily the concentration of WSMV reads.

As expected, the number of 'hits' for WSMV e-probes 20-28 were found to be significantly lower than WSMV e-probes 1-19, which is important for understanding the limits of EDNA to distinguish between differences in virus gene expression during infection. This result is promising as during virus infection *in vivo*, there are variations in virus gene expression depending on how long the plant has been infected (Whitham et al., 2003). This may provide diagnosticians a tool for tracking the progression of virus diseases if e-probes are designed in such a way to target specific virus gene transcripts. Within e-probe sets, it was determined that there was no significant difference between the rate at which any single e-probe produced 'hits' compared to any other probe within the same probe set. This provides more support for the use of EDNA as a quantitative tool to measure virus gene expression as e-probes have been shown to function consistently relative to one another.

Another difference between WSMV probe set one and probe set two was the ratio of 'hits' per e-probe across all tested concentrations of WSMV nucleic acids (Fig. 6). This difference was expected due to the exponential relationship between the number of e-probe 'hits' and the abundance of target within the metagenomic sample. Since the ratio of corresponding nucleic acids for WSMV e-probes 1-19 was higher than that of WSMV e-probes 20-28, a

decrease in the rate of the frequency of 'hits' would be expected with an exponential, non-linear relationship. If the relationship between the number of e-probe 'hits' and the concentration of target within the sample were linear, the rate of the hit to probe frequency would be the same for both sets of e-probes. Considering that the slope of the 'hit' per probe frequency was smaller for the half of the WSMV genome that was in lower concentration relative to the other provides more evidence that the relationship between the number of e-probe 'hits' and target concentration is a non-linear one.

Although there was greater variation in the *in vitro* data when compared to *in silico* models, there was still determined to be no significant difference between the number of 'hits' for any single e-probe compared to one another assuming their corresponding nucleic acid targets are in equal proportion. This result corroborates what was shown in the *in silico* models, suggesting that e-probes do not differ from one another significantly at the length, e-value and percent identity tested in this study. It is possible that changing the length of e-probes or the e-value cutoff for quality 'hits' may change the rate at which e-probes produce 'hits'.

In conclusion, it was determined that *in vitro* sequencing models resembled those produced by *in silico* models for the limit of detection of WSMV and the 'hit' frequency of e-probes at various target concentration. Due to the variation associated with *in vitro* sequencing such as randomness in sequencing and human error during measurement, it is not clear if the model generated here is robust enough ($R^2 = 0.590$) to use for other viruses without further testing. At lower concentrations tested however, such as 0.000001% WSMV, the variation between the number of probe 'hits' per probe were lower than at higher concentrations of 0.00001% and 0.0001%. This provides evidence that *in vivo* models, where the concentration of

WSMV would be expected to be relatively low, would perform better than *in vitro* models at concentrations higher than would be found in nature.

The results of this study indicate that the most important variables that influence the number of e-probe 'hits' are the number of e-probes for the target relative to its genome size and the number of target reads within the metagenomic sample. In the future, a predictive model should be developed that is based on these two variables as they relate to different pathogen targets. The focus of this study was on viruses where the relatively small genomes increase the speed and ease of the curation process due to the ability to include many more near neighbors in the initial curation when compared to prokaryotic or eukaryotic genomes. Studies in the future will also need to take variables such as e-probe length and metagenome read length into consideration as longer probes or shorter read length may adversely impact the limit of detection. Once these variables are standardized, it may be possible to develop a predictive model for a limit of detection range for any potential target.

In order to maintain the validity of the use of EDNA for the detection of viruses of wheat over time, several important factors must be taken into consideration. First, the e-probes developed for this study were derived from type strain sequences that are currently available. It is likely that before this method can be successfully employed in the field, generated e-probes must be tested against virus field samples and variant genomes available on databases. This is because virus genomes as known to vary by region, which may lead to the potential for false negative results (Miyashita & Kishino, 2010; French & Stenger, 2005). In the future, it will be important to update the e-probe set for every virus used in this study to maximize the number of variants

and strains of a given virus that can be detected. This is particularly true of RNA viruses, which are known to have a relatively high genome mutation rate (Holland et al., 1982).

An ultimate goal for the development and validation of EDNA should be to move away from the use of simulated genomes as they do not necessarily accurately reflect the error and randomness that is associated with sequencing and e-probe detection based on the results of this study. Instead, a central model for the limit of detection should be developed that establishes the relationship between the number of target reads in a metagenome and the proportion of the target for which e-probes match. By standardizing the length of reads and e-probes, it may be possible for a model like this to be applied to any pathogen system, not just viruses. This is because EDNA, unlike other detection techniques like ELISA, is based heavily on statistics and the likelihood of nucleotide matches between two or more sequences. Also unlike other detection methods, EDNA is based on the detection of *in silico* targets, which do not inherently differ between one pathogen or another. In other words, metagenomic data that contains nucleotide sequences for a target virus is screened in an identical way to a metagenome containing bacterial target sequences. A standardized model that could predict the number of target reads necessary for the detection of a pathogen based on the percent coverage of e-probes would eliminate the need to perform simulated sequencing.

REFERENCES

Appel, J. A., De Wolf, E., Todd, T., & Bockus, W. W. (2015). Preliminary 2015 Kansas Wheat Disease Loss Estimates. Kansas Coop. *Plant Dis. Survey Rep. http://agriculture. ks. gov/docs/default-source/pp-disease-reports-2012/2015-ks-wheat-disease-loss-estimates35b3d4002e6262e1aa5bff0000620720. pdf*.

Aquino, P., Carrión, F., & Calvo, R. (2000). Selected wheat statistics. *CIMMYT, 2001*, 52–62.

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.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Byamukama, E., Wegulo, S. N., Tatineni, S., Hein, G. L., Graybosch, R. A., Baenziger, P. S., & French, R. (2014). Quantification of yield loss caused by Triticum mosaic virus and Wheat streak mosaic virus in winter wheat under field conditions. *Plant Disease*, *98*(1), 127–133.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, *10*(1), 421.

Curtis, B. C., Rajaram, S., & Gómez, M. (2002). *Bread Wheat: Improvement and Production.* Food and Agriculture Organization of the United Nations (FAO)

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 utilising electronic probes. *International Journal of Data Mining and Bioinformatics*, *12*(2), 115-128.

FAO (Food and Agriculture Organization of the United Nations) (2016) Food and Agriculture Organization of the United NationsCrop Production and Trade Statistics (available from:http://faostat.fao.org/site/339/default.aspx).

French, R., & Stenger, D. C. (2005). Population structure within lineages of Wheat streak mosaic virus derived from a common founding event exhibits stochastic variation inconsistent with the deterministic quasi-species model. *Virology*, *343*(2), 179-189.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., & VandePol, S. (1982). Rapid evolution of RNA genomes. *Science*, *215*(4540), 1577-1585.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. & Thompson, J. D. (2007). Clustal W and Clustal X version 2.0. B*ioinformatics*, *23*(21), 2947-2948.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A. & Dewell, S. B. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376.

Miyashita, S., & Kishino, H. (2010). Estimation of the size of genetic bottlenecks in cell-to-cell movement of Soil-borne wheat mosaic virus and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *Journal of Virology*, *84*(4), 1828-1837.

Prabha, K., Baranwal, V. K., & Jain, R. K. (2013). Applications of next generation high throughput sequencing technologies in characterization, discovery and molecular interaction of plant viruses. *Indian Journal of Virology*, *24*(2), 157-165.

Richter, D. C., Ott, F., Auch, A. F., Schmid, R., & Huson, D. H. (2008). MetaSim—a sequencing simulator for genomics and metagenomics. *PloS One*, *3*(10), e3373.

Seifers, D. L., Martin, T. J., Harvey, T. L., Fellers, J. P., Stack, J. P., Ryba-White, M., Lovat, N. (2008). Triticum mosaic virus: A new virus isolated from wheat in Kansas. *Plant Disease*, *92*(5), 808–817.

Souaze, F., Ntodou-Thome, A., Tran, C. Y., Rostene, W., & Forgez, P. (1996). Quantitative RT-PCR: limits and accuracy. *Biotechniques*, *21*(2), 280-285.

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.

Team, R. C. (2013). R: A language and environment for statistical computing.

Umek, R. M., Lin, S. W., Vielmetter, J., Terbrueggen, R. H., Irvine, B., Yu, C. J. & Chen, Y. P. (2001). Electronic detection of nucleic acids: a versatile platform for molecular diagnostics. *The Journal of Molecular Diagnostics*, *3*(2), 74-84.

United Nations 2016 World Population Prospects: The 2016 Revision (available from: http://esa.un.org/unpd/wpp/index.htm).

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, *35*(suppl_2), W71-W74.

USDA (United States Department of Agriculture) 2011 National Agricultural Statistics Service: County level 'WheatAll' Yield (availablefrom:www.nass.usda.gov/DataandStatistics /indep)

Whitham, S. A., Quan, S., Chang, H. S., Cooper, B., Estes, B., Zhu, T., & Hou, Y. M. (2003). Diverse RNA viruses elicit the expression of common sets of genes in susceptible Arabidopsis thaliana plants. *The Plant Journal*, *33*(2), 271-283.

CHAPTER V

CONFERING RESISTANCE TO *WHEAT STREAK MOSAIC VIRUS* IN SUCEPTIBLE WHEAT THROUGH ACTIVATION OF THE RNA INTERFERENCE PATHWAY

Abstract

Bread wheat (*Triticum aestivum)* is one of the most important crops planted worldwide in terms of human nutrition and economic value. Plant viruses, including *Wheat streak mosaic virus*, are the cause of significant wheat yield loss each year. The use of genetically modified (GM) wheat has been shown to produce varieties that are immune to WSMV but due to poor public perception, are not commercially available. RNA interference (RNAi) is a known regulator of virus gene expression in plants and has been shown to induce resistance to plant viruses such as WSMV. The purpose of this study was to determine if double stranded RNA (dsRNA) injected into a WSMV-susceptible wheat cultivar prior to exposure to WSMV, may lead to resistance comparable to that of a WSMV-resistant wheat variety. dsRNA specific to the replicase gene of WSMV (Nib) was injected into the susceptible wheat variety Gallagher prior to mechanical inoculation with WSMV. WSMV copy number was then calculated using RT-qPCR at seven, fourteen- and twenty-one days post inoculation and compared to the WSMV copy number of two unreleased resistant varieties, OK13804W and OK12621. It was demonstrated that Gallagher treated with Nib-dsRNA prior to inoculation with WSMV had significantly lower WSMV copy numbers than both the control dsRNA treated Gallagher, untreated Gallagher and the resistant variety OK13804W after 21 days post inoculation ($p = 0.019$).

Introduction

Globally, plant viruses are responsible for crop losses that not only have the potential to cause severe economic impact to a region but also may result in famine (Speranza et al., 2008). Cereal crops such as maize, wheat and rice for example, are arguably the most important crops ever cultivated by humans both in terms of economics and nutrition and are considered staple crops for many impoverished nations such as Africa and India (Diao et al., 2010). Unlike fungal pathogens where fungicides are available to treat some diseases, there are currently no treatments available for cereal crops infected with viruses. The most commonly applied management practice of cereal viruses has been the use of resistant cultivars, however, there are significant shortcomings to this strategy. Resistant varieties may take many years to develop, produce unpredictable levels of resistance, often lead to lower yields and may only be effective for a few years (Garcia-Arenal & McDonald, 2003). In contrast, genetically modified (GM) crops, defined as cultivars in which the DNA has been altered to express one or more genes not naturally found in those crops, have been demonstrated to be effective at managing viral diseases. However, GM crops are not currently considered acceptable worldwide (Carpenter, 2010).

In wheat, one of the most economically important viruses is *Wheat streak mosaic virus* (WSMV), which has been known to cause complete crop failure in the Great Plains of the United States (Stenger et al., 2002). WSMV is a member of the Family *Potyviridae* and contains a positive sense, single stranded RNA genome of 9,384 nucleotides (Choi et al., 2001). The WSMV single open reading frame (ORF) polyprotein (350 kDA) encodes the proteins P1, HC-Pro, P3, CI, VPg, Nia, Nib and CP (Shukla et al., 1991). Proteins P1, HC-Pro and Nia function as proteases by cleaving the WSMV polypeptide into its individual constituent proteins (Verchot & Carrington, 1995). P1 is also known to suppress RNA silencing pathways due to a higher affinity

to short interfering RNAs (siRNA), leading to increased disease severity (Young et al., 2012). HC-Pro is an essential protein for the successful transmission of WSMV by its vector, *Aceria tosichella* (Keifer), but plays no role in systemic infection (Stenger et al., 2005). Nib and CP are essential for systemic infection however, as Nib is the RNA dependent RNA polymerase (RDRP) and CP is the coat protein, which is required for cell to cell transport within the host (Tatineni & French, 2014).

One of the mechanisms that is used to mitigate the effects of viruses like WSMV is the cultivation of wheat varieties expressing resistance (R) genes, which operate based on the gene-for-gene model of resistance first described by Flor (1947). Two such resistance genes in wheat to WSMV are *Wsm1* and *Wsm2*, both of which have been determined to inhibit WSMV long-distance transport between cells (Tatineni et al., 2016). Both *Wsm1* and *Wsm2* are single dominant R genes, with *Wsm1* being first derived from an intermediate wheatgrass (*Thinopyrum intermedium*) and *Wsm2* being first identified in wheat germplasm line CO960293-2 (Graybosch et al., 2009; Haley et al., 2002). One flaw of these genes however is that they are temperature dependent, only conferring high levels of resistance to WSMV at 18°C and below (Greybosch et al., 2009). As the external temperature rises above 18°C, resistance begins to fail and long-distance transport of WSMV virions within the wheat host increases (Tatineni et al., 2016). A second problem associated with the use of R genes is the potential for an incurred yield penalty often associated with the gene. Wheat varieties that contain the gene *Wsm1* have been shown to have significant lower yields than varieties without the *Wsm1* gene when WSMV is not present (Baley et al., 2001).

One possible alternative to the use of *Wsm1* and *Wsm2* in regions where early season temperatures are relatively high or WSMV is not always present is the use of transgenic, or genetically modified (GM) wheat varieties. The genes that are used in GM varieties often originate from another variety of the same species, or from a different species using one of several methods including zinc finger nucleases (ZFNs), transcription activator like nucleases (TALENs) or clustered regularly interspaced short palindromic repeats (CRISPR) (Miller et al., 2007; Sander et al., 2011; Cong et al., 2013). All these methods function by first creating a double stranded break (DSB) in the target genome, then inserting donor DNA into the space between where the DSB occurred (Davis et al., 2014). These methods have been used to insert known R genes into known susceptible host varieties and confer immunity to the target pathogen (Fahim, et al., 2010).

For plant viruses such as WSMV, one mechanism for conferring resistance or immunity to the host is by inserting a gene that produces a double stranded RNA (dsRNA) transcript that is homologous to a gene of the target virus (Fahim et al., 2010). These dsRNA transcripts are expressed by the host trigger RNA interference (RNAi) of the target virus through the formation of siRNA that are processed from the dsRNA (Hammond et al., 2001). These siRNA are incorporated into RNA-induced silencing complexes (RISC), which use them as a guide strand to degrade or inhibit translation of homologous RNA transcripts (Campbell & Choy, 2005). In transgenic plants, the resulting degradation of viral RNA leads to decreased viral replication and possible host resistance or immunity. This method of inserting transgenes that produce targeted dsRNA molecules has been used to produce immunity to WSMV in wheat by targeting WSMV genes for Nia, Nib and CP (Fahim et al., 2010; Sivamani et al., 2000; Sivamani et al., 2002).

A significant problem with the use of transgenic methods for inducing resistance to plant viruses is the poor public perception surrounding GM foods and products. The use of transgenic wheat is not currently accepted in any country and as of 2019 there are currently no GM wheat varieties commercially available on the global market. Therefore, it is necessary to develop a method for producing resistance to WSMV in wheat that does not rely on the method of inserting foreign genetic material in the wheat genome. The objective of this study is to determine if RNAi can be used as a protective treatment against WSMV in susceptible wheat treated with dsRNA homologous to WSMV genes Nia, Nib and CP prior to virus infection. It is hypothesized that by triggering the RNAi pathway of wheat against WSMV prior to the establishment of WSMV infection, it may be possible to infer long-term resistance to WSMV without ever modifying the genome of wheat. To assess the effectiveness of dsRNA treatment, WSMV copy number was compared between a dsRNA treated, WSMV-infected susceptible variety and two WSMV-infected resistant wheat varieties at three time points using reverse transcriptase quantitative PCR (RT-qPCR).

2.0 Materials and Methods

2.1 Selection of Susceptible and Resistant Wheat Varieties

To determine the impact of WSMV resistance genes on viral titer of WSMV, three wheat varieties were chosen based on their relative susceptibility to WSMV as determined by wheat variety comparisons performed at Oklahoma State University (Edwards et al., 2012) and Kansas State University (De Wolf et al., 2015) and by acreage planted (USDA, 2016). These wheat varieties include Gallagher, an established and commonly planted WSMV susceptible variety, and two experimental cultivars that have known resistance to WSMV, OK13804W and

OK12621. OK13804W is known to contain the WSMV resistance genes *Wsm1* and *Wsm2*, while the genotypic source of resistance in OK12621 is currently unidentified (USDA, 2016).

2.2 RNAi Construct Design

To produce dsRNA potentially capable of inducing the RNAi pathway in wheat, primers were designed to amplify three targets of the WSMV genome essential to virus replication, the genome regions for Nia, Nib and CP. These primers were designed using WSMV sequences retrieved from NCBI and aligned with MEGA7 (V.7.0.21). The consensus regions obtained for Nia, Nib and CP were then used as the basis for primer design with Primer3 (V.0.4.0). To serve as a negative control, three primer pairs were designed to produce three products that are not homologous to any genes in wheat or WSMV, yet similar in length to predicted amplicons for Nia, Nib and CP. To do this, conserved DNA sequences for the NADH dehydrogenase gene of the mosquito species *Anopheles aegypti* were used. Multiple sequences for this gene were retrieved from NCBI, aligned with MEGA7 and the consensus was used to design primers in Primer 3. To ensure primer specificity, potential amplifications of off-target products for all designed primers were assessed *in silico* using Primer Blast. A table of designed primers for Nia, Nib, CP and three *A. aegypti* NADH targets can be found below (Table 1).

Table 1. Table of designed primers and calculated thermodynamic properties for WSMV genes CP, Nia, Nib and three *Anopheles aegypti* NADH gene regions. Thermodynamic properties were calculated using Primer 3 (V.0.4.0).

Primer	Primer Sequence (5′ - 3′)	Length	Tm(°C)	Size(bp)	Any	3′	Start	Stop
CPF	AACGAGCCCAGAGAACAGAG	20	59.39		2.00	0.00	8776	8795
				435				
CPR	CTGTGCGTGTTCTCCCTC	18	57.73		2.00	0.00	9210	9193
NiaF	CTGGACCGATCGGATTAAGA	20	56.53		8.00	2.00	5554	5573
				696				
NaiR	GGCAAGGTTAATGCTACCAGATCC	24	61.52		4.00	2.00	6249	6226
NibF	GCCGACACAAAGGACAAAGA	20	58.69		2.00	0.00	7757	7776
				216				
NibR	CTTCGGTTCCTTGCTCCTCT	20	59.39		2.00	0.00	7972	7953
AaeF1	CCTTCGAATAAAACCCCGCC	20	59.26		6.00	1.00	6860	6879
				652				
AeaR1	CGAATCGGGGATGTTGCTTT	20	58.91		5.00	0.00	7511	7492
AeaF2	GGTAAAGTCCCTCGAACCCA	20	59.02		4.00	0.00	847	828
				464				
AeaR2	AATAGTGGCGGGGTGATCTT	20	58.79		4.00	0.00	384	403
AeaF3	GGTAAAGTCCCTCGAACCCA	20	59.02		4.00	0.00	847	828
				211				
AeaR3	GCTGAAGGGGAGTCTGAGTT	20	59.02		3.00	1.00	637	656

Synthesis of dsRNA specific to WSMV genes Nia, Nib and CP was performed by extracting RNA from a WSMV positive control (Agdia, Elkhart, IN) using a RNeasy Plant RNA extraction kit (QIAGEN, Hilden, Germany) as per the manufactures protocol. Following RNA extraction, total RNA was reverse transcribed using MMLV reverse transcriptase (Thermo Fisher, Waltham, MA) with random hexamer primers as per the manufacturer's protocol. The resulting cDNA was then used as the template for PCR with primers specific to Nia, Nib and CP under the following reaction concentrations: 10µL of GoTaq (Promega, Madison, WI), 2µL of each 5µM forward and reverse primer, 2µL of template WSMV cDNA and nuclease free water up to 20µL reaction volume total. PCR was then performed under the following thermal conditions: 94°C for 4 min, followed by 40 cycles of 94°C for 20 sec, 55°C for 20 sec and 72°C for 50 sec, followed by a final extension of 72°C for 5 min. Following PCR, amplified products for WSMV genes Nia, Nib and CP were gel purified using a GFX™ PCR DNA and Gel Band

Purification kit (GE Healthcare, Chicago, IL). Each purified PCR product was then sequenced at the Oklahoma State Core Facility with capillary electrophoresis with a 3730xl DNA analyzer (Thermo Fisher, Waltham, MA).

Following product confirmation with sequencing, this amplified product was then used as the template for a second PCR using T7 (5´- TAATACGACTCACTATAGGG -3´) promotor amended primers for Nia, Nib and CP. This second PCR was performed under identical concentrations and thermal conditions to those described above. After this PCR, T7 amended products for Nia, Nib and CP were gel purified using a GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare, Chicago, IL). Following gel purification, this T7 amended, purified PCR product was then used in an RNA transcription reaction using a HiScribe™ T7 High Yield RNA synthesis kit (New England Biolabs, Ipswich, MA) following the manufacture's protocol. To ensure both the positive sense and negative sense RNA strands annealed properly after the transcription reaction, each sample was denatured at 95°C for 10 sec, followed by an incubation of 55°C for 10 min before being placed on ice for 2 min. The resulting dsRNA was then visualized with gel electrophoresis to confirm successful transcription, then gel purified using a Zymoclean™ Gel RNA Recovery Kit (Zymo Research, Irvine, CA).

2.3 Negative Control *A. aegypti* dsRNA

Whole, live *A. aegypti* were placed in -20°C overnight, then were subject to DNA extracted using 10% Cetyltrimethyl ammonium bromide (CTAB) with 1% polyvinylpyrrolidone (PVP) (Shahjahan, 1995). This DNA was then used as the template for PCR using primer pairs Aae1, Aae 2 and Aae3 respectively. Each PCR was performed using the same reaction concentrations and thermal conditions as described in section 2.1 for Nia, Nib and CP. Following PCR, resulting PCR product was then gel purified using a GFX™ PCR DNA and Gel Band

Purification kit (GE Healthcare, Chicago, IL). This purified PCR product was then used for a second PCR with T7 amended primer for Aae1, Aae2 and Aae3 following the same reaction concentrations and thermal conditions as those described in section 2.1. After PCR, T7 amended product was gel purified and used as the template for a transcription reaction with a HiScribe™ T7 High Yield RNA synthesis kit (New England Biolabs, Ipswich, MA) and gel purified following the protocol described above in section 2.1 for Nia, Nib and CP.

2.4 Systemic Movement of Injected dsRNA Transcripts

To establish if designed dsRNA constructs will travel systemically through a wheat seedling once injected into the stem, an assay was performed to first inject, then retrieve dsRNA from the roots, stem and leaves. To do this, 20µL of 0.5µg/µL Nia and Nib dsRNA in water were each injected into three, 10-day old wheat seedlings (Gallugher) with a $26G^{1/2}$ needle (PrecisionGlide, St. Louis, MO). After four hours, sterile scalpels were used to separate the leaves, stem and roots of each of the three wheat seedlings. Total RNA was extracted from leaves stem and roots of each of the seedlings using a RNeasy Plant RNA extraction kit (QIAGEN, Hilden, Germany) as per the manufacturer's protocol. RNA extraction was followed by reverse transcription with MMLV reverse transcriptase (Invitrogen, Carlsbad, CA) using random hexamer primers as per the manufacturer's protocol. This resulting cDNA was then used as a template for PCR using the reaction concentration and thermal conditions described for Nia, Nib and CP in section 2.1. After PCR, amplified product was visualized using gel electrophoresis to confirm if dsRNA had successfully circulated through the wheat seedling.

2.5 Preparation of Susceptible and Resistant Wheat

Wheat seeds of the varieties Gallagher, OK13804W and OK12621 were planted in 2" x 8" containers with silicon enriched horticulture grade soil (50%-60% Canadian sphagnum peat moss) (Sungro Horticulture, Agawam, MA). Planted wheat was maintained in the laboratory under a 12hr light, 12hr dark cycle at 21°C and watered as needed. To serve as a negative control, an equal number of seeds of each variety were also planted using an identical procedure and maintained in identical conditions. After nine days, wheat seedlings of every variety planted were kept in the dark for 24 hours prior to inoculation with dsRNA and WSMV.

2.6 Inoculation of Wheat with dsRNA and WSMV

Ten-day old Gallagher seedlings were injected at the base of the stem with 20µg of purified dsRNA specific to the WSMV gene Nib in water using a $26G^{1/2}$ needle (PrecisionGlide, St. Louis, MO). After four hours, each seedling was mechanically inoculated with WSMV from a positive control (AGDIA, Elkhart, IN) resuspended in 100µL of phosphate buffered-saline (PBS) (pH 7.2) containing 320 grit silicon carbide powder (Alfa Aesar, Haverhill, MA). The mechanical inoculation was performed by cutting the end off a pipette tip and pipetting 10µL of WSMV solution onto the largest leaf of each seedling, then with a gloved hand, spreading the solution across the surface of the leaf between the thumb and index finger.

A total of 21 Nib-dsRNA treated Gallagher seedlings were mechanically inoculated with WSMV using this method. A total of 21, ten-day old seedlings each of OK13804W and OK12621 were also inoculated with WSMV using the same method described above. To serve as a negative dsRNA control, 20µg of purified dsRNA specific to *A. aegypti* control dsRNA Aae3 was injected into the stem of 21, 10-day old Gallagher seedlings using the same method used for injecting Nib dsRNA. After four hours, these Aea3 dsRNA injected Gallagher seedlings

were then mechanically inoculated with WSMV using the method described above. All inoculated seedlings were then maintained in the laboratory at 21°C with 12hr light and dark and watered as needed. A diagram of the inoculation procedure can be found below (Fig 1). A table containing all treatment groups and the variety used for each can be found below (Table 2).

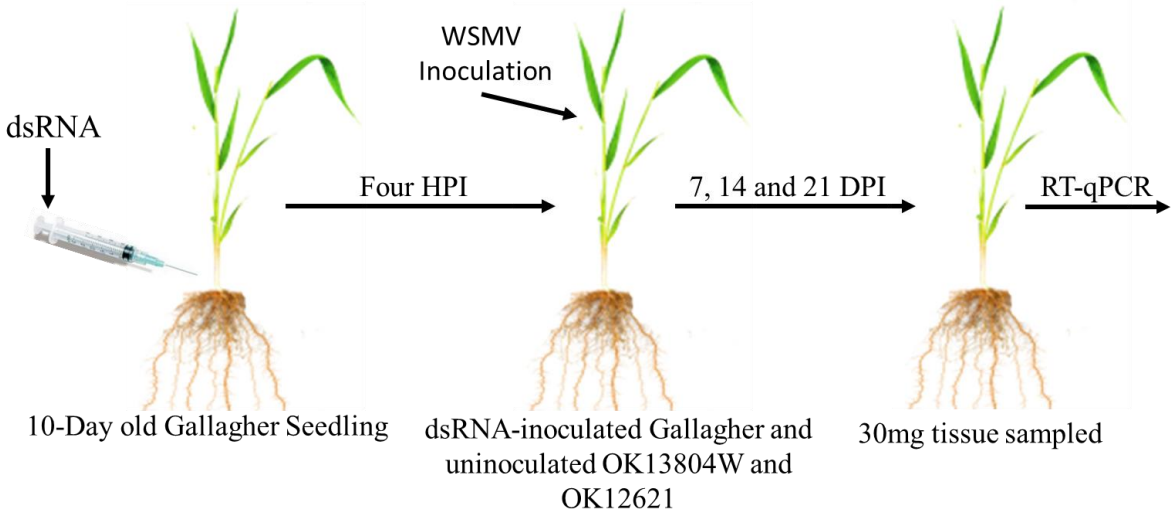


Figure 1. Diagram summarizing experimental design including dsRNA inoculation of Gallagher seedlings, WSMV inoculation of dsRNA-treated Gallagher and untreated OK13804 and OK12621. At 7, 14 and 21 DPI, each plant had 30mg of tissue sampled for RT-qPCR to assess copy number of WSMV.

Table 2. Summary of wheat variety, treatment, presence or absence of known WSMV resistance genes and the number of individual wheat plants within each treatment group (replications).

Wheat Variety	Treatment	Resistance Gene	Replications
Gallagher	WSMV	None	21
Gallagher	Nib-dsRNA + WSMV	None	21
Gallagher	Aae3-dsRNA + WSMV	None	21
Gallagher	No Treatment	None	21
OK13804W	WSMV	*Wsm1* + *Wsm2*	21
OK13804W	No Treatment	*Wsm1* + *Wsm2*	21
OK12621	WSMV	Unknown	21
OK12621	No Treatment	Unknown	21

2.7 Processing of WSMV Infected Tissue

Seven days after inoculation with dsRNA and WSMV, wheat tissue from each seedling among all treatment groups was collected, weighed, and 30mg were placed into a 1.5mL MCT, then stored at -80°C. This process was repeated at 14- and 21-days post inoculation for each treatment. At each of these three time points, wheat tissue was collected by removing the top of the most recently emerged leaf and was never collected from the WSMV inoculated leaf. After all tissue samples for every treatment group were collected across all three time points, every sample had total RNA extracted using TRIzol (Invitrogen, Carlsbad, CA). This RNA was then used in a reverse transcriptase reaction using MMLV reverse transcriptase (Promega, Madison, WI) as per the manufacture's protocol using a WSMV genome specific reverse primer (5´-TCCTGGTACTCGTGGATTTGT-3´). Details for the design and thermodynamic properties of this primer can be found in section 2.4 of the materials and methods section from chapter IV of this dissertation.

2.8 Development of qPCR Standard Curve for WSMV

Total RNA was extracted from a WSMV positive control (Agdia, Elkhart, IN) using TRIzol (Invitrogen, Carlsbad, CA) as per the manufacture's protocol. Resulting RNA was then used in a reverse transcription reaction using MMLV reverse transcriptase (Thermo Fisher, Waltham, MA) and random hexamer primers. cDNA generated from this reverse transcriptase was then used as the template for a PCR with WSMV specific primers WSMVF (5´- CGACAA-TCAGCAAGAGACCA-3´) and WSMVR (5´- TGAGGATCGCTGTGTTTCAG-3´). Details on the design and thermodynamic values for these primers can be found in section 2.1 of the materials and methods from Chapter III of this dissertation. PCR with WSMVF and WSMVR was carried out under the following reaction concentrations: 10µL GoTaq (Promega, Madison,

WI), 2µL each of 5µM WSMVF and WSMVR, 2µL WSMV template cDNA and 4µL nuclease free water to bring the total volume of the reaction to 20µL. PCR was carried out in a Biometra thermal cycler (Göttingen, Germany) under the following thermal conditions: 94°C for 4 min, followed by 40 cycles of 94°C for 20 sec, 58°C for 20 sec, 72°C for 20 sec, followed by a final extension at 72°C for 3 min.

The resulting 198bp PCR product was then gel purified using a GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare, Chicago, IL) and used as the template for a second enrichment PCR under identical reaction concentrations and thermal conditions as those described above. The resulting enriched 198bp product of WSMV was then gel purified with the GFX™ PCR DNA and Gel Band Purification kit and had its concentration measured using Nanodrop 2000 (Thermo Fisher, Waltham, MA). A 10-fold dilution series was then generated using this enriched, purified WSMV PCR product to be used as a standard curve in qPCR.

2.9 qPCR of dsRNA- and WSMV-Treated Wheat Samples

The cDNA obtained from wheat samples in section 2.7 above, were used as the template for qPCR in a RotoGene (Corbett, Mortlake, Australia) using the following reaction concentrations: 10µL of 2X Forget-Me-Not™ qPCR master mix with EvaGreen® (Biotium, Fremont, CA), 2µL each 5µM WSMVF and WSMVR primers, 1µL template cDNA, 3µL of 1X ROX reference dye and 2µL of nucleotide-free water. The thermal conditions for each qPCR was as follows: 95°C for 2 min, followed by 40 cycles of 95°C for 5 sec, 56°C for 10 sec and 72°C for 20 sec. qPCR for each sample was performed in duplicate alongside the standard curve described in section 2.8 and an NTC. WSMV copy number for each sample was calculated by comparing the fluorescence of each sample to the standard curve using the Rotor Gene 6000

Software (V.1.7.0), then normalizing the resulting number to the ROX reference dye in each reaction. ANOVA with Tukey's HSD was performed in Excel (Microsoft, Redmond, WA) to determine if a significant difference existed between the mean WSMV copy number for any treatment at 7-, 14- and 21-days post inoculation.

3.0 Results

3.1 dsRNA Design, Synthesis and Reclamation from Injected Wheat

Primers specific to a 696bp region of the WSMV gene Nia and a 216bp region of the WSMV gene Nib were successfully amplified with PCR. Primers specific to the WSMV CP gene failed to produce any PCR product (Fig. 1). Following PCR, gel purified PCR products for Nia and Nib were successfully amplified in a second PCR using T7-amended, Nia and Nib primers respectively. Following a second gel purification, T7-amended Nia and Nib product were then used as the template for transcription of Nia and Nib dsRNA (Fig. 2). After dsRNA for Nia and Nib were injected into 10-day old Gallagher seedlings, only dsRNA corresponding to the 216bp segment of Nib was reclaimed from the roots, stem and leaves of seedlings four hours after injection. The 696bp PCR product of Nia was not detected after attempted reclamation from dsRNA injected Gallagher seedlings after four hours (Fig. 3). Only Nib-dsRNA was selected for further study and used for the dsRNA-WSMV trial.

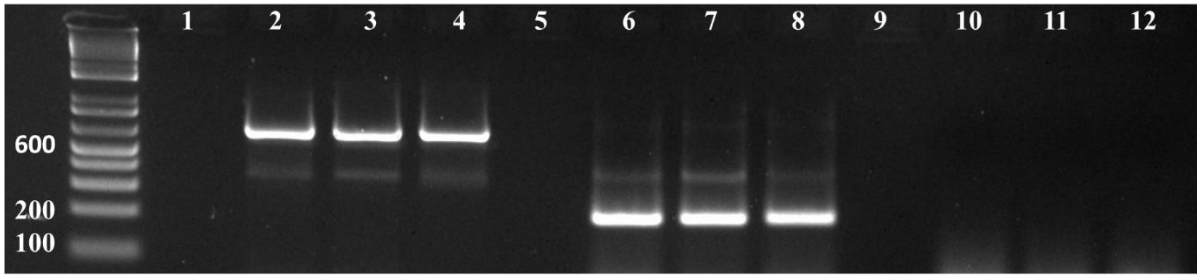


Figure 2. Gel image of PCR product for WSMV gene targets Nia (2-3), Nib (6-7) and CP (10-12) alongside a 100bp ladder (New England Biolabs, Ipswich, MA). Wells 1, 5 and 9 were empty. Gel electrophoresis was performed in a 2% agarose gel and run at 95V for 50 min. Gel image reveals successful amplification of predicted products for Nia and Nib, but not for CP.

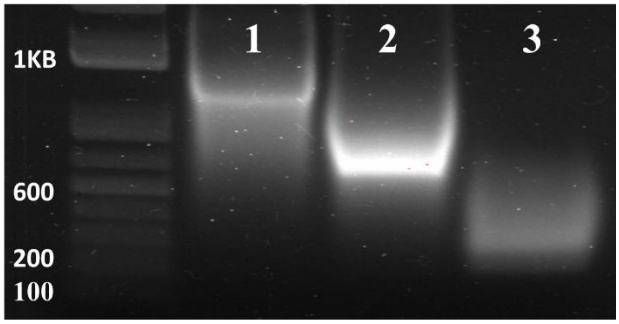


Figure 3. Gel confirmation of dsRNA for HiScribe™ T7 high yield RNA synthesis kit control dsRNA (1), Nia dsRNA (2) and Nib dsRNA (3). Gel electrophoresis was performed using 2% agarose and run at 95V for 45 min. Ladder used is a 100bp ladder (New England Biolabs, Ipswich, MA).

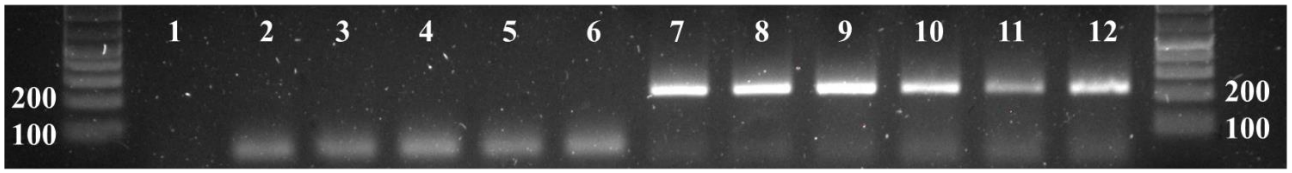


Figure 4. Gel confirmation of retrieval of Nia and Nib dsRNA from roots, stem and leaves of injected 10-day old Gallagher seedlings. Lanes 1-6 represent attempted retrieval of Nia target from roots (1-2), stem (3-4) and leaves (5-6). Lanes 7-12 represent attempted of retrieval of Nib dsRNA from roots (7-8), stem (9-10) and leaves (11-12). Results indicate successful retrieval of dsRNA of Nib with presence of predicted band at ~216bp for root, stem and leaf tissue. Bands corresponding to Nia of ~700bp were not seen, suggesting reclamation failure of Nia dsRNA. Gel electrophoresis was performed with 2% agarose and run at 95V for 45 min. Ladders used were NEB 100bp ladder (Ipswich, MA).

3.2 Development of Standard Curve and dsRNA Trial

A standard curve was successfully developed to quantify the number of copies of WSMV within a sample. It was comprised of a 10-fold dilution series of the following calculated concentrations of the 198bp, WSMV target region: $2.71x10^9$, $2.90x10^8$, $2.75x10^7$, $2.70x10^6$, $2.73x10^5$ and $2.29x10^4$. The background level of florescence as measured by the calculated NTC value was 4.66. The threshold of the standard was calculated to be 0.3652 normalized relative fluorescence units (RFU), the R value was 0.999 and the R^2 value was calculated at 0.999. Figure four below contains two graphs plotting the normalized fluorescence to the PCR cycles and the critical threshold values to the standard concentration respectively.

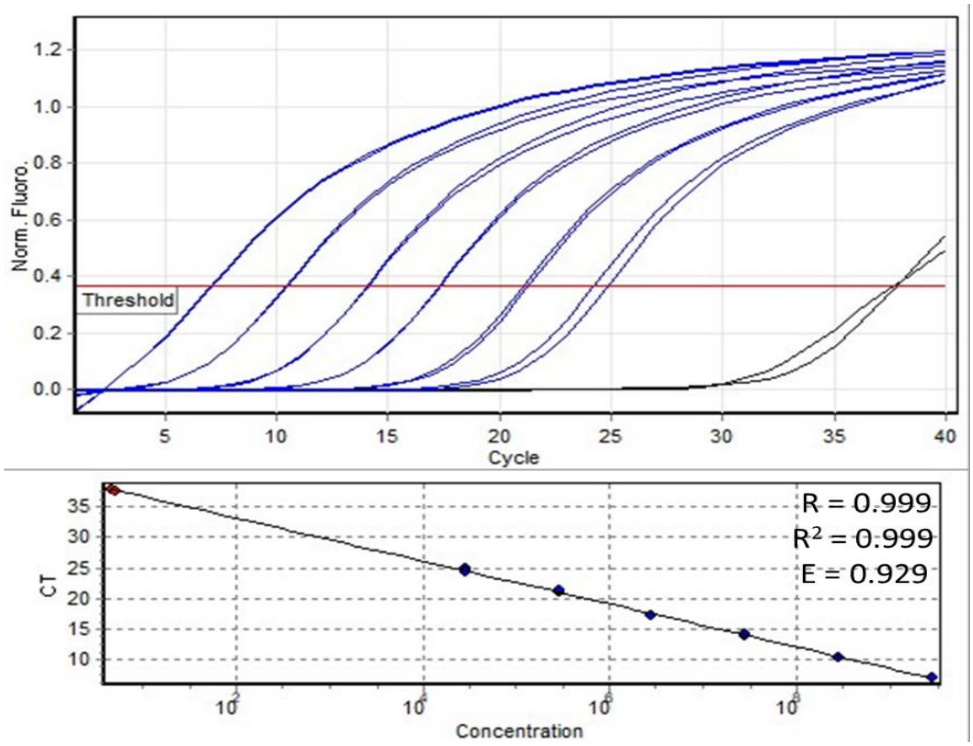


Figure 5. Designed standard curve for the quantification of WSMV within dsRNA- and WSMV-treated wheat samples. The standard curve is comprised of six known concentrations of WSMV in duplicate and a NTC in duplicate. The calculated values for R, R^2 and reaction efficiency (E) were determined to be 0.999, 0.999 and 0.929 respectively.

After seven days, WSMV was detectable in all treatment for which data was available. ANOVA and Tukey's HSD analysis revealed that there was no significant difference in WSMV copy number for any of the treatment groups with the exception of OK13804W, which was significantly higher than all other treatment groups ($p = 0.033$). At 14 DPI, the number of WSMV copies among all treatment groups did not significantly increase overall, but more differences were determined between treatments. ANOVA and Tukey's HSD were used to determine that Gallagher treated with WSMV dsRNA had significantly fewer WSMV copies than untreated Gallagher ($p = 0.001$), Gallagher treated with control dsRNA ($p =0.002$) and OK13804W ($p = 0.001$). It was also noted that OK12621 had significantly fewer WSMV copies than OK13804W ($p = 0.005$).

At 21 DPI, the total number of WSMV copies significantly increased for all treatments ($p = 0.001$). Gallagher treated with WSMV dsRNA ($p = 0.685$) had significantly fewer WSMV copies compared to all other treatment groups except for OK12621 ($p = 0.273$). Analysis of other treatments revealed that untreated Gallagher and Gallagher treated with control dsRNA had significantly more WSMV copies than all other treatment groups ($p = 0.001$) and that no significant difference was determined between OK13804W and OK12621 ($p = 0.684$). Lastly, it was determined that no significant difference in WSMV copy number was found between untreated Gallagher and Gallagher treated with control dsRNA. Graphs comparing the WSMV copy number across all treatment groups at 7, 14 and 21 DPI can be found below (Fig. 5). Graphs comparing the change in WSMV copies within the treatment groups Gallagher, OK13804W and OK12621 at 7, 14 and 21 DPI can also be found below (Fig. 6). Finally, a comparison in the number of WSMV copies for treatment groups of Gallagher treated with WSMV dsRNA and Gallagher treated with control dsRNA can be found below (Fig. 7).

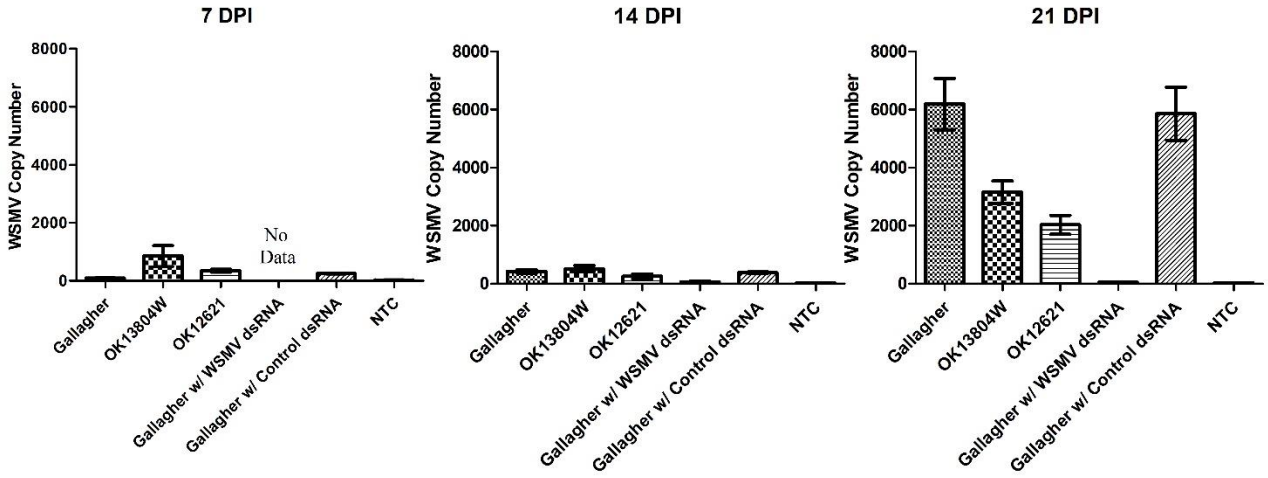


Figure 6. Graphs comparing the transcript copies of WSMV for dsRNA and WSMV inoculated wheat plants. At 7 DPI, only one treatment group, OK13804W, was found to have a significantly higher WSMV copy number when compared to Gallagher ($p = 0.033$). No data is available for Gallagher treated with WSMV dsRNA at 7 DPI due to loss of samples. At 14 DPI, significant differences in WSMV copy number were found between: Gallagher and Gallagher with WSMV dsRNA ($p = 0.001$), OK13804W and OK12621 ($p = 0.005$), OK13804W and Gallagher with WSMV dsRNA ($p = 0.001$) and between Gallagher with WSMV dsRNA and Gallagher with control dsRNA ($p = 0.002$). At 21 DPI, significant differences in WSMV copy number were found between: Gallagher and OK13804W ($p = 0.008$), Gallagher and OK12621 ($p = 0.001$), Gallagher and Gallagher with WSMV dsRNA ($p = 0.001$), OK13804W and Gallagher with WSMV dsRNA ($p = 0.019$), OK13804W and Gallagher with control dsRNA ($p = 0.023$), OK12621 and Gallagher with control dsRNA ($p = 0.001$) and Gallagher with WSMV dsRNA and Gallagher with control dsRNA ($p = 0.001$). For all comparisons $n = 21$ and a combined NTC from all qPCRs is present. All error bars represent standard error measurement.

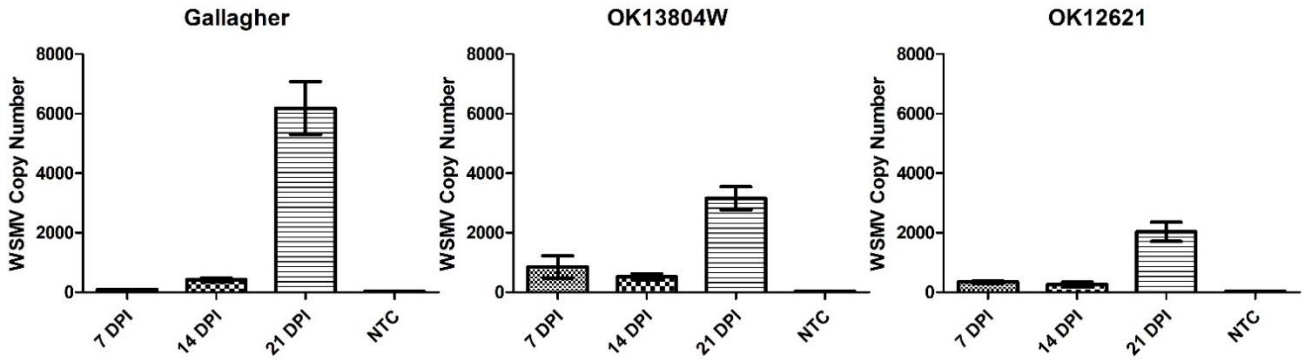


Figure 7. Graphs comparing the mean WSMV copy number for Gallagher, OK13804W and OK12621 at 7, 14 and 21 DPI. For Gallagher, OK13840W and OK12621, the WSMV copy number at 21 DPI was significantly higher than both 7 and 14 DPI ($p = 0.001$) ($n = 21$). No significant difference in WSMV copy number existed for any treatment between 7 and 14 DPI ($p = 0.848$, 0.713 and 0.899 respectively). The mean background florescence level for qPCRs can be seen in the NTC. All bars represent standard error of the mean.

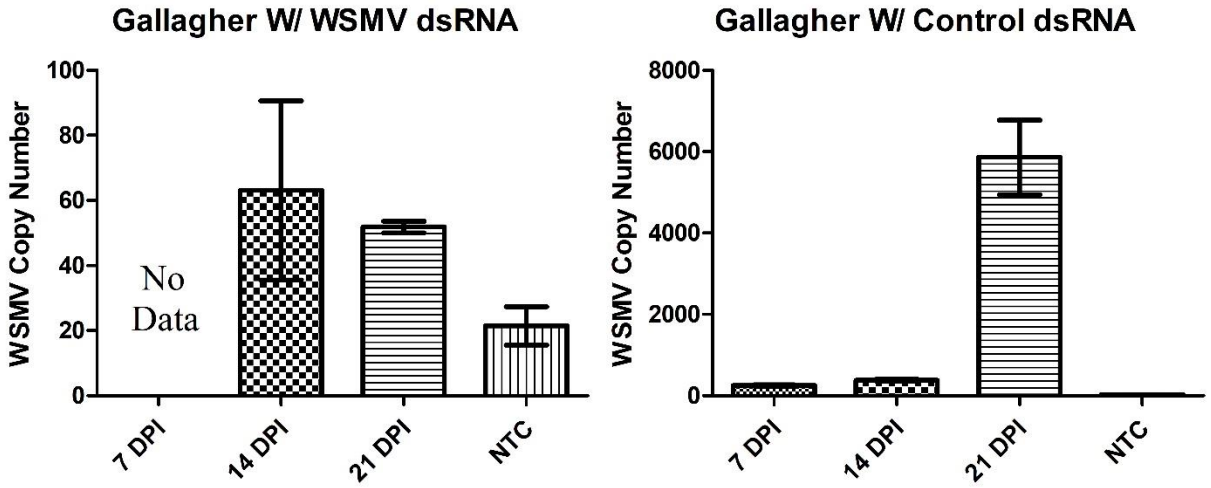


Figure 8. Graphs displaying the mean WSMV copy number for treatments Gallagher with WSMV dsRNA and Gallagher with control dsRNA and 7, 14 and 21 DPI. For Gallagher with WSMV dsRNA, the data for 7 DPI is not available due to sample loss, and no significant difference was observed between WSMV copy number at 14 and 21 DPI ($p = 0.685$). For Gallagher treated with control dsRNA, the mean WSMV copy number at 21 DPI was significantly higher than WSMV copy number at both 7 and 14 DPI ($p = 0.001$). For all comparisons, $n = 21$ and all bars represent the standard error of the mean.

Discussion

Plant viruses pose a unique challenge to agriculture when compared to bacteria and fungi in part due to the relatively limited options available for their management. The traditional methods of management of diseases caused by plant viruses include the management of arthropod vectors and the use of resistant varieties. These strategies have proven effective in many cases but have limits to their success like any management strategy. Over time, arthropods may become resistant to specific chemistries, leading to a decrease in the effectiveness of the control and a corresponding increase in transmission of the pathogens they carry (Roberts & Andre, 1994). Breeding resistant plant varieties offers another defense against virus infection through resistance to arthropod vectors or to the virus itself. Much like chemical control, this resistance can break down over time due to adaptation within the vector species or against the virus itself for various reasons (Acosta-Leal et al., 2008; Tatineni et al., 2016). While it is possible for new chemical treatments and resistant varieties to be developed to cope with these changes, it is a process that takes many years and significant capital investment to accomplish (Mayer & Furtan 1999).

The use of GM crops provides a solution to some of the shortcomings of chemical control and breeding resistance. One possible solution to this is the use of transgenes specifically targeted against a specific arthropod vector or plant virus while minimizing the potential for off-target effects often associated with pesticide use (Bird et al., 1996). Another advantage of the use of GM resistant varieties over traditionally bred resistant varieties is the ability to choose which genes become part of the plant's chromosomes and which do not. When using traditional methods for breeding resistance, there is often a yield penalty or other undesirable traits associated with the variety (Van Beuningen & Kohli, 1990). The ability to select for resistance

genes while simultaneously excluding undesirable ones can increase the rate at which a resistant variety is released. The problem with GM crops however is the poor public perception of them worldwide (Daniell 1999), resulting in them being banned in many countries worldwide.

The objective of this study was to provide an initial proof of concept for a possible middle ground between the use of single dose treatments and the use of transgenic varieties. It was hypothesized that by injecting dsRNA specific to WSMV directly into a susceptible wheat host prior to inoculation with WSMV, it would be possible to change its phenotype to that of a WSMV resistant variety. After designing and testing three potential dsRNA constructs specific to WSMV genes Nia, Nib and CP, only the dsRNA for Nib was determined to be a potential candidate. For CP, failure occurred during the initial PCR meant to amplify a 435bp segment of the gene preventing its use downstream as a candidate to synthesize dsRNA from. One possible reason for this failure may have been a discrepancy in the nucleotide sequences used to design the primers used in this study and the CP gene sequence in the template used for PCR. It is also known that in the case of many viruses, the CP gene is less conserved than other genes involved in replication, such as Nib, which may account for possible nucleotide polymorphisms found there (Schneider & Roossinck, 2001).

Reclamation failure of Nia may be caused by one of two reasons. First, the dsRNA product for Nia was larger (696bp) than the dsRNA product for Nib (216bp). While it is known that during RNA silencing within a plant host, siRNA (21bp-25bp) are able to move freely through the plasmodesmata of neighboring cells, it is not clear what the limit of nucleic acids movement from cell to cell is (Hamilton & Baulcombe, 1999). It is possible that the ~700bp Nia-dsRNA was too large to move freely through the plant while the smaller, ~200bp Nib-dsRNA

was. However, if it were the case that Nia-dsRNA were unable to travel systemically due to its size, that would not explain why it was not detected in the stem tissue where it was originally injected. It is possible that either its larger size, or some architectural property of its nucleic acid composition made it less stable relative to Nib-dsRNA.

When sampling wheat tissue at 7-, 14-, and 21DPI, only the top 30mg of tissue from the youngest, most recently emerged wheat leaf was taken. This was done for two reasons. First, it limited the chance of measuring any WSMV particles that may have been left on the surface of older leaves during inoculation and second, it maximized the ability to detect WSMV for each treatment due to increased virus replication in younger, faster growing tissue (Cooper & Jones, 1983). During the processing of samples, reverse transcription was performed using a reverse primer specific to WSMV as opposed to using random hexamer primers. This was done to limit the cDNA present in each sample to that belonging to WSMV, as any off-target amplification differences between varieties may have reduced the accuracy of qPCR and the reliability of results.

Over the course of 21 days, WSMV copy number was shown to increase significantly from the time of inoculation except for Gallagher treated with WSMV-dsRNA. It was determined that treating Gallagher with WSMV-dsRNA had reduced virus expression of WSMV to the point where there was no significant difference between the treated Gallagher and OK12621, a known resistant variety. When comparing both resistant varieties (OK13804W and OK12621) to the susceptible variety (Gallagher) the rate of WSMV replication appeared to be significantly slower for the resistant varieties. All treated plants in this study were maintained in the laboratory at 21°C, which is above the known effective temperature for *Wsm1* and *Wsm2*, the

two resistant genes found in OK13804W. It is not known how much this difference in temperature affected the results of this study, but it was shown that when compared to a susceptible variety, WSMV copy number was still significantly lower in OK13804W. Currently, the genetic source of resistance for OK12621 is unknown, so it is not clear if temperature may have played a role in its resistance to WSMV or not.

There is strong evidence to suggest that treatment with dsRNA specific to the Nib gene of WSMV lead to an increase in resistance against WSMV in Gallagher. When compared to the dsRNA control, WSMV copy number was significantly lower in the Nib-dsRNA treated Gallagher and there was no significant difference in the WSMV copy number of untreated Gallagher and Gallagher treated with control dsRNA. This control was present to determine if the presence of dsRNA alone or the injection treatment method was enough to produce an effect on WSMV copy number. It was determined that neither the non-WSMV specific dsRNA or the injection method significantly changed WSMV copy number over time. However, it is important to note that there are several elements of the design of study which prevent any conclusion regarding the use of dsRNA treatments to induce virus resistance in natural settings.

The purpose of this study was to determine if the presence of dsRNA specific to WSMV prior to infection would significantly reduce virus replication. It was determined that this occurred for at least the 21 days after inoculation that were studied. More work still needs to be done to assess how long an effect like this may last, as well as how differences in dsRNA dose or the time between dose and virus inoculation influence WSMV expression. Another important factor to consider is the inoculation method used. In this study, inoculation was performed only once, and mechanically. In the field, WSMV infection primarily occurs via its mite vector, where

many vectors may be transmitting low titers of WSMV over time, as opposed to a high titer at one time. It is not clear from the results of this study just how long this resistance may last if this injection method would be applied in a field study. It may be the case that to ensure long-term resistance to WSMV, dsRNA may have to be applied several times over the growing season. Something that is not feasible when considering the large acreages on which wheat is planted. Therefore, a strategy for a long-term and efficient delivery of dsRNA to wheat would be required.

One possible strategy for the long-term delivery of dsRNA would be the use of a transformed endophyte. It has been shown that endosymbiotic bacteria can be transformed to transcribe dsRNA hairpins capable of inducing post transcriptional gene silencing within their host (Whitten et al., 2016). If an endophytic bacteria species could be transformed to produce dsRNA specific to WSMV then deliver it to its host plant. This strategy would provide many of the benefits associated with the use of transgenic crops, without ever having to alter the genome of wheat. The next step would be to select an endophyte that is readily transformable and would not end up in the final grain product. One such genus of endosymbiont is *Streptomyces*, a highly studied and commonly found root endosymbiont of wheat and a bacterium known for the wide variety of transcripts it can produce (Coombs et al., 2004). Based on current laws and regulations in the United States, wheat produced in this way would not be considered GM and could be sold commercially.

REFERENCES

Acosta-Leal, R., Fawley, M. W., & Rush, C. M. (2008). Changes in the intraisolate genetic structure of Beet necrotic yellow vein virus populations associated with plant resistance breakdown. *Virology*, *376*(1), 60-68.

Baley, G. J., Talbert, L. E., Martin, J. M., Young, M. J., Habernicht, D. K., Kushnak, G. D. & Bruckner, P. L. (2001). Agronomic and end-use qualities of Wheat streak mosaic virus resistant spring wheat. *Crop Science*, *41*(6), 1779-1784.

Bird, S. L., Esterly, D. M., & Perry, S. G. (1996). Off-target deposition of pesticides from agricultural aerial spray applications. *Journal of Environmental Quality*, *25*(5), 1095-1104.

Campbell, T. N., & Choy, F. Y. (2005). RNA interference: past, present and future. *Current Issues in Molecular Biology*, *7*, 1-6.

Carpenter, J. E. (2010). Peer-reviewed surveys indicate positive impact of commercialized GM crops. *Nature Biotechnology*, *28*(4), 319.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N. & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819-823.

Coombs, J. T., Michelsen, P. P., & Franco, C. M. (2004). Evaluation of endophytic actinobacteria as antagonists of Gaeumannomyces graminis var. tritici in wheat. *Biological Control*, *29*(3), 359-366.

Cooper, J. I., & Jones, A. T. (1983). Responses of plants to viruses: proposals for the use of terms. *Phytopathology*, *73*(2), 127-128.

Daniell, H. (1999). GM crops: public perception and scientific solutions. *Trends in Plant Science, 4*(12), 467-469.

Davis, L., & Maizels, N. (2014). Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of the National Academy of Sciences*, *111*(10), E924-E932.

De Wolf, E., Bockus, W., Whitworth, J. (2015) Wheat Variety Disease and Insect Ratings 2015. Kansas State Research Extension, Kansas State University Agricultural Experiment Station and Cooperative Extension Service, MF991

Diao, X., Hazell, P., & Thurlow, J. (2010). The role of agriculture in African development. *World development*, *38*(10), 1375-1383.

Edwards, J., Hunger, R., Carver, B., Royer, T. (2012) 2012 Wheat variety comparison. Oklahoma Cooperative Extension Service. Division of Agricultural Sciences and Natural Resources, Oklahoma State University, PSS-2142.

Fahim, M., Ayala-Navarrete, L., Millar, A. A., & Larkin, P. J. (2010). Hairpin RNA derived from viral NIa gene confers immunity to wheat streak mosaic virus infection in transgenic wheat plants. *Plant Biotechnology Journal*, *8*(7), 821-834.

Flor, H. H. (1947). Inheritance of reaction to rust in flax. *J. agric. Res*, *74*, 241-262.

García-Arenal, F., & McDonald, B. A. (2003). An analysis of the durability of resistance to plant viruses. *Phytopathology*, *93*(8), 941-952.

Graybosch, R. A., Peterson, C. J., Baenziger, P. S., Baltensperger, D. D., Nelson, L. A., Jin, Y. & Martin, T. J. (2009). Registration of 'Mace' hard red winter wheat. *Journal of Plant Registrations*, *3*(1), 51-56.

Haley, S. D., Martin, T. J., Quick, J. S., Seifers, D. L., Stromberger, J. A., Clayshulte, S. R. & Gill, B. S. (2002). Registration of CO960293-2 wheat germplasm resistant to Wheat streak mosaic virus and Russian wheat aphid. (Registrations of Germplasm). *Crop Science*, *42*(4), 1381-1383.

Hamilton, A. J., & Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, *286*(5441), 950-952.

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., & Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science*, *293*(5532), 1146-1150.

Mayer, H., & Furtan, W. H. (1999). Economics of transgenic herbicide-tolerant canola: The case of western Canada. *Food Policy*, *24*(4), 431-442.

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I. & Gregory, P. D. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnology*, *25*(7), 778.

Roberts, D. R., & Andre, R. G. (1994). Insecticide resistance issues in vector-borne disease control. *The American Journal of Tropical Medicine and Hygiene*, *50*(6_Suppl), 21-34.

Sander, J. D., Cade, L., Khayter, C., Reyon, D., Peterson, R. T., Joung, J. K., & Yeh, J. R. J. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature Biotechnology*, *29*(8), 697.

Schneider, W. L., & Roossinck, M. J. (2001). Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. *Journal of Virology*, *75*(14), 6566-6571.

Shahjahan, R., Hughes, K., Leopold, R., & Devault, J. (1995). Lower incubation temperatures increase yield of insect genomic DNA isolated by the CTAB method. *Biotechniques, 19*(3), 332-4.

Shukla, D. D., Frcnkel, M. J., & Ward, C. W. (1991). Structure and function of the potyvirus genome with special reference to the coat protein coding region. *Canadian Journal of Plant Pathology, 13*(2), 178-191.

Sivamani, E., Brey, C. W., Dyer, W. E., Talbert, L. E., & Qu, R. (2000). Resistance to wheat streak mosaic virus in transgenic wheat expressing the viral replicase (NIb) gene. *Molecular Breeding*, *6*(5), 469-477.

Sivamani, E., Brey, C. W., Talbert, L. E., Young, M. A., Dyer, W. E., Kaniewski, W. K., & Qu, R. (2002). Resistance to wheat streak mosaic virus in transgenic wheat engineered with the viral coat protein gene. *Transgenic Research*, *11*(1), 31-41.

Speranza, C. I., Kiteme, B., & Wiesmann, U. (2008). Droughts and famines: the underlying factors and the causal links among agro-pastoral households in semi-arid Makueni district, Kenya. *Global Environmental Change*, *18*(1), 220-233.

Stenger, D. C., Hein, G. L., Gildow, F. E., Horken, K. M., & French, R. (2005). Plant virus HC-Pro is a determinant of eriophyid mite transmission. *Journal of Virology*, *79*(14), 9054-9061.

Tatineni, S., & French, R. (2014). The C-terminus of Wheat streak mosaic virus coat protein is involved in differential infection of wheat and maize through host-specific long-distance transport. *Molecular Plant-Microbe Interactions*, *27*(2), 150-162.

Tatineni, S., Wosula, E. N., Bartels, M., Hein, G. L., & Graybosch, R. A. (2016). Temperature-dependent Wsm1 and Wsm2 gene-specific blockage of viral long-distance transport provides resistance to Wheat streak mosaic virus and Triticum mosaic virus in wheat. *Molecular Plant-Microbe Interactions*, *29*(9), 724-738.

USDA (United State Department of Africulture), (2016) Crop Production 2016 Summary. National Agricultural Statistics Service, ISSN: 1057-7823.

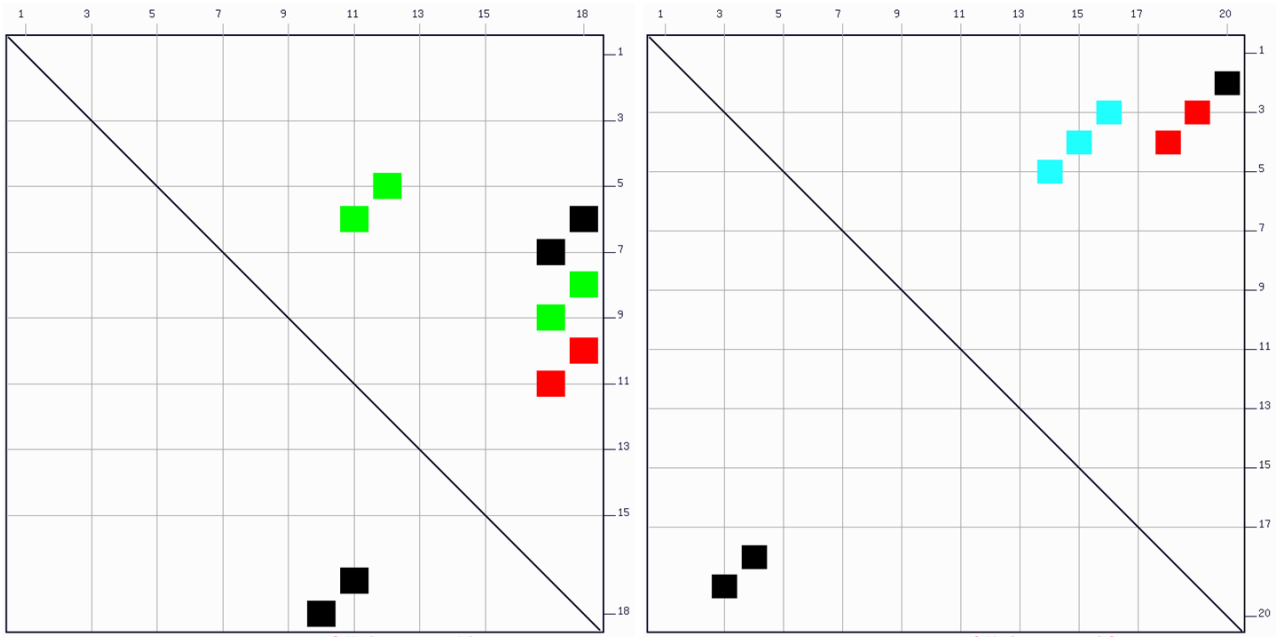
Van Beuningen, L. T., & Kohli, M. M. (1990). Deviation from the regression of infection on heading and height as a measure of resistance to Septoria tritici blotch in wheat. *Plant Disease*, *74*(7), 488-493.

Verchot, J., & Carrington, J. C. (1995). Evidence that the potyvirus P1 proteinase functions in trans as an accessory factor for genome amplification. *Journal of Virology*, *69*(6), 3668-3674.

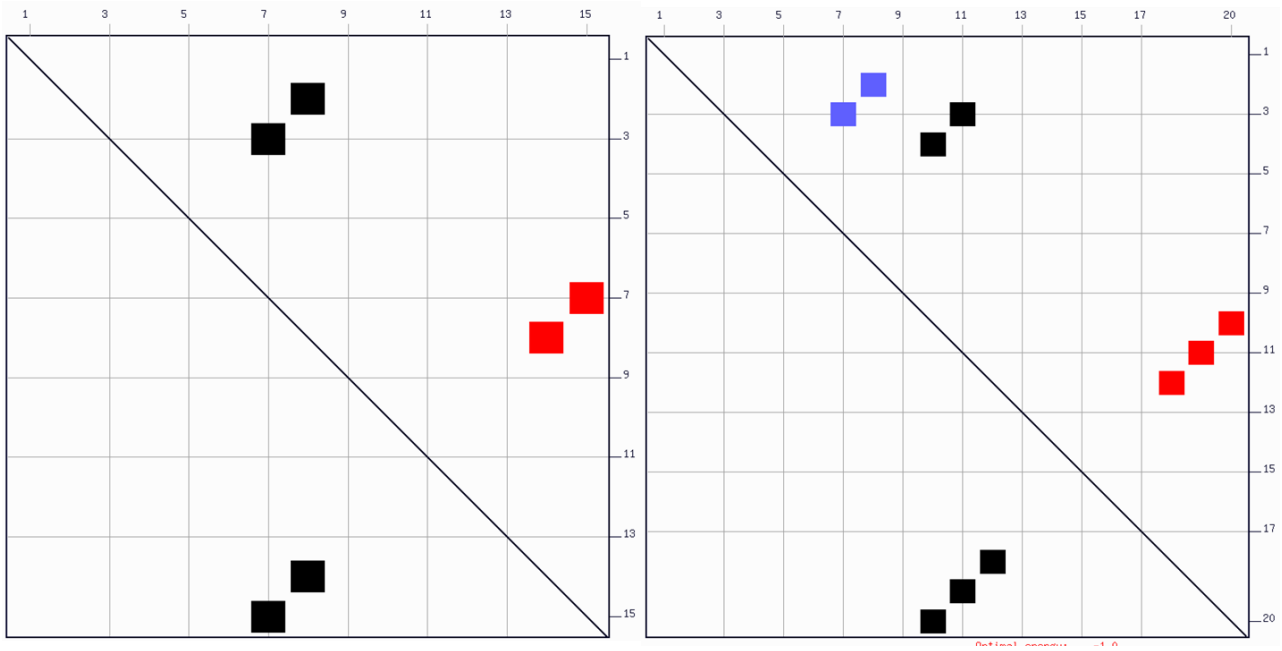
Whitten, M. M., Facey, P. D., Del Sol, R., Fernández-Martínez, L. T., Evans, M. C., Mitchell, J. J. & Dyson, P. J. (2016). Symbiont-mediated RNA interference in insects. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1825), 20160042.

APPENDICIES

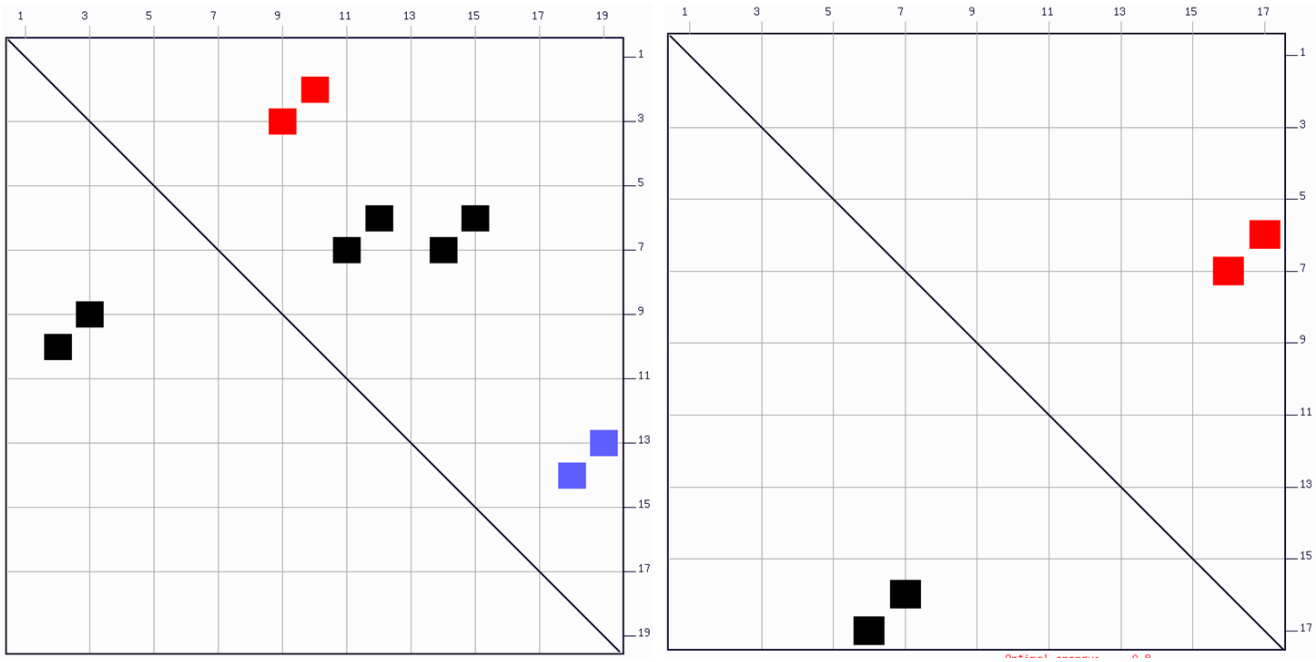
Appendix 1.1. mFold evaluation of primer pair MMVF (ΔG = 0.7) and MMVR (ΔG = 0.2) for the detection of MMV in multiplex.



Appendix 1.2. mFold evaluation of primer pair WSMVF (ΔG = 0.5) and MMVR (ΔG = 1.0) for the detection of WSMV in multiplex.



Appendix 1.3. mFold evaluation of primer pair BYDVF (ΔG = 0.6) and BYDVR (ΔG = 0.0) for the detection of BYDV in multiplex and the discrimination of BYDV-PAV, BYDV-PAS and BYDV-MAV using HRM.



Appendix 2.1. Designed, curated e-probes for the detection of *Agropyron mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
AgMV	AGMV1	CGAAACTGGAACTGGAATCACGTTCGATCA	160	189
AgMV	AGMV2	GAGTGAGTTTAAGGTCATACCAAAACAGCG	415	445
AgMV	AGMV3	ACACGGAACATGTAAAACCTATTACCAAGG	687	717
AgMV	AGMV4	ATTGCAGAAACAATGCTAGTCGACGCACGT	791	820
AgMV	AGMV5	CACATTTCCAATGCGGAATGTGTGCTGCAA	1008	1037
AgMV	AGMV6	CGGGTATACAAATAACGCAAGAGATTGCAA	1185	1215
AgMV	AGMV7	TAGAAGCATTATGCAAATCACGCAAGTACT	1246	1276
AgMV	AGMV8	TCGGTGAATGGCCAACCATGCGGTCAGTTG	1971	2001
AgMV	AGMV9	GGATTACATTTGTCACACCTGAAATTGGAA	2964	2993
AgMV	AGMV10	ACTCGAAAAGGAGGACTTGCTACTTGACAT	3094	3124
AgMV	AGMV11	CCAATAACAGAACAATGCTTAAGGAATTCA	3915	3945
AgMV	AGMV12	TGCGATGCGAAACCGATCTTTCCCGTCAAA	4070	4100
AgMV	AGMV13	ATGATGCAACAGTGCTTGGACACGACATAT	4164	4194
AgMV	AGMV14	TATTAGTTTATGTTGCGAGCTACAATGAAG	4191	4221
AgMV	AGMV15	AATGGCGTCACCCTCGCAGTCGATTATCTG	4367	4397
AgMV	AGMV16	ACGCATCCAGCGAATTGGTCGCGTTGGAAG	4483	4513
AgMV	AGMV17	AGAGTCAATTGCCACTACAGCAGCCTTCAA	4573	4603
AgMV	AGMV18	GAATGTGCTCTCTCGGTGTACCCGTGAGCA	4648	4677
AgMV	AGMV19	CGATATAAACTAAGGGAATCAGAGATTAAG	4778	4808
AgMV	AGMV20	GTAAGTCTCAGCGCGATAGCACAATGTCTG	5165	5195
AgMV	AGMV21	GAATTACCCTTGCGTTACTCTTGTAGAACA	5314	5344
AgMV	AGMV22	AAGTCATTTGTAGCTTGGAGTACTATTTTG	6132	6162
AgMV	AGMV23	CACGGTGCAAAACATGCAGAAGCTCAAGAT	6277	6307
AgMV	AGMV24	TGTGATCACATGCCCTAAGGACATGCAACC	6334	6364
AgMV	AGMV25	CCCAATGGTCTCGGAAGCAAGCGTGACAAC	6451	6481
AgMV	AGMV26	GCGGGCACTGTGGCTTGCCGATAGTTTCAC	6528	6558
AgMV	AGMV27	TTGGAGCTCTTTATGGCGGTAAGAAGGAGA	7194	7224
AgMV	AGMV28	TCTGGACACACTTCTTGCAGCGAAGGGTTG	7390	7420
AgMV	AGMV29	AACCAATTCTATAGCAAGCACCTCGAAGGC	7436	7465
AgMV	AGMV30	CGTTGACAATACATTAATGGTTGTCTTAGC	7759	7789
AgMV	AGMV31	TTATCCGGTACTTTGCTAACGGTGATGATT	7842	7872
AgMV	AGMV32	TGGAGTGGGACCGAAGTCATGAGCCTGAGT	8067	8097
AgMV	AGMV33	ATGCCTTCGACTTCTATGAAATAACATCCG	9156	9786
AgMV	AGMV34	CAAATCGTGTGCGTGAGGCGCACTTGCAGA	9192	9222
AgMV	AGMV35	GTTTCAACACGAGGGCATGAAACTTGTTGC	9498	9528

Appendix 2.2. Designed, curated e-probes for the detection of *Brome mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
BMV	BMV1	GTCAAGTTCTATCGATTTGCTGAAGTTGAT	76	106
BMV	BMV2	CAGGTTGCGCAACAGTTATCTGCGCAGATT	155	185
BMV	BMV3	TGCAAGAAAGCGATGATTTCGATGAAGTCC	516	546
BMV	BMV4	GAGACACTACGTCACTGTGTATGGTTTGAA	965	995
BMV	BMV5	GTTATTATTAACGGTCAGGCTATCATGGCT	1184	1214
BMV	BMV6	AGAACATTGGCTAGCAGATTTCCCCTACTA	1397	1427
BMV	BMV7	GAGTTCGAGCCCTGAGTCCGTCAGTGATGA	1660	1690
BMV	BMV8	TGAGGTACCAACGGACCCTCGTGGCATATC	1735	1764
BMV	BMV9	TAAAACCACTGCCATAAAAGATGCATTCCG	2143	2173
BMV	BMV10	TGCATTCCGTATGGGAGAGGACCTAATTGT	2164	2193
BMV	BMV11	CTTAACACGACACAAGAAGTCCTTTGAGTA	2878	2908
BMV	BMV12	CGCCTTTGAGAGTTACTCTTTGCTCTCTTC	3103	3133
BMV	BMV13	GGGATGATGATTTCGTTCGCCAGGTCCCGT	3353	3382
BMV	BMV14	TGCAGGAGCCGGCAGACGGAGTTGCCATTG	3443	3472
BMV	BMV15	GGGCTCTATTTGCGACACCGTCCAACAAAT	3558	3588
BMV	BMV16	ACAAATGGTTCAACAGTTCACCGATAGACC	3582	3611
BMV	BMV17	GATGACTGGTATCCCGAGGATACTAGTGAT	3688	3718
BMV	BMV18	TTCTTGAAGCGGACCTAAGCAAATTTGATA	4718	4748
BMV	BMV19	ACTGTGACTGTGCAATATTTTCAGGAGATG	4982	5011
BMV	BMV20	GCGCTTAGCTAAGCGAAAGATTCTGCGTGA	5187	5217
BMV	BMV21	TCCTTCTGTGATCGAATGAAGTTTATTAAT	5245	5274
BMV	BMV22	CTCCCTGTCAAACGGATCGGACGTTTACAT	5608	5638
BMV	BMV23	GTCGCTGTTCTCTGAAAAGGCTGTGAAAGA	6285	6315
BMV	BMV24	ATAGACCTGGTGCCAAAGTCACACGTATCT	6379	6408
BMV	BMV25	CATGAACGTTCCACGCATCGTTTGTTTTCT	6468	6497
BMV	BMV26	GCGATTCTGGTAAGGCTGCTCGTGCTGGAG	6548	6578
BMV	BMV27	ATCAGGAGGCCACAATTCAGTTGTCGGCTT	6596	6626
BMV	BMV28	CAGTTACTCATGCGTATTGGCAAGCTAATT	6758	6788
BMV	BMV29	TTTCAAAGCGAAGCCCAACAACTATAAGTT	6786	6815
BMV	BMV30	GGAACTGGTAAGATGACTCGCGCGCAGCGT	7355	7384
BMV	BMV31	ATCGTTGGACCGCTAGGGTCCAACCAGTAA	7404	7433
BMV	BMV32	AGGAGCTTAAGGTCGGCAGGGTGCTGCTTT	7590	7620

Appendix 2.3. Designed, curated e-probes for the detection of *Barley stripe mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
BSMV	BSMV1	AAGGACCATCGTCGATTCCGTGGATAAGAA	210	240
BSMV	BSMV2	ATAAGAAAGCGGTCAGTCGCAAGCATGAAG	233	262
BSMV	BSMV3	CTCGATGGAGATGTGTTTTGTGAGAACACT	682	711
BSMV	BSMV4	GAAAAGGTATAACACAGGCTTATGGGTGTT	818	847
BSMV	BSMV5	GGTCAGAAGGAAGGCATTTTACCTTCCGTG	874	903
BSMV	BSMV6	ATCACGGATGTCACCGCAGCAATGTATCAT	1087	1117
BSMV	BSMV7	AATCTCCAAGTTGGCAGTGACTCTGTACCT	1401	1430
BSMV	BSMV8	TAACAGGGACGGTGACCGAATGCGAGAAGG	1796	1826
BSMV	BSMV9	CACTCAAGTTCAGATTCTGAACACTCTATG	1963	1993
BSMV	BSMV10	TATCAACTGTGTAGCGTGATTTGTGAAAGG	2167	2197
BSMV	BSMV11	TTTGTGACAACACTAAATTGTGTAACAATT	2498	2527
BSMV	BSMV12	TGTACCGCAAGCTGATAGGTTTCATTTTGA	2796	2826
BSMV	BSMV13	TGGTATTGCACATACACCTATTCGTACTTT	3462	3491
BSMV	BSMV14	AGGTCCCGTCCTTACGCTTTCATCACCTGT	3903	3932
BSMV	BSMV15	GGTGAAACGGTTGCTACAGGAACAACCCCG	3972	4001
BSMV	BSMV16	AGGTCCCGTCCTTACGCTTTCATCACCTGT	4269	4298
BSMV	BSMV17	GGTGAAACGGTTGCTACAGGAACAACCCCG	4338	4367
BSMV	BSMV18	TAAAGGTTCCGGAGGTAGGTACTATCCCAG	4697	4726
BSMV	BSMV19	ACGTTTCCTGGTAACTCCGTTAGGTTTTCT	4750	4780
BSMV	BSMV20	AAGAAGGCGTTAAAGGTATTCTTCTTTGAA	5029	5059
BSMV	BSMV21	TGTGATGTTGACTTCACGACTTATAACTTC	5167	5197
BSMV	BSMV22	ATGACTGCTGATGAACTGAATGAAACAGTT	5374	5404
BSMV	BSMV23	CTTCAAGTACCCCGCATTTTGTGGTAAATT	5817	5847
BSMV	BSMV24	AATTTCTGCTGTGCATAGATGGAAAATATC	4843	5873
BSMV	BSMV25	ATCTTTGACTTTAATAAGTTTAAGTTGCTG	6085	6115
BSMV	BSMV26	AACAAGAAATAAGAGATTGGAACTTTACAA	6306	6335
BSMV	BSMV27	AGAAGATGCAGGAGCTGAAACTTTCTCATA	6533	6562
BSMV	BSMV28	AGGTCCCGTCCTTACGCTTTCATCACCTGT	7067	7096
BSMV	BSMV29	GGTGAAACGGTTGCTACAGGAACAACCCCG	7136	7165
BSMV	BSMV30	AGGTCCCGTCCTTACGCTTTCATCACCTGT	7433	7462
BSMV	BSMV31	GGTGAAACGGTTGCTACAGGAACAACCCCG	7502	7531
BSMV	BSMV32	TAAAGGTTCCGGAGGTAGGTACTATCCCAG	7861	7890
BSMV	BSMV33	ACGTTTCCTGGTAACTCCGTTAGGTTTTCT	7914	7944
BSMV	BSMV34	AAGAAGGCGTTAAAGGTATTCTTCTTTGAA	8193	8223
BSMV	BSMV35	TGTGATGTTGACTTCACGACTTATAACTTC	8331	8361
BSMV	BSMV36	ATGACTGCTGATGAACTGAATGAAACAGTT	8538	8568
BSMV	BSMV37	CTTCAAGTACCCCGCATTTTGTGGTAAATT	9891	9011
BSMV	BSMV38	AATTTCTGCTGTGCATAGATGGAAAATATC	9007	9037
BSMV	BSMV39	ATCTTTGACTTTAATAAGTTTAAGTTGCTG	9249	9279
BSMV	BSMV40	AACAAGAAATAAGAGATTGGAACTTTACAA	9470	9499
BSMV	BSMV41	AGAAGATGCAGGAGCTGAAACTTTCTCATA	9697	9726

Appendix 2.4. Designed, curated e-probes for the detection of *Barley yellow dwarf virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
BYDV	BYDV1	TGGGCACATGGCTGACATTGAGGACTTCGA	302	332
BYDV	BYDV2	GAAAAGGGCCGCGGCTTCATCAACTGCTGG	597	627
BYDV	BYDV3	CTACAGTTTCACGGAGGTCAAGAAGGTGGA	746	776
BYDV	BYDV4	GTCCATCAAGGACACCGCCGAAGCCAGGGA	857	887
BYDV	BYDV5	AGAAGAGGTGGGATACTGTAAAACATATCC	1453	1483
BYDV	BYDV6	AACTATGTAATAACGGAGACGACTGTGTCA	2106	2135
BYDV	BYDV7	AAGGCCCGATAGCATAGGCAAAGATAGCAC	2299	2329
BYDV	BYDV8	AGATAGCACAACACTACTGAGCATGCTCAA	2320	2349
BYDV	BYDV9	CTGGAGTACCCATACTTGAAAGTTTCTATA	2409	2439
BYDV	BYDV10	CTATAAATGCCTATATAGGAGCTCGGGGTA	2434	2463
BYDV	BYDV11	ACGTACCTATAATGAAACACCTATCACAAA	2533	2563
BYDV	BYDV12	AGAGGAGGGGCAAATTTTGTATTTAGACCA	2998	3027
BYDV	BYDV13	CTCAGTTGACAACCTTAAAGCCAACTCCTC	3054	3083
BYDV	BYDV14	TCTATCGCAATGCCCAGCGCTTTCAGACGG	3108	3138
BYDV	BYDV15	TCAGTTGACCACATCGGTGGCAACGAGGAC	3811	3841
BYDV	BYDV16	AGCTTCTACCTCATGGCTCCCAAAACAATG	4060	4090
BYDV	BYDV17	CTGACAAATACAACTATGTTGTCTCATATG	4094	4124
BYDV	BYDV18	AGGCACGCTGAAACGCCCATACGTTCTAAA	4207	4237
BYDV	BYDV19	CAAGGAACCTGAAGTACTTGGGACATACCA	4500	4530
BYDV	BYDV20	GACTTGTAGAAGCGAATAGGTCCCCTACTA	4658	4688
BYDV	BYDV21	CCGATGAAATGAGGGTGGAGTGAGCGGAGT	5265	5295
BYDV	BYDV22	CTGACACTCGAAAGAGCAGTTCGGCAACCC	5647	5677
BYDV	BYDV23	TGAACTTTTGATCGGCGCTAGCGTCAAAGC	5826	5856
BYDV	BYDV24	GGGCAATTTAAGGCCCACGACGCTTTTGTC	5938	5968
BYDV	BYDV25	ATCCAGGACCTCTACGCCTTCACCAAGGTT	6412	6442
BYDV	BYDV26	CAAAATGAAGATGGTGATCCAGAGCCCTGC	6756	6785
BYDV	BYDV27	TCGGCTTGGTATACAAAGCCCCAAATGCCT	6967	6997
BYDV	BYDV28	TTGGAACTCGTTTAAAATTCAACGAGAAGA	7357	7387
BYDV	BYDV29	TGTGCTTTCTGGCTATGACAACTTCACTCA	7427	7456
BYDV	BYDV30	CACTGCTCTCGAACATCAACTTGTAAATAA	7607	7637
BYDV	BYDV31	ATATAAACACCAGCATGGGAAACAAACTGA	7702	7731
BYDV	BYDV32	CAAACTGATAATGTGTGGCATGATGCACGC	7724	7753
BYDV	BYDV33	CGTCAGAATAGATGGCAAATACAGGATGGT	7943	7973
BYDV	BYDV34	TGTCAGCTGTGGCTCAGTGTGGGCTCGTTC	8050	8080
BYDV	BYDV35	GGCTCGTTCTGAACGCGGGTGTACCCATTC	8071	8100
BYDV	BYDV36	TTCTCAGTCGACAACCTTAAAGCCAACTCT	8729	8759
BYDV	BYDV37	AACCGGCGTGATTTCAACGCGAGAAAACTC	9274	9304
BYDV	BYDV38	AAATAGTTGAAAGGGATGGCGTGATATCTT	9678	9708
BYDV	BYDV39	CAAACGGATGGAGTTCGGTACCATATCTGT	9814	9844
BYDV	BYDV40	ATTACCCACCATAATCGACCAAGGCTTGTG	9946	9976
BYDV	BYDV41	AGCTGCCACAGCTGAAATACCTGATGCTGA	10072	10101
BYDV	BYDV42	AACGAGATGAGCAACACTCGTTAAAAGTTC	11038	11068
BYDV	BYDV43	AGTCGGCCCGACCCGTAGGCATACCCTCGA	11084	11114
BYDV	BYDV44	TCGATACGAAACGAGGGTTCCTAGGAGCCA	11110	11140
BYDV	BYDV45	GGGGTAAGAGTCTTAGCAAGCTCTGTACCT	11158	11188
BYDV	BYDV46	CTGCTTGCAACAGTATCCCTTAGAAGCAAT	11286	11316
BYDV	BYDV47	AGGCTTGGACCGACTTCTTTACGAAGTCGA	11772	11801
BYDV	BYDV48	CCCAGCGCAACTCCTGGCTAACAGCTATAG	12816	12846
BYDV	BYDV49	GCACGCTATCGATAGTGTGTTCGGATCCCC	13053	13083
BYDV	BYDV50	AAGGCCCGATAGCATAGGCAAAGATAGCAC	13635	13665
BYDV	BYDV51	AGATAGCACAACACTACTGAGCATGCTCAA	13656	13685
BYDV	BYDV52	GGCCGTGGCTCAGTGCGGCTTGGTGCTCAA	13713	13743
BYDV	BYDV53	CAGAAGCTCTGGGTACAAGAAAGTGAGTGA	13785	13814
BYDV	BYDV54	GCGCAATGTTTATTGAACTCGACACTTGGT	14551	14581
BYDV	BYDV55	ATTCACCATCTCAAAATCAGCAACCAAAAC	14615	14645
BYDV	BYDV56	CATTGCCGTTAAGGCGATGAATGATCAATC	14942	14972

BYDV	BYDV57	AAGAACCGTCAAGGCATGGTGGAACTCCAA	15005	15034
BYDV	BYDV58	CCCGCTGGTGAGTACAGCGTGAATATCTCT	15075	15104
BYDV	BYDV59	TATGTTGTGTCGTATGGAGGGTACACAGAT	15417	15446
BYDV	BYDV60	CTAACGACCCTCAGTAATTGGCTGGTTCTG	16462	16491
BYDV	BYDV61	GTTGTACACTGCCCCGGAGCCCACCGGGTT	16564	16594

Appendix 2.5. Designed, curated e-probes for the detection of *Barley yellow dwarf virus*-MAV in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
BYDV-MAV	MAV1	CGAGAGGTGGCGTCTACGCGCCACTTTCCT	427	456
BYDV-MAV	MAV2	GGATATATAATGGGGTCTTTGGAGACAGCG	1857	1886
BYDV-MAV	MAV3	AGCTCGTCGACGCACAAGAGGAGGAGGGGC	2940	2970
BYDV-MAV	MAV4	CATTGCCGTTAAGGCGATGAATGATCAATC	3570	3600
BYDV-MAV	MAV5	CCCGCTGGTGAGTACAGCGTGAATATCTCT	3703	3732
BYDV-MAV	MAV6	GCTCGGGTAAACTAGTCCTTTACCGCCGTA	4856	4885

Appendix 2.6. Designed, curated e-probes for the detection of *Barley yellow dwarf virus*-PAS in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
BYDV-PAS	PAS1	TACTTCGCAGAGTCCATAATTGACAAGGTG	1433	1463
BYDV-PAS	PAS2	CGTCAGAATAGATGGCAAATACAGGATGGT	2266	2296
BYDV-PAS	PAS3	GGCTCGTTCTGAACGCGGGTGTACCCATTC	2394	2423
BYDV-PAS	PAS4	AAAGCTTCTACCGCTGTTTGTACAGAAGCT	2427	2456
BYDV-PAS	PAS5	CCCCACAACCCACTCCAACCCCACAGCCCA	3509	3539
BYDV-PAS	PAS6	AACCGGCGTGATTTCAACGCGAGAAAACTC	3597	3627
BYDV-PAS	PAS7	GGCTACTGGATTGGGCTCATCGCTTACTCC	3841	2871
BYDV-PAS	PAS8	AAGTACTTGGAACGTATCAGGGTATGAATA	4511	4540
BYDV-PAS	PAS9	AACGAGATGAGCAACACTCGTTAAAAGTTC	5361	5391
BYDV-PAS	PAS10	AGTCGGCCCGACCCGTAGGCATACCCTCGA	5407	5437
BYDV-PAS	PAS11	CTGCTTGCAACAGTATCCCTTAGAAGCAAT	5609	5639

Appendix 2.7. Designed, curated e-probes for the detection of *Barley yellow dwarf virus*-PAV in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
BYDV-PAV	PAV1	AGACCTTCCGGTCAGAGGCAATTAATGGGA	3308	3338
BYDV-PAV	PAV2	GACTTGTAGAAGCGAATAGGTCCCCTACTA	4658	4688

Appendix 2.8. Designed, curated e-probes for the detection of *Barley yellow striate virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
BYSMV	BYSMV1	AGGATCCCGGTTTACACTCCTGCAGAATTG	274	304
BYSMV	BYSMV2	ATTGAATGATAATCAACTAGTTGCTGCTGC	300	329
BYSMV	BYSMV3	GAACGTGACTGTGGCTGACAGTGAGAAGGA	510	539
BYSMV	BYSMV4	GGATCACGCCAACAGGAGAGGCATGGAAAT	1795	1825
BYSMV	BYSMV5	AAGCCAAGATTGCAAAGGATGTGGACCAGA	2028	2057
BYSMV	BYSMV6	CAACGGACAAGAAGACTACATTTGCAATCT	2787	2817
BYSMV	BYSMV7	TCCAGCTACTAGCCGCACTCACTTCTCTTT	3244	3274
BYSMV	BYSMV8	GTAGACTTTGGAGAAGGAGTGATCCCGATC	3966	3996
BYSMV	BYSMV9	CTCAGGAGGTGAACAACGAGACCAGGAGTA	4258	4288
BYSMV	BYSMV10	ATGGACGAGGTGTTATTCAGAGAGTGCCTG	5495	5525
BYSMV	BYSMV11	TGGGGTTACATATCTTCGTTATAGTATCTT	6012	6042
BYSMV	BYSMV12	CAGCGATTGTCCCAATCGAGATCAACTCTC	6239	6269
BYSMV	BYSMV13	ATGTGATAACACAGGTATTTGACTTATTCG	6848	6878
BYSMV	BYSMV14	GAAGCCACCTACTACACCTTCTTGGAGATA	7900	7930
BYSMV	BYSMV15	ACCGCTTGATGGGTCAAGGTGACAATCAGG	8174	8203
BYSMV	BYSMV16	GAGATCCTTGCAAGACACTTTCAGCGAGGT	8301	8331
BYSMV	BYSMV17	GGGACTGAGCCTTTAATAACAAGACCATTG	9601	9631
BYSMV	BYSMV18	GAATCTCCTCCGAGCAGTTACTGATCTTGA	9693	9723
BYSMV	BYSMV19	ATTGCAGAAATTCTATCTGATAAGGAGATT	10036	10065
BYSMV	BYSMV20	AGAGTAGCATCTGGAGACATATATCCCTCG	10387	10417
BYSMV	BYSMV21	AATAACTCATGAGTTGGCGAACCAGTCAAC	10971	11001
BYSMV	BYSMV22	AGGTGTCATATCAGATGCCCGGAGAAGAAT	11018	11048
BYSMV	BYSMV23	CAGACCCAAAGATGACTACTCTGACAAGTA	11055	11085
BYSMV	BYSMV24	GTCTGTACGATCCCTATATGGATCCTGGAA	11571	11601
BYSMV	BYSMV25	TTCCAGGATGCGTCCATAATTCATTGGAGA	11848	11878
BYSMV	BYSMV26	ATTAGATTCTGTCCTCCTGGGAAGAATACC	12262	12292
BYSMV	BYSMV27	AACTCCCTATTGACCATAGTATAAATAATA	12333	12363
BYSMV	BYSMV28	TATAGATAAGGAAAACCGTCTATGTGATGA	12410	12439
BYSMV	BYSMV29	ATTAACTAGTAGGATATTGGGACACACCAA	12644	12673
BYSMV	BYSMV30	GATGGTACAGATTGCGGCTCACTTGGTCGT	12676	12706

Appendix 2.9. Designed, curated e-probes for the detection of *Cocksfoot mottle virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
CoMV	CoMV1	CGTTCCCCTCCCGAATCAGAGGTTGAGAAG	31	60
CoMV	CoMV2	CAGGCGATTCCACAATTGTCAAGTCCATTC	130	160
CoMV	CoMV3	TGATTCCATCCCTGCCCAAAGGCTTCAAAG	301	330
CoMV	CoMV4	TGTTGGAAACTGCAAGTCCGTGATGTTGAT	442	471
CoMV	CoMV5	TCTGCTGATTCGGACTTGGACTTTGTCCTT	1034	1064
CoMV	CoMV6	TTTGTCCTTGTGTCCGTGCCGAGGAACGCT	1055	1084
CoMV	CoMV7	GAAAATCCCTTCCGCATAGTCACGAAAGTG	1205	1235
CoMV	CoMV8	AACCTCCGACCGAGAGTGGACAACAGCTGA	1510	1539
CoMV	CoMV9	CCTTTAAACTACCAGCGGGCGGGCTCCCTA	1631	1660
CoMV	CoMV10	CTTGGTGTGTGAGGCTGTGGTGGAAAGGCT	2151	2180
CoMV	CoMV11	GAAAAGAGGATTGTGCGATCCAGTGAGACT	2238	2268
CoMV	CoMV12	TCCAGTGCTCCCCAACGGAGATTTGAAGGA	2880	2909
CoMV	CoMV13	TCTCGCCGCTGGGGCCAGATAGTCAGATAC	3034	3063
CoMV	CoMV14	GTTTAAGTTATCTTGGAGCACTCCTGAGGA	3673	3703

Appendix 2.1.1. Designed, curated e-probes for the detection of *Cereal yellow dwarf virus*-RPV in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
RPV	RPV1	CCTGGTTAGATTCCCGCGCGCTTTGCTTTA	489	519
RPV	RPV2	TCTGCAGGGTTCTTACCTGCAGAAAATTCT	798	828
RPV	RPV3	ATCCCAACCCACACCCACAACCTCGACGAC	1717	1746
RPV	RPV4	TCGATAATAGATCAGGTCGCTGCACAAGCT	1880	1910
RPV	RPV5	CCAGTACTATCCCCGTGCCCGATCAACAAC	2049	2079
RPV	RPV6	ATGACTTTCGGATTGACTTCCTCGAAGCCA	2333	2362
RPV	RPV7	CTTGATGAGGGCCGCTACCGCCTCATCATG	2605	2635
RPV	RPV8	ACAAATCCGAGATTGCATTGTGGAGTGCAA	2687	2717
RPV	RPV9	GACGATATGGAGGTAAGAAATCGCCTTACC	6893	2922
RPV	RPV10	ATGATGCCCTCGAAGCGCCAGATACAGATT	3134	3164
RPV	RPV11	GGGTACAATCCGGAATGTGGAAACGCAGAG	3295	3325
RPV	RPV12	GTCAATTACCTCAATGCAGCCAGCTCAGTG	3331	3361
RPV	RPV13	CCAACGGCCCAGCCAGGCGCGGAAGACGCC	3740	3770
RPV	RPV14	TACTCTGGGGTGGCTCACACGGTCATCACT	4348	4378
RPV	RPV15	TTTAACCAAAACCAAGTTTTGGAGAAAGAT	4750	4779
RPV	RPV16	CTTCAAAGTACAATTACGCCGTCTCATATG	4856	4886
RPV	RPV17	TGCTCAAACGGGCGAAGGAAAATGCCTTAA	5497	5526

Appendix 2.1.2. Designed, curated e-probes for the detection of *Digitaria striate mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
DiSMV	DiSMV1	GCAGTAGTACCTTCAATTCGGGCTTTGGGC	196	226
DiSMV	DiSMV2	GAGGTTGAGTCGTCATGCCTGCTTCGTCGA	337	367
DiSMV	DiSMV3	CCCGTTCCAGCGCTAGCCAGGATGCACTGC	436	466
DiSMV	DiSMV4	ACGGGAGTACCTTCACGGCGGATTACACGA	832	861
DiSMV	DiSMV5	CGAAGCTGCCGGTGCCCAATACAGACAACC	859	888
DiSMV	DiSMV6	GAAAATGCGCGGCGCGTGCCGACGGAGGAG	1247	1277
DiSMV	DiSMV7	TACATTGTATTCAGCGTCGGGAATGATGTT	1538	1567
DiSMV	DiSMV8	AGTAGTTATGCTTGCCCAGACTTCTGGCCC	1591	1620
DiSMV	DiSMV9	TTCGGCAGCGCCCGGAGCGCGTTCGGCCAG	2374	2404
DiSMV	DiSMV10	GGTGCTTTCATGCCGGGCCCTATCAGCCCA	2688	2717
DiSMV	DiSMV11	GTGGAGGCTTACGACATACGGCAGTCACTT	3152	3182
DiSMV	DiSMV12	CAAAACGCCGTCGCGTCTATAAACAAGCAG	3375	3405
DiSMV	DiSMV13	CTTACAAGTTATGCCCGCGGTTCAGCTGAG	3500	3530
DiSMV	DiSMV14	TTAGTGGCACTATGCAACAATACTGTGTGT	3591	3621
DiSMV	DiSMV15	AATCAGTGATTACCGTTCATGGTTTATGAA	4034	4064
DiSMV	DiSMV16	TAATATGCCGCCCTAGCGGAGAAGCTATGC	4211	4241
DiSMV	DiSMV17	GCGGACCTCGTTCCTGGTTACATCAGACAT	4639	4669
DiSMV	DiSMV18	ATTGGGCACCTTGGGTATGTCAGGAAGACA	5279	5309
DiSMV	DiSMV19	AGCGAGGATCGAACATGACTGAGTCACTGA	5379	5409

Appendix 2.1.3. Designed, curated e-probes for the detection of *Maize dwarf mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
MDMV	MDMV1	AAAGGGCCAGGTCTATGACGAGAAGCGAGC	231	261
MDMV	MDMV2	AACCTCAGTTAATAAACTAGTGAGAAAGAC	444	474
MDMV	MDMV3	TCTCGAGTTATCGAGATGGTATAAGAATAG	1320	1350
MDMV	MDMV4	CTCGATTCAAATGGTAACTTTGTTTGGGGA	1444	1474
MDMV	MDMV5	TAACTACAGAGTTTGGACAACCTGCATACT	1727	1757
MDMV	MDMV6	GAGGATTCAGCTAAGGATTACACAAAATTC	1915	1945
MDMV	MDMV7	GCAGTACGAGTCGATGGAAAGCGAGATGAG	2172	2201
MDMV	MDMV8	TGGTACAATAACACACAAATCATTTTCCAC	2217	2246
MDMV	MDMV9	GAAGCGTAGATTTAGGCGCTGTCTACAATA	2801	2831
MDMV	MDMV10	GCAGAAAAGTCGCAGTCGAGTCAGCTCTAC	2862	2892
MDMV	MDMV11	TATTATCAACAGTAGCTAATACTATTATAG	3017	3046
MDMV	MDMV12	ATAACGAACTCACTTGCGAACAGTTCCTGC	3155	3185
MDMV	MDMV13	GGCTATTGGAACACTAATCACAATGATATT	3297	3327
MDMV	MDMV14	CTCGGAGGAGCAGTCGGTTCGGGAAAATCA	3715	3744
MDMV	MDMV15	TCGTGCTTGAACTTGGGACTGGATCAAAAG	4163	4193
MDMV	MDMV16	AATACGGTAACAACATTCTGGTTTACGTGG	4205	4234
MDMV	MDMV17	AGGTTATAACACATAACGTTTCAACTACAC	4652	4681
MDMV	MDMV18	CCATTCATTATGTCTGAATTAGTCAAATTT	4738	4767
MDMV	MDMV19	TGGTTAACAGTTAAAGACTACAACAAAATT	4876	4905
MDMV	MDMV20	TTTACCACGAACAATCGCTATTGTGAATCA	5088	5118
MDMV	MDMV21	GCAATATGTTAGCCTCAAGGTACATGAAGG	5210	5239
MDMV	MDMV22	ACAGTCATTCACCAAGGCTTAGATTCCACC	5359	5389
MDMV	MDMV23	TAATTTTGATCCACAAGATTACAATCTAAT	5721	5750
MDMV	MDMV24	TTTATTCAAATCCAGGAATTAAAGCATATT	5882	5912
MDMV	MDMV25	ACTATACTTATCCACGCACGAGGAAGCTGA	6945	6975
MDMV	MDMV26	GTCATACAACTCCTACGCAATGTTGGGATT	7117	7147
MDMV	MDMV27	AAGAGGATAGAGCAGAGATAATAAAGCAAT	7262	7292
MDMV	MDMV28	ACAGTTGTTGGCAACACACTAATGGTGATC	7771	7801
MDMV	MDMV29	ATTACTTGCAGTGAACCCAACACATGTTGA	7890	7920
MDMV	MDMV30	GAAAACACTTCGCAGCGTTAGGGTTAAATT	7937	7967
MDMV	MDMV31	ATCAAGTACGAAGAAATGTATATCCCAAAA	8026	8056
MDMV	MDMV32	TTCAAACACACGGATGTTTGGTCTTGATGG	9138	9168
MDMV	MDMV33	ACCACGAGGATGCAGCGAGTTTCGTGGTGA	9464	9494

Appendix 2.1.4. Designed, curated e-probes for the detection of *Maize streak virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
MSV	MSV1	CAGCTGCCCTGGAGTCATTTCCTTCATCCA	1460	1490
MSV	MSV2	GCGAGTATTATTGTAGGCTTAGACTTCTTC	1506	1536
MSV	MSV3	AGTAATTATGAACCCCTAGGCTTCTGGCCC	1693	1723
MSV	MSV4	ATAATGTCTCGCATTATTTCATCTTTAGAA	2096	2126
MSV	MSV5	CAAAGAACCTTGAGTCAGATATCCTTACCG	2289	2319
MSV	MSV6	AGTGGTTGTAAATGGGCCGGACCGGGCCGG	2627	2657

Appendix 2.1.5. Designed, curated e-probes for the detection of *Northern cereal mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
NCMV	NCMV1	AACCACAAAGTGCGCTGCGGGATATCAATA	341	371
NCMV	NCMV2	AGCTGGTAAGTCAGAGGCGGAGAAGAAGAA	524	554
NCMV	NCMV3	GGTGAGATCTCAAGATGATGAGACTATCAT	605	635
NCMV	NCMV4	CAATCTAATTTGGGGATGCAGTGGCTCCTA	978	1008
NCMV	NCMV5	CAAGAATCATGAGTCCCGCGTACTTCCAGG	1117	1147
NCMV	NCMV6	CACAACGTAATATGGGGATGTACTCAGACA	1321	1351
NCMV	NCMV7	GGATGAAGTGTTCGGAATGTAACCTGCTCC	1415	1445
NCMV	NCMV8	GATTGGGAAATTCAATGGGGTCCTCAATGA	2079	2109
NCMV	NCMV9	GATCAAACGACTGAAACGAGCAGGGAGAAG	2167	2197
NCMV	NCMV10	TGGTCCCTGTGGAACGTCTAGAGAAGTTAT	3341	3371
NCMV	NCMV11	AAGAACCTCTTTAACGAGAGGAGACTTTAG	1359	1389
NCMV	NCMV12	TGTTTTCTGATGACACTTCAGGGATAACTA	3608	3638
NCMV	NCMV13	CCGCCGTTCTCGTTGGTTCCGAGGTCTCAG	2818	2848
NCMV	NCMV14	GCAAAGCTCACCCTTAGCTGCGGGTACAAA	4045	4075
NCMV	NCMV15	CGCTACAAATATATATTCCCCGAATATCTT	4594	4624
NCMV	NCMV16	CCGGGTATCCAGATCCAAGATGCAGAAATT	4976	5006
NCMV	NCMV17	GTCTTGTCTGGATAATCACCCTTTCCGTAT	5008	5038
NCMV	NCMV18	ATCTTGTTCTGATACTATATTAGGAGAGCA	5117	5147
NCMV	NCMV19	AGTTGTCGGCAGATGTATCTGGAGGATCAT	5189	5219
NCMV	NCMV20	TAAGTGAGGATCTCGAGATAGAGCCTGAGT	5350	5380
NCMV	NCMV21	CAACTATACTAAGGTTGATACATCATGTCC	5663	5693
NCMV	NCMV22	CGTAGACGGGTTCTATTACGATCCACTCCC	5891	5921
NCMV	NCMV23	GACTTCTACTTCACGTTTTGATTGTTGTGT	6498	6528
NCMV	NCMV24	AGAATCCTATCTGCGTTCCCTAATATTATT	6880	6910
NCMV	NCMV25	TGATGGGGTTGTTTCGGCTATGGGGTCATC	7670	7700
NCMV	NCMV26	CTGGGGCTGTAATCAATACAAAAGATCCTA	7895	7925
NCMV	NCMV27	TTGTCTGATTGGGATTTGGTTGAATCAAAG	7933	7963
NCMV	NCMV28	TGATGGACTATTAAGGGGAATCGATAAGAA	8133	8163
NCMV	NCMV29	AAGAGAGATGAACCCTGTTGCGAGGATGTT	8214	8244
NCMV	NCMV30	GATTAGGGCAACTCGATCTCAAGGGAAACA	8370	8400
NCMV	NCMV31	TGGGAGATTATTCGGGTTACCAACTTTGTA	8496	8526
NCMV	NCMV32	GATTACCCTTAAAACCATTAGAAACATGGG	8900	8930
NCMV	NCMV33	GAGGATATCTAGGATATTTTACTTCTCAAA	9000	9030
NCMV	NCMV34	CTTAAGAATTGGTATAGGGTTGTTCTATCT	9451	9481
NCMV	NCMV35	GGCGACTCTGTATGGATACTCTGAATCAAT	9729	9759
NCMV	NCMV36	ATCAGGGTGAGGAGGTAGTGAGTCTTTGTC	9884	9914
NCMV	NCMV37	CCATTGATCACGCGTCCTATAAAACTTCTA	10177	10207
NCMV	NCMV38	TGGCACTCGCCCACGGATCATTATGGATGC	10355	10385
NCMV	NCMV39	GAGAGTTTGCCTATTATCCACAGAAGCTAT	10714	10744
NCMV	NCMV40	GGAGATCAAGATAAACTCATCCTGGATAGT	11196	11226
NCMV	NCMV41	TAGTGTCACCACTTGTCAAAGTAGACCCCG	11222	11252
NCMV	NCMV42	GATCTTTCTCGAATCAGATGCAATCAGGGT	11365	11395
NCMV	NCMV43	TGGGTTCGGATATTCCTCAGTTATGTGCAA	11751	11781
NCMV	NCMV44	TCTACAATAGACCTTCTGTCAGATGTATCC	11908	11938
NCMV	NCMV45	TGGGCACCAGCTCTGTCTACAGAAGGGCAC	12697	12727

Appendix 2.1.6. Designed, curated e-probes for the detection of *Paspalum striate mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
PSMV	PSMV1	CAGCGTCTGAGGTTTCAATTCCCGTAAGGG	259	289
PSMV	PSMV2	GTGGTTGGTGTTTAGTTTCCATGGCAGCTT	321	351
PSMV	PSMV3	AAGGGATCCACATACGGTGGAGGAGTGGAA	1825	1855
PSMV	PSMV4	TCCTTACCATCGAGAGATAGTCCTCCTTGT	1922	1952
PSMV	PSMV5	ATTCCAGGAGGAATACAGCCGTGAGGTTGC	2282	2312
PSMV	PSMV6	GAGGTTGCACTTACTGTATGTGAGGAACAG	2304	2334

Appendix 2.1.7. Designed, curated e-probes for the detection of *Soil-borne wheat mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
SBWMV	SBWMV1	GTCTTCACCACTATGCCGATCGACAGCAGT	89	119
SBWMV	SBWMV2	TTTGCCGACGATGGACGGTTTCTTTGAAGT	892	922
SBWMV	SBWMV3	TTGTCCAAAGGGCGATAGAGTACACAGGCA	1251	1281
SBWMV	SBWMV4	GCGCCCGTTTATGTTGATGCGGAACAAGCT	2051	2081
SBWMV	SBWMV5	TGCTGAGGACGAAAGCGACGATAGTTACTA	2614	2644
SBWMV	SBWMV6	AGACAAGGTACAAAAGAGTGCTTTACCAAA	2665	2695
SBWMV	SBWMV7	AAGATCGTTTCTGACAGGACTACATTGAGG	2768	2798
SBWMV	SBWMV8	AAATGAACCAGAAGCCTGTTAACACAAGGG	3024	3054
SBWMV	SBWMV9	GTTGGTCATTTATCGTAATTTGCAACAGGT	3100	3130
SBWMV	SBWMV10	ACAACAAATACCTTTCATAAATCGTGTTGA	3481	3511
SBWMV	SBWMV11	GCCCAAGTTGCAAGAGGCAGTCAATGAATT	4390	4420
SBWMV	SBWMV12	TCAAGTTCGCACCTGATCCTATGAAGTTGA	5286	5316
SBWMV	SBWMV13	ACAGGACCACTGAAGATCCGACTGCTGCAT	5451	5481
SBWMV	SBWMV14	TATGGGCTAGTATGTTTAATTATAAGGGTG	5517	5547
SBWMV	SBWMV15	TTGTAATTTGTGCAGACATGGGCTCACAGG	5635	5665
SBWMV	SBWMV16	GGACTAAGACTAATAAGGAGGAAGGTACTC	5848	5878
SBWMV	SBWMV17	GCTGGGAAAAGGTCCGCATGTGATGTGTTT	6023	6053
SBWMV	SBWMV18	TCATGCATTACAGTATTCCGTTAAATGATA	6055	6085
SBWMV	SBWMV19	AGGAGGTTCGTTGGGAGGTCTCTCTCCTTA	6282	6312
SBWMV	SBWMV20	AAAGGTTACACTGGTTACAACAAAGAGCTT	7444	7474
SBWMV	SBWMV21	AGCTGCACCTGGCACTAGTCAAGTAGAGAA	7776	7806
SBWMV	SBWMV22	AAGAATTTCAGAATCGACTCATAATTGCTG	8003	8033
SBWMV	SBWMV23	GAACCAGAGGTAGTTTATTCGGAGCTAGAA	8086	8116
SBWMV	SBWMV24	GCGCTATGGCAATTTAGGCGTCGTGCGAAA	8460	8490
SBWMV	SBWMV25	ACGTGGTTTTAAGTGAAGGCATGTTGTCTT	8786	8816
SBWMV	SBWMV26	GCTCTTTGATAGAGCCGGTACATTCGAAGA	9255	9285
SBWMV	SBWMV27	CCACACTTGTGCGAGTTGCGTTGATGGACC	9780	9810
SBWMV	SBWMV28	ATTGTGGTATGCCTGCGGCTTTTGTTTTAG	9908	9938
SBWMV	SBWMV29	CGGAGGTTCACGATGTTACTGGAAAGAAGC	10166	10196
SBWMV	SBWMV30	CAGGTGTGCATTACGGGTTTAGTTGATGTA	10195	10225
SBWMV	SBWMV31	CAGTACGTTTAAACTGTAGGTTCCAACCTC	10529	10559

Appendix 2.1.8. Designed, curated e-probes for the detection of *Rice stripe virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
RSV	RSV1	AAGACTATTTATGGTGCTGAGTTTGAGAGT	437	467
RSV	RSV2	TTTGAGAGTACTATTAGAATTAGGAACTAT	458	488
RSV	RSV3	GGACACAAGAGCTTTAGTAATGTGGATGAA	647	677
RSV	RSV4	TTGTTGATAATCATATAGTGATAAAATTGA	1539	1569
RSV	RSV5	GTGATTCGATAAAGCAAGATTTTGTATATC	1596	1626
RSV	RSV6	CTGAGACTGACAAGAGCTTTAAGGGTGCTG	1686	1716
RSV	RSV7	ATCTCTACTAAGATCAGTAGAGGTTAAAAG	1840	1870
RSV	RSV8	TAGAAGGACAAGTGGCTAAGAAAGTGGAGG	2601	2631
RSV	RSV9	TTGGAAGAACCAGATGTGAAATTTTATCGA	2735	2765
RSV	RSV10	TAAATGTAGAGGATCTAGTGTTATCTGCTC	2877	2907
RSV	RSV11	AAGAAGACACCTATGTGGTTAAGAAACTCA	3045	3075
RSV	RSV12	TTTCGTCTCTGCTAATGTGAGCAAGTTAGT	3223	3253
RSV	RSV13	GTGGCCCCTAGCATTGAGGAACAAGATGGA	3317	3347
RSV	RSV14	GGCGGAGCCAGACTCAGTATGTCAGATGAT	3349	3379
RSV	RSV15	TTAGCACTCAGTTTATTGAGACGATGGCAT	3984	4014
RSV	RSV16	AAAATAGTGGTAGATCTCTTACCGAAGGCT	4223	4253
RSV	RSV17	TGAATGTATGCACATTTGTATTTTCAAGAA	4273	4303
RSV	RSV18	GTTTAGAATACCTGAATTGCACAACATGGA	4438	4468
RSV	RSV19	AATGAGTATATGACAATATCTACTAGTGAT	4487	4517
RSV	RSV20	GACTGAGACAGGTATGATGCAGGGAATTCT	4783	4813
RSV	RSV21	ATGAAGCAGCACAATTGTACCTGCTCTGTG	5001	5031
RSV	RSV22	TCATCAGTGTGGCCAGCATATGAAACTCTT	5351	5381
RSV	RSV23	TAATCCTAAATGTGCTGGCTTGTTGGGATT	5425	5455
RSV	RSV24	CATGACTATCTGCAGGAACCGCATACATTG	5945	5975
RSV	RSV25	AAGAATCAAAATACACACTGGGAGAGAAGA	6055	6085
RSV	RSV26	ACTTATCCTTGAAAGAAACAATTGACCAAG	6426	6456
RSV	RSV27	TAGAACTATCTGGAAAGTCTACAAAGAACA	6616	6646
RSV	RSV28	ACTACTTGTCAATCTGCACCACTCTGTCTC	6900	6930
RSV	RSV29	GGCCTGACCAAGACACAAGTCAGAAACAGC	7083	7113
RSV	RSV30	CTAACAGAGAAGGTGTTTCACGATTATCTG	7157	7187
RSV	RSV31	TACCATGTAGACCCTAAAGCAAACTGGATT	7409	7439
RSV	RSV32	GTCAACTAGAAGCCAAACTTATATCAACTG	7945	7975
RSV	RSV33	CCAGGTTAGAAGAAATAAGGCTGCAAGGGA	8239	8269
RSV	RSV34	AGTTAGAAAAGGGCACCTTAAAACAGATCA	8475	8505
RSV	RSV35	AAATTATGTTAGGTACATGATTCAAGAGAT	8617	8647
RSV	RSV36	AAATGATGCCTATTTTGTTAAACAGGAAGA	8764	8794
RSV	RSV37	ATCTGGCTCATGGGATGCTGTGAGGAGTTC	9181	9211
RSV	RSV38	CTACACACTTAAGCTTAAGAGAAGATTACT	9881	9911
RSV	RSV39	AGTTGTCTTCTTAAATTTCCTCTCTATACC	10353	10383
RSV	RSV40	CTAAATCATTTGCGGTCCTTAAATAAGCTT	10700	10730
RSV	RSV41	CTAGTAATGCCTCCTAGTTCCTTCAACACA	10816	10846
RSV	RSV42	TTCCATCAGTTCTGCAACCTTCCCCTTCAA	11045	11075
RSV	RSV43	TTTATCAGAAACATTAGTGTTAGGTGCAAT	11400	11430
RSV	RSV44	TCTACAACTAATAGGCCCTCCTCTGCTTCA	11437	11467
RSV	RSV45	TCCCAAGAAACCAGTTTTGAGATTTGAGTA	12000	12030
RSV	RSV46	TAGACGTGACCACCGAACTTAGAGAAAGAA	12142	12172
RSV	RSV47	TTTTCAATGAAATATCCATTCAGCTCATAA	12259	12289
RSV	RSV48	TGGTGTAGATGAAATATGATTTAAAATGCA	12422	12452
RSV	RSV49	CAATATGCCAATCAGAGTTGCAGCATCGTA	14730	14760
RSV	RSV50	AATAGGATGTCTAAGGACCAAATTAAAATG	15422	15452
RSV	RSV51	CTTCTTGTATGAAAATACGAGTATACTATA	15646	15676
RSV	RSV52	CATCCGGATGTGGTGCGTAGCACCATTTTC	16014	16044
RSV	RSV53	CGTAAAATTACTAAAGCTGGACAAATGGTT	16382	16412
RSV	RSV54	TAGCCTTATCTTCTAGTGGTTCAAATTCAA	16414	16444
RSV	RSV55	CTGAGAAGCTGAAGGGTTCCAATCCAAGCA	16900	16930

Appendix 2.1.9. Designed, curated e-probes for the detection of *Triticum mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
TriMV	TriMV1	TTAGTAAGACCTAATGTTCGTTTGTGATAC	510	540
TriMV	TriMV2	TCTCCTGACCATTCACGATGTCGTCGAAGA	721	751
TriMV	TriMV3	ACCAGTTAGCAACCCGAGTGATTGTCGTGA	854	884
TriMV	TriMV4	CCTCTGAGAATTGAGAAACAACTGCAATTT	1284	1314
TriMV	TriMV5	ATAAGGGAAACGATTGATCTTTGTGACGAG	1497	1527
TriMV	TriMV6	TTGTCGAACAAGGGGACATGTACAATGAAG	1624	1654
TriMV	TriMV7	CTCGAACTGATTATGAGTGGGAAGAACTTG	1963	1993
TriMV	TriMV8	AAGAACTTGCTATGTGGGTTAATGCCGTTT	1984	2014
TriMV	TriMV9	ATTCACGCATGGATGAGTCTTTTGAATTTG	2527	2557
TriMV	TriMV10	GCTAGTATAGCTCCGATGGCGTTTAGAACG	2583	2613
TriMV	TriMV11	AGTGGTCGGTTTCAAACAGATATTGATTAT	2673	2703
TriMV	TriMV12	GCTGTTGCGTTAGCTCAGCTAGTAGCAATC	3468	3498
TriMV	TriMV13	TTTGAAGTGAGTAAACAGTTGCAAAACGTG	3549	3579
TriMV	TriMV14	AGGGAGCGGTTTAAAGAGTACAGTCAAGTC	3821	3851
TriMV	TriMV15	TTACTCTTGCCTTATTGTGCGCAATTAAAT	4075	4105
TriMV	TriMV16	TGTTTTAATTCTGGTCTCGCCAATGATATT	4245	4275
TriMV	TriMV17	TAAAACAGCACACTCAGACTCATTCGTTTT	4316	4346
TriMV	TriMV18	AATATAGCAACAGGGGCAGGCAATGAATTT	4569	4599
TriMV	TriMV19	GGCGTGTGATTATGTTTGTGCCATCTCGAA	5116	5146
TriMV	TriMV20	GAAGGGAGTGTGAGTTAGCGCGATCCAGTC	5143	5173
TriMV	TriMV21	ATCGAGCAGCGGCGACACAAGCAACAAAGC	5209	5239
TriMV	TriMV22	GGTGTAAGGAGGCGCAACATTAATCCAGGT	5388	5418
TriMV	TriMV23	CGAAATAAACCAGGTAAGTTTATACAAGTT	5448	5478
TriMV	TriMV24	ACATTATGTAAGGAGAGATGGAAGAATGCT	5660	5690
TriMV	TriMV25	AGCTTAAAGGGCTTTTATTGCAAACAAGCG	5707	5737
TriMV	TriMV26	GTAAGGTACACACGCGTGAAACTGGCTTGA	6445	6475
TriMV	TriMV27	GCGAAGCATATAAGAGAAGGTTTATTGGTC	6559	6589
TriMV	TriMV28	GTTGTTGTTGAGTGCCCTGATTTTGACCTT	6702	6732
TriMV	TriMV29	TGTGGAAGACGGCTGTGTTGTGGGTTTTCA	7484	7514
TriMV	TriMV30	AATGACGATTTGATTTCGTGGAAAGGCGTC	7629	7659
TriMV	TriMV31	GTCTTGCAGTCGTTTAATACACGACATGTT	7785	7815
TriMV	TriMV32	AGCAACTGGTAAGTTAGGATTATGGAAAGC	8195	8225
TriMV	TriMV33	ACTAGCACGTCGATTTAATCCAGATTGGAA	8417	8447
TriMV	TriMV34	TCAAGGAATTAGGTCTAACTTACGAGTTTG	8821	8851
TriMV	TriMV35	CCAAAACTTAAGCCTGAGAGAATAGTGAGT	8925	8955
TriMV	TriMV36	ACCTAACCAAACAACTGATCCTGTGAACAA	9272	9302
TriMV	TriMV37	AATTTGGGCGTTGGGCTAATGCGGCAGCAA	9481	9511
TriMV	TriMV38	AACCATATATGAAATGTCTCCAATGTTTGA	9686	9716
TriMV	TriMV39	CTGAGCTATAGAGCGGAGTACCTCATGTTT	10138	10168
TriMV	TriMV40	CTCATGTTTATCATTGCCACGTAGTGGAGT	10159	10189

Appendix 2.2.0. Designed, curated e-probes for the detection of *Wheat dwarf virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
WDV	WDV1	AGGAAGTGCGCGGTAGCCCATCTCGATGGA	145	175
WDV	WDV2	ATGATGCAGAGCCGAAACAGGCAATGCCAG	775	805
WDV	WDV3	CGGAGGCGAACGAGTAGTTGATGAAAGACT	1395	1425
WDV	WDV4	CAAACGGAAAACGATTGCAAACCATGCTGA	2027	2057
WDV	WDV5	TTTCATATCCGCATCACGGTCTTTACGACC	2094	2124
WDV	WDV6	TGTGCCAGAAAACTCTATGCTCTACCCTGC	2555	2585

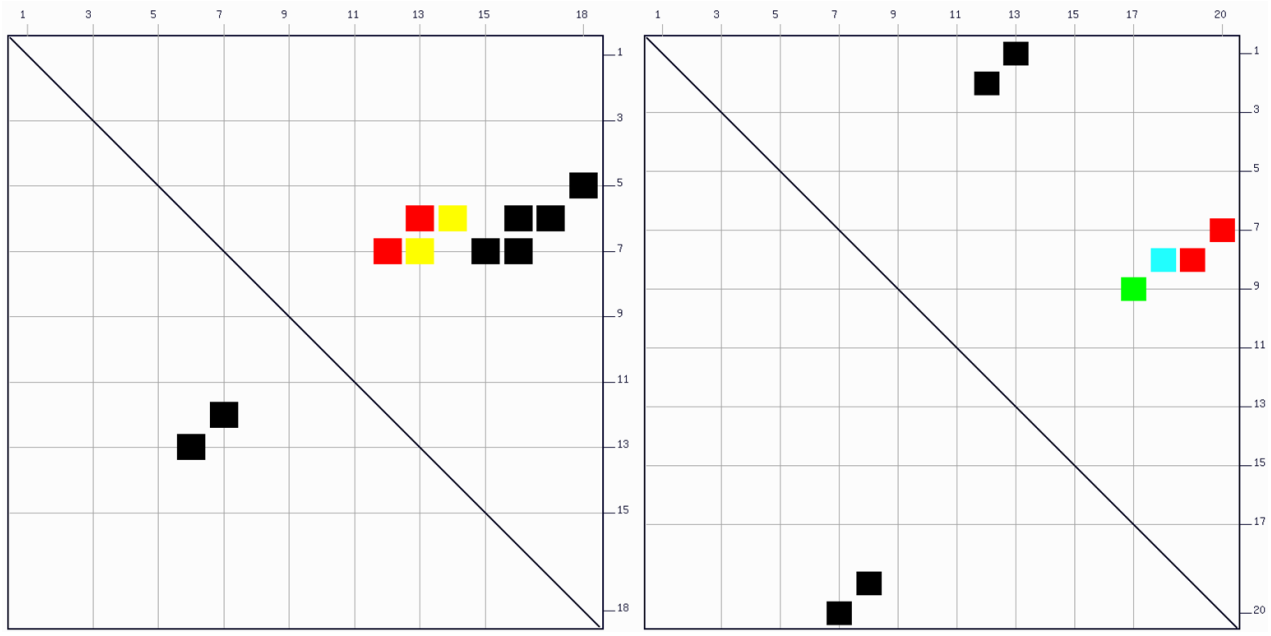
Appendix 2.2.1. Designed, curated e-probes for the detection of *Wheat streak mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
WSMV	WSMV1	ATGTGGTGGTTAACCACCACGGACCAGGAA	248	277
WSMV	WSMV2	GCGACAAACTCGTGGAGAAGAACTGCGCGT	443	473
WSMV	WSMV3	ACAACAACTGCGCAAGTTACCTGGAAGCCT	1235	1264
WSMV	WSMV4	GAACAGCCTCAGAGCATTTGCAATTGTTAC	1352	1382
WSMV	WSMV5	GCATTCTTGCAAGCACTAGGGCCAGATGAA	1441	1470
WSMV	WSMV6	CGAGGATCTGGGTCCTTGGCCAATGTTTGG	2088	2118
WSMV	WSMV7	GGTGATGTACTAAGACAACTTGACTGGATG	2116	2146
WSMV	WSMV8	ACGTGCCAACACCATATGGTATAAAGCAAT	2216	2246
WSMV	WSMV9	ATCAGAGTCAACACAGTGCTAGAGTTGATA	2260	2290
WSMV	WSMV10	CATGCGTCAAGAGCAGAAAAGAATTTGTGC	2369	2399
WSMV	WSMV11	AGCTTTACTAAACTTGCGCCAAATAGCTTT	2532	2562
WSMV	WSMV12	CGAATTTGGTGAAGTTCAGAACCATATTTG	3266	3296
WSMV	WSMV13	CAAGTAGCAGTTCCACCAGCGATATTGACA	3409	3438
WSMV	WSMV14	CAAGTGCAAACGCACAAGGCAAAGGAGTTT	3553	3582
WSMV	WSMV15	ACCCGCGTGCTAGTCACAAATTTGCAGGAT	3670	3700
WSMV	WSMV16	ATTCAATGCTAGCAACGAGAAATCTAAGCA	3698	3727
WSMV	WSMV17	AATATAACAGTAACTACGTATGGGTATGCA	3769	3799
WSMV	WSMV18	GCACAACGTGGGAGGGCAAGCTCATCAAGT	3911	3941
WSMV	WSMV19	AAACGTGGCCGGTTATGGATCACAGAACAT	4001	4031
WSMV	WSMV20	GAAGGAGACACGCATGTTGACACAGTTTGG	6141	6170
WSMV	WSMV21	ACGTACAATCTTGGAAAACTTACCAACAAG	6172	6202
WSMV	WSMV22	ACGGAGCAACAAGCGACGAGATCAATGAAT	7109	7139
WSMV	WSMV23	TTCTTTGAACTGAGTGCAGCGAAGCTACTC	7138	7167
WSMV	WSMV24	ATGATTTGCTTATTAATGCCGACACAAAGG	7739	7769
WSMV	WSMV25	CGAAGGCTATTCAGAGCGCTATTATTGCAG	7967	7996
WSMV	WSMV26	AGCATACGTGGAAGCTTTCGGTTATGATGA	7995	8025
WSMV	WSMV27	AGGGGCTTCGACAAGGCCGGTGTGCTAAGC	8977	9007
WSMV	WSMV28	AGTAAACAGGCTACGCGGAATACAGAACAA	9093	9122
WSMV	WSMV29	TTGTTACTAGGTGTGTACTTCTCCACGAGA	9310	9340

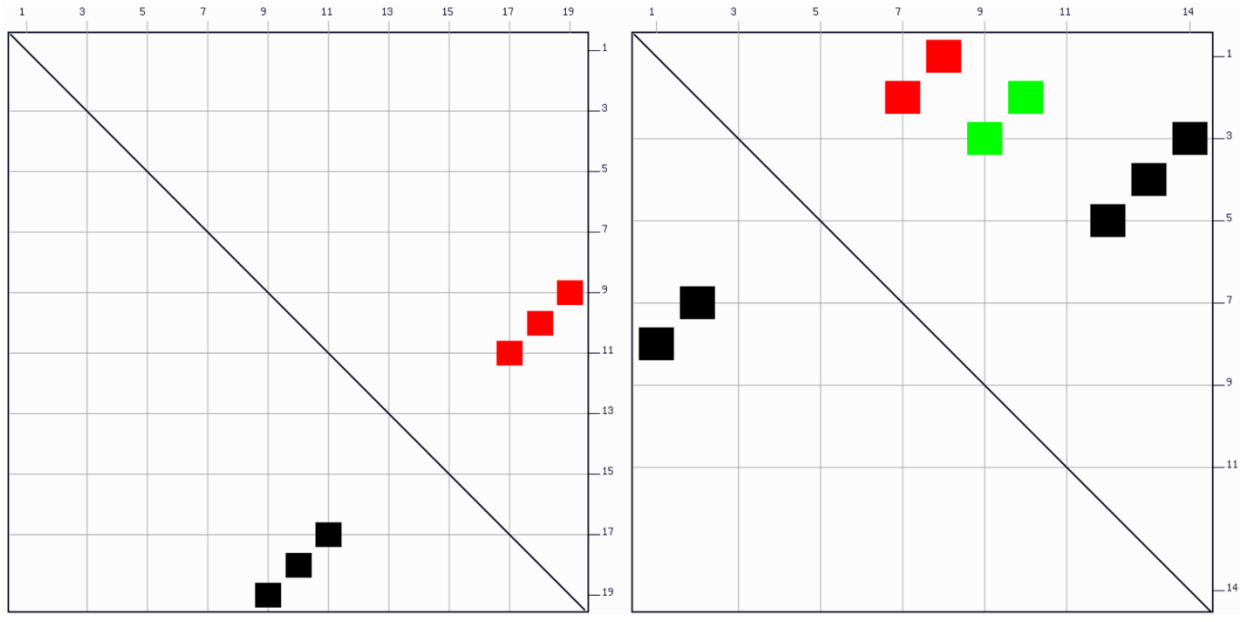
Appendix 2.2.2. Designed, curated e-probes for the detection of *Wheat yellow mosaic virus* in unassembled metagenomes and metatranscriptomes.

Virus	Name	Sequence (5´-3´)	Start (bp)	End (bp)
WYMV	WYMV1	CGAGAAGATCAAGAGTATCATAAAGATGCG	296	326
WYMV	WYMV2	AAAGATGCGTAGCACGACGCGCATCATTGA	317	347
WYMV	WYMV3	GATCCAGCATTGTACCTATTGCAAACTCAG	439	469
WYMV	WYMV4	GCTGCCAACTTAGGGAGAGCTGTGTATGGA	1755	1785
WYMV	WYMV5	TTTCAACGCTGTCTAAATTCGACTGGAATG	2899	2929
WYMV	WYMV6	TTGCATGTGAGCAGCATCGCCCAACAGTGA	2938	2968
WYMV	WYMV7	CAACAGTGATAAGCTCATTCCAAGGCCTTG	2959	2989
WYMV	WYMV8	CCATCTCTGCAACATGCGTAACAATTACAA	3050	3080
WYMV	WYMV9	TGTTGACTGATGAGACACTGTCAAACGCTC	3328	3358
WYMV	WYMV10	ATACATGGAACGAGAAAGCTAAGGAGAAAA	3763	3793
WYMV	WYMV11	TAAAAGCCTTCCTTGACCCAAAACCAATCA	3979	4009
WYMV	WYMV12	GGGATCCGAAAGTTTTCTCAATCCGACTGG	4593	4623
WYMV	WYMV13	ATGTGTGGATGCCCTGTTTTGGATGTTGGA	4680	4710
WYMV	WYMV14	CCGCCCTGCCTGTGGATACAAGCAAGGAGA	4847	4877
WYMV	WYMV15	TCACAGAGCTAGGTTTGACGTACGAGTTTG	6118	6148
WYMV	WYMV16	GAATCTCAGCCATTTACGAAAGCTTCAACA	6301	6331
WYMV	WYMV17	CTGACATCCTAGCTGCGATGACCGGAACAG	6391	6421
WYMV	WYMV18	GAAGGAGGAAGCTCGACTGGCAGCCGCCAC	6521	6551
WYMV	WYMV19	CGCCACCAAAGAGAAATGGTCACTTCCAGA	6725	6755
WYMV	WYMV20	ATTCAGAAGTTAAGACGTGGTCAGACGCTG	6856	6886
WYMV	WYMV21	GAACCCACAGGACATCGAAGTTGCGAAACA	7223	7253
WYMV	WYMV22	CCTTTCACGGCCACGGTTACGATCTTTAAT	7435	7465
WYMV	WYMV23	GGGACGGTTCTATGCACAATTATGCTTCGA	7486	7516
WYMV	WYMV24	CACACACACCGCTCTATCATCTGAGCACAC	7679	7709
WYMV	WYMV25	CTCTACCACAGCGCATATCTGAAGCTTGGA	7892	7922
WYMV	WYMV26	GTTTACCAGCAACACGATTCAACGCTTACA	7988	8018
WYMV	WYMV27	GAAAACAACCATTTACTTAATGCGTGTCTT	8355	8385
WYMV	WYMV28	CTATAGTGGTATGGTCCGGCCAACAAAACG	8742	8772
WYMV	WYMV29	ACCTCTCCGACAGAGCGCCATTTAAAGCTA	8813	8843
WYMV	WYMV30	TATGACAATCCAAATGCATGCTTCACACGA	8947	8977
WYMV	WYMV31	TGTTTTGCAGTCGAGCGCCCATGGATCTGG	9609	9639
WYMV	WYMV32	ACGAATGGCATCTACGCTCCAACGAATTGC	9897	9927
WYMV	WYMV33	TCCCGCTGCTTGCTAAAGCTGAGGCCTCGC	10037	10067
WYMV	WYMV34	CTACTTGCGCCTATAAAGCCCAACTGTCTG	10877	10907
WYMV	WYMV35	TCTCGAACTCTAGTTACGGTACGTCATGGA	11017	11047

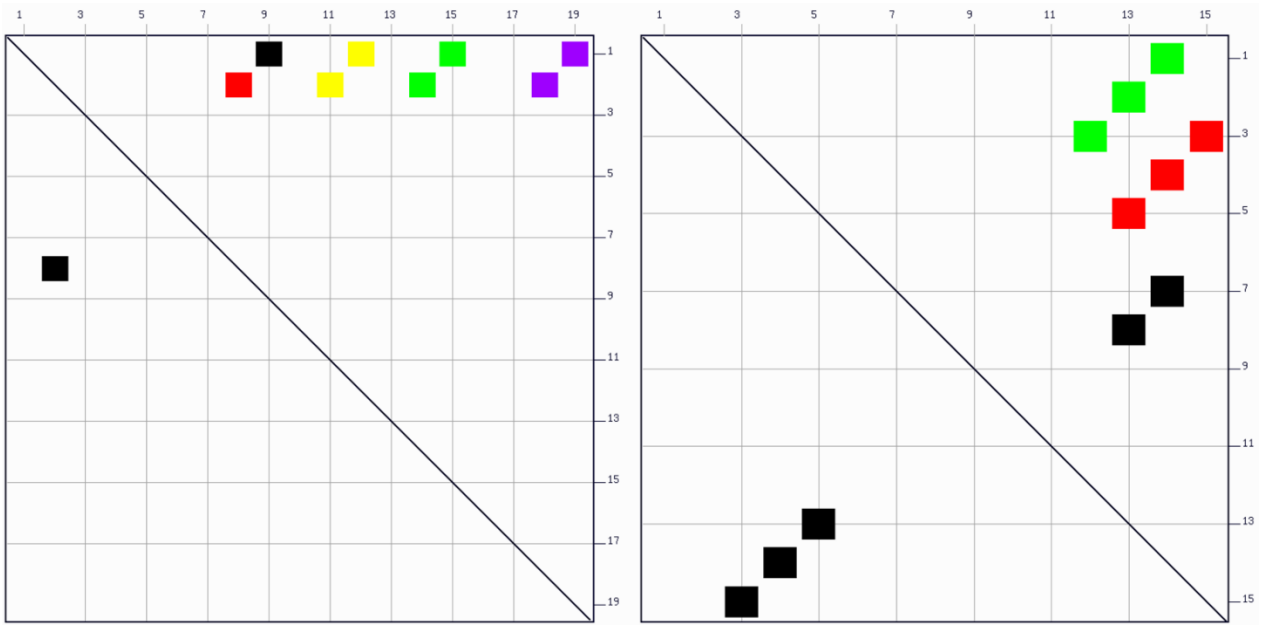
Appendix 2.2.3. mFold evaluation of primer pair WSG1F (ΔG = 0.9) and WSG1R (ΔG = 0.9) used in the amplification of the full genome of WSMV.



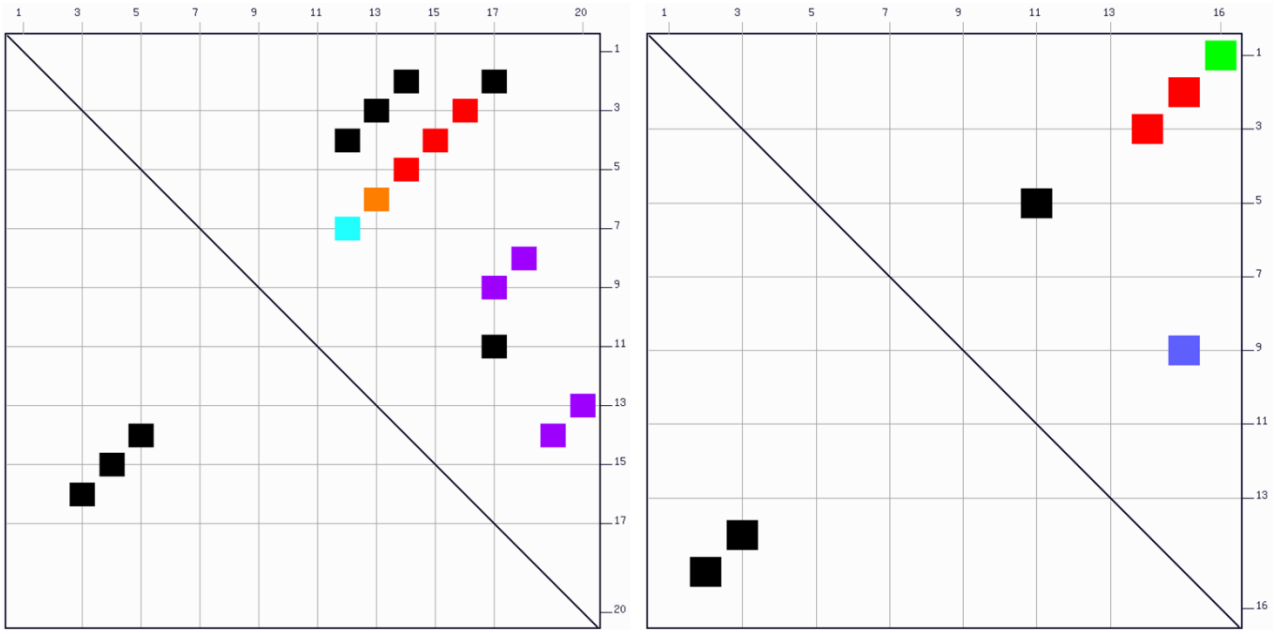
Appendix 2.2.4. mFold evaluation of primer pair WSG2F (ΔG = 0.0) and WSG2R (ΔG = 0.7) used in the amplification of the full genome of WSMV.



Appendix 2.2.5. mFold evaluation of primer pair WSG3F (ΔG = 0.6) and WSG3R (ΔG = 0.5) used in the amplification of the full genome of WSMV.



Appendix 2.2.6. mFold evaluation of primer pair WSG4F (ΔG = 1.0) and WSG4R (ΔG = 0.8) used in the amplification of the full genome of WSMV.



Appendix 2.2.7. Table of the wheat chromosome sequences used to generate mock metagenomic databases and their corresponding NCBI accession numbers.

Wheat Chromosome	NCBI Accession
1A	LS992080.1
1B	LS992081.1
1D	LS992082.1
2A	LS992083.1
2B	LS992084.1
2D	LS992085.1
3A	LS992086.1
3B	LS992087.1
3D	LS992088.1
4A	LS992089.1
4B	LS992090.1
4D	LS992091.1
5A	LS992092.1
5B	LS992093.1
5D	LS992094.1
6A	LS992095.1
6B	LS992096.1
6D	LS992097.1
7A	LS992098.1
7B	LS992099.1
7D	LS992100.1
Un	LS992101.1

Appendix 2.2.8. Summary table of the relationship between the abundance of WSMV in simulated metagenomes with the total average number of probe 'hits' and the average number of probes per 'hit'. The standard error measurement is provided for each mean.

Abundance of WSMV	Total Average Probe Hits	Average Hits Per Probe
Transcriptome	0.75 ± 0.25	0.03 ± 0.01
0%	0.00 ± 0.00	0.00 ± 0.00
0.000001%	0.33 ± 0.21	0.01 ± 0.01
0.00001%	7.00 ± 0.37	0.24 ± 0.01
0.0001%	111.00 ± 1.67332	3.83 ± 0.06
0.001%	1087.00 ± 13.18	37.48 ± 0.45
0.01%	6996.33 ± 0.56	241.25 ± 0.02

Appendix 2.2.9. Summary table of the relationship between the abundance of WSMV in *in vitro* metagenomes with the total average number of probe 'hits' and the average number of probes per hit for WSMV e-probe set one and two. The standard error measurement is provided for each mean.

Abundance of WSMV	Total Average Probe	Hits/Probe Set 1	Hits/Probe Set 2
0%	82.75 ± 16.67	73.5 ± 14.75	9.25 ± 2.17
0.000001%	336.83 ± 48.23	15.97 ± 2.21	3.52 ±0.60
0.00001%	1488.33 ± 642.28	70.07 ± 30.33	17.44 ± 7.35
0.0001%	3521.67 ± 424.79	166.87 ± 19.96	39.02 ± 5.13
0.001%	4864.00 ± 279.78	225.11 ± 10.66	65.24 ± 9.23
100%	7000.00 ± 0.00	250.00 ± 0.00	250.00 ± 250.00

VITA

Patrick Michael Rydzak

Candidate for the Degree of

Doctor of Philosophy

Dissertation: DETECTION OF ARTHROPOD TRANSMITTED VIRUSES OF CEREAL CROPS and RNAI INDUCED RESISTANCE TO *WHEAT STREAK MOSAIC VIRUS*

Major Field: Plant Pathology

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Plant Pathology at Oklahoma State University, Stillwater, Oklahoma in July, 2019

Completed the requirements for the Master of Science in Biology at University of Texas at Tyler, Tyler, Texas in April, 2015.

Completed the requirements for the Bachelor of Science in Marine Biology at Texas A&M University, Galveston, Texas in May, 2012

Experience:

Vector entomology, DNA and RNA extractions, RT-qPCR, primer design, gel electrophoresis, sanger sequencing, Illumina sequencing, capillary electrophoresis, insect colony rearing, viral isolation, RNA interference, molecular cloning, BSL-2 and BSL-3 laboratory SOP, pesticide application, microbial culturing and microscopy.

Professional Memberships:

American Phytopathology Society
Entomological Society of America