Evolutionary studies of wheat streak mosaic-associated viruses and characterization of the wheat virome

by

Carla Bianca Dizon Redila

B.S., University of Guam, 2018

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Approved by:

Major Professor
Dr. Shahideh Nouri

Copyright

Abstract

Wheat viruses including *Wheat streak mosaic virus* (WSMV), *Triticum mosaic virus* (TriMV), *High Plains wheat mosaic emaravirus* (HPWMoV), and *Barley yellow dwarf virus* (BYDV) cost substantial losses in crop yields annually. Although there have been extensive studies conducted on known wheat viruses, currently, there is limited knowledge about all viral components associated with wheat *(Triticum aestivum L.)* including potential novel viruses. Wheat streak mosaic (WSM), a disease of cereals and grasses, costs Kansas farmers millions in yield losses. Although WSMV is considered as the main causal agent of WSM, TriMV and HPWMoV have also been reported in mixed infections. While resistant varieties are utilized to minimize the effects of the disease, genetic variation in associated viruses increases the emergence of potential resistance-breaking isolates. Currently, little is known about the genetic composition of populations of WSM-associated viruses in the field. This study first aims to analyze the genetic variation and characterize the evolutionary mechanism(s) applied by WSM-associated viruses in the field using complete genomes sequences, and also to determine and characterize all viral populations associated with wheat. Field collections of 24 WSM-like symptomatic and asymptomatic wheat samples were used for total RNA deep sequencing, along with 5 historic WSMV samples for the evolutionary studies. Through bioinformatics analysis, sequences were mapped to available reference genomes and *de novo* assembled to identify new viruses. Results of the 2019 field survey showed WSMV as the predominant virus followed by mixed infections of WSMV+TriMV. Recombination was observed to be a major evolutionary force for WSMV but not for TriMV isolates. Phylogenetic analyses based on the obtained full genome sequences demonstrated that, unlike other isolates from the United States, Kansas isolates are widely distributed in sub-clades. Moreover, the phylogenetic studies suggested that TriMV field isolates may be under selection pressure to introduce genetic variations due to the use of resistant varieties in the fields. The full genome sequence of a new Kansas HPWMoV isolate was reported here. In addition to known wheat viruses, viral sequences sharing significantly low (<40%) similarity with potyviruses and recently discovered virga-like viruses were identified. Our analysis showed that these viral sequences belong to a novel putative virus tentatively named *Wheat-associated vipovirus* (WaVPV). Additionally, fungal-associated viruses, such as *Mitovirus*, was also identified. The findings of this study provide a better understanding of the current status of the

genetic structure of WSM-associated viruses in Kansas fields. This, in turn, can aid in developing efficient and durable disease control strategies. The discovery and characterization of novel viruses potentially associated with wheat is important to determine if they may pose a threat to the wheat industry or have the potential to be used as new biological control agents for wheat pathogens.

Table of Contents

List of figures

List of tables

Acknowledgements

I would like to acknowledge my advisor, Dr. Shahideh Nouri, for her guidance and support. I am grateful for the opportunity she has given me to learn and pursue this path and her encouragement to always strive to become the best female scientist that I can be. I would also like to thank Dr. David Cook and Dr. Sanzhen Liu for serving in my committee and providing helpful feedback both relating to my bioinformatics analyses and conveying my results and ideas to an audience effectively.

Thank you to those who've shared wheat samples with me and accompanied me during my surveys of Northwestern Kansas: Dr. Erick DeWolf, Judy O`Mara, Nar Ranabhat, Dylan Mangel, Christian Webb, Jeffrey Ackerman, Dr. Marty Draper, Dr. Sarah Zukoff, Dr. Robert Whitworth, Dr. Punya Nachappa (Colorado State University), and Dr. Stephen Wegulo (University of Nebraska-Lincoln). Thank you to Kansas State University's Integrated Genomics Facility for their assistance in conducting QC analyses and sequencing my libraries.

Thank you to the previous members of the Nouri lab: Alexandria Kieffaber for providing me guidance in learning basic lab etiquette and protocols, former undergraduate (current MSc. student) Savannah Phipps from the REEU program for giving me the opportunity to become a (very proud!) mentor and aiding me with a majority of my viral screenings, Dr. Ferdinand Nanfack-Minkeu for giving me great advice pertaining to research and life, and Dr. Maryam Rastegar for helping with experiments and lab duties during the past few challenging months. Thank you to Jun Huang and Dr. Veerendra Sharma for providing helpful troubleshooting advice for molecular biology techniques and cheering me on during some of the toughest days in lab. Thank you to Dr. Wei Zhang for aiding me in improving my library preparation protocol and being there to answer many of my questions about writing codes and my bioinformatics pipeline.

I would like to extend my appreciation to Dr. G. Curt Fiedler (University of Guam) for being the first person to believe in my potential be a scientist, giving me my first opportunity to pursue research, and being an amazing mentor and undergraduate advisor. Dr. Wei Xiao (University of Guam), the first female scientist role model that I aspired to become, thank you for always listening to me, giving me great advice, introducing me to the world of plants and bioinformatics, and being an awesome plant mom! Thank you Dr. Katharine Lofdahl (University of Guam) for showing me compassion when I needed it the most during my undergraduate years and giving me the push I needed to power through.

Thank you to my family both in Guam and Texas for being very supportive of me throughout this journey. To the friends I've continued to share my troubles and triumphs with back home and the new friends I have made here in Kansas, thank you for always lending an ear and cheering me on. To everyone in the Kansas State Plant Pathology program, from the students to office admins to the faculty, thank you for the encouragement and support and for making this journey a wonderful experience.

Dedication

This thesis is dedicated to the five people I hold dear to my heart.

I would like to dedicate this thesis and all of my hard work to my parents: Rogelio and Angelina Dizon. To my incredibly resilient dad, you gave me the strength I needed during the most difficult times. My mom, the true definition of a strong, educated, and loving woman. I aspire to become even just half of the woman that you are now. I would also like to give this dedication to my grandparents, who helped raise me and my siblings. My grandmother, Felicidad Puno and my late grandfather, Carlos Puno, have provided me the solid foundation I needed as a child to strive to be the best version of myself. I miss you, tatay and nanay.

Finally, I dedicate this work to my husband, Raiden Redila. Thank you for standing by me through all of the happiness, tears, frustrations, and late night writing and study sessions. You helped keep me sane and grounded throughout this crazy journey.

I hope I make you all proud.

Chapter 1 - Literature review

Wheat (*Triticum aestivum*) is one of the most economically important crops grown across the world. Along with rice, maize, and soybeans, wheat accounts for 50% of the world food calories (Nelson et al. 2010). In the United States, wheat is primarily grown in the Great Plains, with Kansas as the second leading producer in 2019 (NASS 2020). Kansas harvested millions of bushels amounting to $1.9 billion in 2019 (NASS 2020). The projected demand for wheat is expected to increase to 40% by 2050, making it crucial to increase crop supply to ensure food security globally (Fischer et al. 2014). However, plant pathogens including viruses continue to have a significant negative impact on the production of wheat in Kansas and other parts of the world.

Wheat streak mosaic disease

Wheat streak mosaic (WSM) is one of the most economically important viral diseases in Kansas. WSM causes a significant yield loss annually. In 2017 alone, WSM had a devastating impact in Kansas wheat production and cost Kansas farmers $76 million in yield losses (Hollandbeck et al. 2017). WSM is distributed throughout the Great Plains, Pacific Northwest, and Ohio. Due to the expansive distribution of WSM and its economic impact to the wheat industry, comprehensive studies have been done to gain a better understanding of this disease.

WSM is a disease complex which consists of three documented RNA viruses: *Wheat streak mosaic virus (*WSMV), *Triticum mosaic virus* (TriMV), and *High Plains wheat mosaic emaravirus* (HPWMoV). The viruses share similar symptoms, such as yellow streaks on the leaves, chlorosis, and in severe cases, stunting. Single and mixed-infections of WSM were observed in previous surveys of the Great Plains. WSMV has been reported in many countries,

such as Australia, Argentina, Canada, Mexico, and Eastern Europe and is considered the main causal agent of WSM due to its higher prevalence in comparison to TriMV and HPWMoV (Hadi et al. 2011).

In 2008, a survey of wheat viruses in the Great Plains utilizing enzyme-linked immunosorbent assay (ELISA) tests kits found that single infections of WSMV was the most prevalent at an average of 47%, followed by single infections of HPWMoV at 19% and TriMV at 17% (Burrows et al. 2009). For mixed infections, the average was all equal at 13% for WSMV + TriMV, HPWMoV + WSMV, and TriMV + HPWMoV and 5% for the infection of WSMV + TriMV + HPWMoV (Burrows et al. 2009). When focusing solely on the Kansas results, both the single and mixed infection rates were not equally distributed. The mixed infection of WSMV + TriMV was at 21%, WSMV + HPWMoV at 15%, and HPWMoV + TriMV at 13% and for single infections, the percent of samples infected for WSMV, TriMV, and HPWMoW were 62%, 30%, and 38% respectively. (Burrows et al. 2009). In 2010-2011, another survey was conducted in the Great Plains again using ELISA, spanning over the summer and spring seasons. This study determined that WSMV was once again predominant in the fields at 73%, followed by HPWMoV at 15%, and TriMV at 6% (Byamukama et al. 2013). In the same study, the double and triple infections of WSM were also observed throughout the Great Plains: TriMV + WSMV at 20%, HPWMoV + WSMV at 11%, and TriMV + WSMV + HPWMoV at 3% (Byamukama et al. 2013). Unlike the survey conducted by Burrows et al. (2009), there was no presence of mixed infection of TriMV + HPWMoV detected in the later survey. When examining just the Kansas results from the survey by Byamukama et al. (2013), similar results of the order of virus prevalence for both single and mixed infections were found.

The main distribution of WSM in the fields are dependent on the arthropod vector shared by all three viruses: wheat curl mites (Slykhuis 1955; Seifers et al. 1997, 2009).

Wheat streak mosaic virus (WSMV)

WSMV was first described as yellow mosaic in Nebraska in 1922 (McKinney 1937). WSMV is the type species classified under the family *Potyviridae* and the genus *Tritimovirus* (Rabenstein et al. 2002). Along with the vector transmission, WSMV can also be transmitted through mechanical inoculation and is also seed borne (Jones et al. 2005). The 9.3 kb genome is a monopartite, single-stranded, positive-sense RNA, encapsidated in filamentous virions (Stenger et al. 1998). The 5' and the 3' terminals contain a VPg and a Poly (A) tail, respectively (Singh et al. 2018). The genome encodes a large polyprotein, which is proteolytically processed into 10 individual proteins: P1 (protein 1), HC-Pro (helper component protein), P3 (protein 3), 6K1, CI (cytoplasmic inclusion), 6K2, NIa-VPg (nuclear inclusion a viral protein genome-linked), NIa-Pro (nuclear inclusion a protease), NIb (nuclear inclusion b), and CP (coat protein) (Stenger et al. 1998; Choi et al. 2002; Chung et al. 2008; Singh et al. 2018). Unlike other potyviruses, the P1 of WSMV acts as an enhancer of pathogenicity and a suppressor of RNA silencing (Tatineni et al. 2012; Young et al. 2012). The HC-Pro has been discovered to be required for vector transmission and one of the proteolytic processors of WSMV (Stenger et al. 2005). The P3 region is less understood, but observed to bind to P1, HC-Pro, and CI for WSMV, suggesting that this protein may play a role in the viral replication and movement (Choi et al. 2000; Choi et al. 2005). P3N-PIPO, a small overlapping protein typically found in the P3 region of potyviruses, was also discovered to be translated from the P3 region of WSMV and has been found to be essential in cell-to-cell movement and internal RNA (Chung et al. 2008). Previous studies of other potyviruses propose that the 6k1 and 6k2 proteins are important in the formation of viral vesicles and play essential roles in the viral replication (Nigam et al. 2019;

Cui and Wang 2016). The NIa-Pro is a major proteinase which aids in processing the polyprotein into 10 individual proteins (Singh et al. 2018). The NIb is presumed to be the replicase due to the presence of conserved polymerase motifs within this protein (Choi et al. 2000). The NIa-VPg binds with the eukaryotic translation initiation factor 4E (eIF4E), which initiates the translation of the viral genome and has also been found to suppress the production of the host proteins (Urcuqui-Inchima et al. 2001; Adams et al. 2005). The CP is multi-functional and is utilized for host- and strain- specific long-distance transport, virus assembly, and cell-to-cell movement (Tatineni and French 2014, Tatineni et al. 2011, 2014a).

Triticum mosaic virus (TriMV)

TriMV was discovered in Kansas in 2006 after RonL, a wheat variety resistant to WSMV, was found to exhibit viral symptoms, such as yellow streaks on the leaves (Seifers et al. 2008). TriMV is a type species classified under the same family as WSMV in *Potyviridae* but under the different genus *Poaceavirus* (Fellers et al. 2009; Tatineni et al. 2009). Similar to WSMV, the genome of TriMV encodes a large polyprotein which is cleaved into 10 different proteins (Fellers et al. 2009). The genome spans to ~10.2 kb and contains an unusually long ~739 nt 5' untranslated region (UTR) which was discovered to be a strong translation element (Fellers et al. 2009; Roberts et al. 2015). Similar to WSMV, TriMV encodes the P1 to function as an RNA silencing suppressor and pathogenicity enhancer (Tatineni et al. 2012; Young et al. 2012).

High Plains wheat mosaic emaravirus (HPWMoV)

HPWMoV was first discovered infecting maize in the Great Plains in 1993 (Jensen et al. 1996). HPWMoV was classified under the family *Fimoviridae* and the genus *Emaravirus* (Stewart 2016). The genome of HPWMoV was recently sequenced which consists of 8

negative-sense, single-stranded RNA genomes: RNA1 (6981 nt)- is predicted to encode the RdRp, RNA2 (2211 nt)- glycoprotein, RNA3 (1439 nt or 1441 nt)- nucleocapsid protein, RNA4 (1682 nt)- movement protein, RNA5 (1715 nt)- no known function, RNA6 (1752 nt)- no known function, RNA7 (1434 nt)- RNA silencing suppressor, and RNA8 (1339 nt)- RNA silencing suppressor (Tatineni et al. 2014b).

In the first study conducted to characterize and obtain the complete genome of HPWMoV, two variants of RNA 3 were found and were referred to as 3A (1439 nt) and 3B (1441nt) (Tatineni et al. 2014b). RNAs 3A and 3B had differences of 12.5% at the nucleotide and 11.1% at the amino acid level (Tatineni et al. 2014b). Only RNA 3 contained variants in the sample, so mixed infection was ruled out and the ratio of the sequences was 5:1, making 3A the major nucleocapsid protein and 3B minor (Tatineni et al. 2014b). The specific role of the two variants of the nucleocapsid protein is not known. Another study confirmed the presence of both 3A and 3B in the samples, and also samples which encode only the 3A major protein (Stewart 2016). A phylogenetic study of the amino acid of RNA 3 was conducted to determine the relationship between the variants and if host or geographic regions play a role. The analysis resulted into two clades in which one clade contained two sub-clades of RNAs 3A and 3B separately and the other major clade consisted of RNA 3 with a similarity of 77-85% at the amino acid level with the 3B (Stewart 2016). This relationship suggests that the host or geography do not play a role in the variation. Kansas, Nebraska, and Ohio HPWMoV isolates from wheat were found grouped together in 3A and 3B sub-clades and the other clade consisted of isolates from Kansas, Texas and Ohio with a mixture of collection hosts, such as wheat, barley, and maize (Stewart 2016).

The wheat curl mite (WCM)

All documented WSM-associated viruses are transmitted by an eriophyoid mite called wheat curl mite (WCM; *Aceria tosichella*) (Slykhuis 1955; Seifers et al. 1997, 2009). WCM has two different biological types throughout the United States: Biotype 1 and Biotype 2 (Slykhuis 1955; Malik 2001; Hein et al. 2012). Both biotypes are able to transmit WSMV, but biotype 2 is much more efficient in transmitting the virus (Wosula et al. 2016). In a study to determine the efficiency of WCM transmissibility of TriMV, biotype 2 populations were able to transmit TriMV but biotype 1 was not (McMechan et al. 2014). Interestingly, the transmission efficiency of TriMV increases in mixed infections with WSMV (Seifers et al. 2009). Although less information is available about HPWMoV transmission, the result of a study showed that WCM biotype 1 is a very poor vector for HPWMoV and only biotype 2 is able to transmit the virus (Seifers et al. 2002). Similar to TriMV, the transmission rate increases when HPWMoV is in mixed infections with WSMV (Seifers et al. 2002). The declining incidences of single infection of TriMV from previous WSM surveys may have been attributed to the antagonistic effect of the virus infection on the vector's survival (McMechan et al. 2014; Burrows et al. 2009; Byamukama et al. 2013).

The differential transmission rates of WSM by WCM suggests that the distribution of WSM may be strongly influenced by the proportion of WCM biotypes found in the fields. Both WCM biotypes were found to co-occur in the field and the same grain head of wheat (Khalaf et al. 2020). A survey of WCM reported that different counties in Kansas exhibited varied ratios of biotypes 1 and 2, but the mean ratio of biotype 1:2 has been found to be even (Khalaf et al. 2020). In addition to this, there was difference observed in the comparison of the same county surveyed in different years so determining the most dominant biotype would prove to be difficult (Khalaf et al. 2020).

WCM is an important contributing factor in the distribution of WSM in the field. However, other factors such as the virus-virus, vector-virus, and virus-plant-vector interactions may also be at play. The current knowledge about these interactions is limited. The temperature and precipitation conditions conducive to the survival of one biotype over the other also plays an indirect role in determining the dispersal of WCM biotypes and in turn, the spread of single and mixed infections of WSM in the fields. Along with controlling WSM, WCM has also been the target of many disease management strategies to decrease the spread of WSM.

Synergistic relationships of plant viruses

The interaction of two viruses in the same host resulting in the increase of titer of one or both viruses or the increase of pathogenicity and disease symptoms is referred to as the synergistic relationship of viruses (Pruss et al. 1997). Many potyviruses engage in synergistic relationship with other related or unrelated viruses. When *Potato virus X* (PVX) is co-inoculated with *Potato virus Y (*PVY), the symptoms lead to necrosis and the titer of PVX increases ten-fold compared to the single infection (Vance 1991). In other cases, synergism produces a devastating viral disease such as the maize lethal necrosis disease (MLN), which consists of a mixed infection of *Maize chlorotic mottle virus* (MCMV) and a potyvirus: WSMV, *Sugarcane mosaic virus* (SCMV), or *Maize dwarf mosaic virus* (MDMV) (Mbega et al. 2016). This synergistic relationship can lead to a yield loss ranging from 10% to 90% (Mbega et al. 2016).

In co-infections, WSMV and TriMV engage in a synergistic relationship in which WSMV replicates at a higher rate in the early stages of infection and conversely, TriMV accumulates higher titer in the late stages of infection (Tatineni et al. 2014a). The synergistic effect of being in a mixed infection with WSMV may also play a big factor in the increased prevalence of mixed infections of WSMV + TriMV found in Kansas fields. To date, there has

been no study conducted to determine if there is a synergistic relationship between WSMV + HPWMoV, HPWMoV + TriMV, and WSMV + TriMV + HPWMoV, so the effects of mixed infections to the viral titer of WSM when in mixed infection with HPWMoV remained unclear.

WSM management strategies

Cultural practices and the use of resistant varieties are the main strategies used to control both WCM and WSM. WCM require a living host to survive during the post-harvest period and summer fallow (Webb 2018). The volunteer wheat or WSM susceptible grass species utilized by WCM before wheat planting are referred to as "green bridges" (McKinney 1937). The disease cycle of WSM has been predicted to be very similar considering the shared vector: WCM uses green bridges over the summer, travels through air to land on the crop after the green bridge starts to senesce, and WCM transmits and over-winters on the winter wheat seedlings. Due to the increased viral and WCM spread posed by these green bridges, many farmers have been encouraged to remove volunteer wheat and grasses before planting and to plant at later dates (Velandia et al. 2010).

Host resistance genes have been utilized to develop resistant varieties to control WSM. Currently, there are three known resistance genes for WSM: *Wsm1, Wsm2,* and *Wsm3 (*Friebe et al. 1991; Haley et al. 2002; Lu et al. 2011; Liu et al. 2011). Both *Wsm2* and *Wsm3* genes are resistant against WSMV and TriMV and *Wsm1* is resistant against WSMV (Kumssa et al. 2019). *Wsm1* and *Wsm2* genes are applied in different wheat varieties and deployed in Kansas fields, but unlike *Wsm3*, these resistant genes are temperature-sensitive and have been proven to be ineffective in temperatures above 25 $^{\circ}$C (Seifers et al. 2006, 2007; Haley et al. 2002). Wheat varieties containing *Wsm3* are not commercially available yet. Although the use of resistant varieties is promising, a recent discovery of a WSMV isolate overcoming the *Wsm2* resistance and potential WSMV and TriMV isolates overcoming resistance have been reported

in Kansas (Kumssa et al. 2019; Fellers et al. 2019). These events indicate that the viral populations of WSM might be undergoing evolutionary changes in Kansas fields to overcome this resistance and emphasizes the importance of studying the evolutionary mechanisms of WSM-associated viruses.

Plant virus evolution

Virus evolution is the process in which viruses undergo genetic changes over time as a result of adaptations in their environment. The three major evolutionary forces, which help shape genetic diversity of viruses are: mutation, recombination, and reassortment (Roossinck 1997). In RNA viruses, the RNA dependent RNA polymerase (RdRp) involved in RNA replication lacks a proofreading mechanism. This makes the RdRp prone to errors or mutations. The estimated error rate for the RdRp is an average of 10^{-4} (Roossinck 1997). Due to the high error rate of the RdRp, multiple copies of genome variants or viral quasispecies are created and may be subjected to selection pressure (Domingo and Perales 2019). Recombination occurs through template switching in which the RdRp falls off the original template and continues replication using a different template (Roossinck 1997). This creates a recombinant which consists of complementary strands from two source of templates. Recombination rates for some positive-sense, single-stranded RNA viruses, such as potyviruses, bromoviruses, and coronaviruses ranges between 1.38×10^{-4} to 1.4×10^{-5} per site per generation (Simon-Loriere et al. 2011.) Reassortment as one of the major evolutionary forces is restricted to multipartite viruses only such as *Tomato spotted wilt virus* (Fontana et al. 2019), *Rice grassy stunt virus* (Miranda et al. 2000), and *Alfalfa mosaic virus* (Bergua et al. 2014) and is the genetic exchange of RNA segments before packing the segments into the viral particles (virions) (Worobey and Holmes 1999).

Despite the error-prone RdRp and the other methods of introducing genetic variation to viral genomes, previous evolutionary studies indicate that genetic stability is favored by plant viruses (Garcia –Arenal et al. 2001). Negative or purifying selection is a process in which variants containing deleterious or less fit traits are selected against and they are decreased in the population (Worobey and Holmes 1999). In a genetic diversity study of *Barley yellow dwarf virus* (BYDV), the viral population was observed to be stable and purifying selection played a major role in shaping the population's diversity (Wu et al. 2011). Genetic bottleneck is also another approach used to maintain stability, which is done through a sharp reduction in the population size. Genetic bottleneck may occur through infection of a new host, vector transmission, or systemic infection of the host (Garcia Arenal et al. 2001).

Regarding the WSM-associated viruses, previous phylogenetic study based on the coat protein sequences of WSMV isolates divided them into four different clades (Stenger et al. 2002). Clade A contains an isolate from Mexico, Clade B contains isolates from Europe, Clade C contains an individual isolate from Iran, and Clade D contains isolates from USA, Argentina, and Turkey (Stenger et al. 2002). Clade D was further divided into 4 sub-clades in which Kansas isolates were found in 3 sub-clades, grouped with Colorado, Kentucky, Missouri, and Nebraska. Similar to the Stenger et al. (2002) study, the phylogenetic analysis of the coat protein of isolates from Brazil was also done and resulted in a similar topology (Mar et al. 2013). Clades A to D contained similar groups, but the isolates from Brazil were clustered with Argentina in a sub-clade in Clade D. In the phylogenetic tree generated in this study, reference Kansas isolates were found in two sub-clades of Clade D along with Idaho, Nebraska, and Turkey isolates (Mar et al. 2013). Another phylogenetic study of the coat protein included isolates from Australia produced the similar found clades, with the addition of the Australia isolates in the sub-clade D2 with the U.S. Pacific Northwest and Argentina reference isolates (Dywer et al. 2007; Stenger and French 2009). Interestingly, Kansas and other Great Plains

isolates were found distributed throughout most of Clade D's sub-clades but were not found in sub-clade D2, which is the sub-clade containing isolates from the U.S. Pacific Northwest, Argentina, Australia, and Brazil. This suggests that the U.S. Pacific Northwest isolates probably share origins with isolates from Brazil, Argentina, and Australia possibly through common source of seeds or farming material. It also indicates that the other U.S. isolates, such as isolates from Kansas may have had a separate origin or introductory source for WSMV.

In contrast, a phylogenetic study also focusing on the coat protein was utilized to study isolates from Washington (Robinson and Murray 2012). The topology was divided into two clades: clade I contained isolates from Washington and European countries and clade 2 contained other Washington isolates along with reference isolates from Argentina, other states from the U.S. Pacific Northwest, Texas, and Kansas (Robinson and Murray 2012). In the same study, recombination analysis of the coat protein detected recombination events and determined that the hotspots were at the 3' end of the sequence (Robinson and Murray 2012). The similarities with the Argentina, Brazil and Europe isolates raises even more questions about the origin of the Washington isolates and whether or not they had been introduced from these countries.

All previous studies of WSMV utilized one of ten encoded proteins- the coat protein, which may not provide comprehensive information of the phylogenetic relationships among WSMV isolates. To date, there has been no evolutionary study done using the whole genome sequence of WSMV or evaluating the selection pressures acting upon the virus genome.

Although TriMV is not as widely studied as WSMV, there have been previous studies done to determine the diversity of the virus population through the sequencing of the coat protein (Fuentes- Bueno et al. 2011). The results exhibited that TriMV has a homogenous population and had very little variation in comparison to WSMV (Fuentes-Bueno et al. 2011).

Another study examining the population dynamics of TriMV using the P1 and CP indicated that TriMV populations maintain stability through genetic bottlenecking through systemic movement and vector transmission (Bartels 2014). There is currently no phylogenetic studies conducted to examine the relationships among TriMV isolates.

Compared to TriMV and WSMV, HPWMoV is poorly studied and to date, there are no evolutionary studies conducted on this virus. HPWMoV is poorly studied due to the difficulty of upkeep of maintaining a virus source for experimental studies. Unlike WSMV and TriMV, HPWMoV cannot be inoculated mechanically and requires the use of infected WCM or vascular puncture inoculation in order to conduct studies (Tatineni et al. 2014b).

The wheat virome

Metagenomics is the study of a microbial community in an environmental sample using sequencing technologies (Roossinck et al. 2015). Through plant viral metagenomics, studies have been conducted to gain a better insight in the virome, which are all viral populations within a holobiont or ecosystem (Roossinck et al. 2015). The three mentioned viruses, WSMV, TriMV and HPWMoV are the documented viruses associated with the WSM. However, if these are the only viruses causing WSM is not clear. In an innovative world filled with numerous sequencing methods such as next generation sequencing (NGS), conducting metagenomics analysis with the purpose of identifying all known and unknown viral sequences associated with a specific host has become more feasible and economic. Two recent studies used the NGS technique to identify all viral sequences associated with wheat in Czech Republic and Ohio. Using small RNA deep sequencing, WSMV was determined as the most dominant wheat virus in Czech Republic (Singh et al. 2020). The same study also reported the first observation of *Cereal yellow dwarf virus* (CYDV) and *Wheat yellow dwarf virus* (WYDV) in the Czech Republic (Singh et al. 2020). Moreover, other wheat known viruses such as *Barley yellow dwarf*

virus (BYDV), one of the causal agents for the barley yellow dwarf disease has also been found across the Great Plains and are also considered economically important (Miller et al. 1997; Lister and Ranieri 1995). In Ohio, high-throughput RNA sequencing was conducted and BYDV and *Wheat spindle streak mosaic virus* (WSSMV) were detected the most (Hodge et al. 2020). Additionally, complete genome sequences of two HPWMoV were obtained in this study (Hodge et al. 2020).

Although these wheat virome studies are informative, their focus was only on known viruses; while, virome analysis may also be used to discover unknown, novel viruses which may be either pathogenic viruses and associated with WSM or persistent non-pathogenic viruses which can be viable for biocontrol.

Objectives

All previous studies focused on only one gene encoding the coat protein to study the genetic diversity and phylogenetic analyses of WSMV; while, no robust whole-genome analysis has been yet conducted. Moreover, there are currently no phylogenetic or evolutionary studies performed on TriMV and/or HPWMoV. A better understanding of the genetic variation of the WSM-associated viruses through comprehensive whole-genome phylogeny will enable a better picture of the current status of WSM populations in Kansas fields which eventually improve the disease management. In the other hand, the current knowledge about the wheat virome has obtained by focusing only on the known wheat viruses; however, the virome is composed of all viral communities associated with wheat. Hence, we used two different research pipelines here to fill these knowledge gaps by following these two aims:

1. Determine the distribution of WSM in Kansas fields and study the phylogenetic relationships and identify the major evolutionary force shaping the viral populations using the whole genome sequences.
2. Identify and characterize metagenomics-driven wheat virome.

References

Adams, M. J., Antoniw, J. F., & Beaudoin, F. (2005). Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Mol Plant Pathol*, *6*(4), 471–487. https://doi.org/10.1111/j.1364-3703.2005.00296.x

Bartels, M. S. (2014). Population dynamics of Triticum mosaic virus in various host species. PhD. Dissertation. University of Nebraska.

Bergua, M., Luis-Arteaga, M., & Escriu, F. (2014). Genetic Diversity, Reassortment, and Recombination in Alfalfa mosaic virus Population in Spain. *Phytopathol, 104*(11), 1241-1250. https://doi.org/10.1094/PHYTO-11-13-0309-R

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 Prog, 10*(1), 14. https://doi.org/10.1094/php-2009-0706-01-rs

Byamukama, E., Seifers, D. L., Hein, G. L., De Wolf, E., Tisserat, N. A., Langham, M. A. C., Osborne, L. E., Timmerman, A., & Wegulo, S. N. (2013). Occurrence and Distribution of Triticum mosaic virus in the Central Great Plains. *Plant Dis, 97*(1), 21-29. https://doi.org/10.1094/PDIS-06-12-0535-RE

Chare, E. R., Gould, E. A., & Holmes, E. C. (2003). Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J Gen Virol, 84*(Pt 10), 2691-2703. https://doi.org/10.1099/vir.0.19277-0

Choi, I.-R., Horken, K. M., Stenger, D. C., & French, R. (2002). Mapping of the P1 proteinase cleavage site in the polyprotein of Wheat streak mosaic virus (genus Tritimovirus). *J Gen Virol, 83*(2), 443–450. https://doi.org/10.1099/0022-1317-83-2-443

Choi, I.-R., Horken, K. M., Stenger, D. C., & French, R. (2005). An internal RNA element in the P3 cistron of Wheat streak mosaic virus revealed by synonymous mutations that affect both movement and replication. *J Gen Virol, 86*(9), 2605–2614. https://doi.org/10.1099/vir.0.81081-0

Chung, B. Y.-W., Miller, W. A., Atkins, J. F., & Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *PNAS, 105*(15), 5897–5902. https://doi.org/10.1073/pnas.0800468105

Cui, H., & Wang, A. (2016). Plum Pox Virus 6K1 Protein Is Required for Viral Replication and Targets the Viral Replication Complex at the Early Stage of Infection. *J Virol, 90*(10), 5119–5131. https://doi.org/10.1128/jvi.00024-16

Domingo, E., & Perales, C. (2019). Viral quasispecies. *PLOS Genet, 15*(10), e1008271. https://doi.org/10.1371/journal.pgen.1008271

Dwyer, G. I., Gibbs, M. J., Gibbs, A. J., & Jones, R. A. C. (2007). Wheat streak mosaic virus in Australia: Relationship to Isolates from the Pacific Northwest of the USA and Its Dispersion Via Seed Transmission. *Plant Dis, 91*(2), 164–170. https://doi.org/10.1094/pdis-91-2-0164

Fellers J.P., Seifers D., Ryba-White M., and Martin T.J. (2009). The complete genome sequence of Triticum mosaic virus, a new wheat-infecting virus of the High Plains. *Arch Virol, 154*, 1511-1515. https://doi.org/10.1007/s00705-009-0462-1

Fellers, J. P., Webb, C., Fellers, M. C., Shoup Rupp, J., & De Wolf, E. (2019). Wheat Virus Identification Within Infected Tissue Using Nanopore Sequencing Technology. *Plant Dis*, *103*(9), 2199–2203. https://doi.org/10.1094/pdis-09-18-1700-re

Fischer, T., Byerlee, D., & Edmeades, G. (2014). Crop yields and global food security: Will yield increase continue to feed the world? Australia: Australian Centre for International Agricultural Research, Canberra, Australia.

Friebe, B., Mukai, Y., Dhaliwal, H. S., Martin, T. J., & Gill, B. S. (1991). Identification of alien chromatin specifying resistance to wheat streak mosaic and greenbug in wheat germ plasm by C-banding and in situ hybridization. *Theor Appl Genet*, *81*(3), 381–389. https://doi.org/10.1007/BF00228680

Fuentes-Bueno, I., Price, J. A., Rush, C. M., Seifers, D. L., & Fellers, J. P. (2011). Triticum mosaic virus Isolates in the Southern Great Plains. *Plant Dis*, *95*(12), 1516–1519. https://doi.org/10.1094/pdis-04-11-0281

García-Arenal, F., Fraile, A., & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol*, *39*, 157–186. https://doi.org/10.1146/annurev.phyto.39.1.157

Hadi, B. A. R., Langham, M. A. C., Osborne, L., & Tilmon, K. J. (2011). Wheat Streak Mosaic Virus on Wheat: Biology and Management. *J Integr Pest Manag*, *2*(1), 1–5. https://doi.org/10.1603/ipm10017

Haley, S., Martin, T., Quick, J., Seifers, D., Stromberger, J., Clayshulte, S., Clifford, B., Peairs, F., Rudolph, J., Johnson, J., Gill, B., & Friebe, B. (2002). Registration of CO960293-2 wheat germplasm resistant to Wheat streak mosaic virus and Russian wheat aphid. *Crop Sci, 42*, 1381-1382. https://doi.org/10.2135/cropsci2002.1381

Hein, G. L., French, R., Siriwetwiwat, B., & Amrine, J. W. (2012). Genetic Characterization of North American Populations of the Wheat Curl Mite and Dry Bulb Mite. *J Econ Entomol, 105*(5), 1801–1808. https://doi.org/10.1603/ec11428

Hodge, B. A., Paul, P. A., & Stewart, L. R. (2020). Occurrence and High-Throughput Sequencing of Viruses in Ohio Wheat. *Plant Dis, 104*(6), 1789–1800. https://doi.org/10.1094/PDIS-08-19-1724-RE

Hollandbeck G.F., DeWolf E., and Todd T. (2017). Kansas cooperative plant disease survey report preliminary 2017 Kansas wheat disease loss estimates. https://agriculture.ks.gov/docs/default-source/pp-disease-reports-2012/2017-ks-wheat-disease-loss-estimates.pdf?sfvrsn=d49587c1_0

Jensen, S.G., Lane, L.C., & Seifers, D.L. (1996). A new disease of maize and wheat in the high plains. *Plant Dis, 80,* 1387-1390.

Jones, R. A. C., Coutts, B. A., Mackie, A. E., & Dwyer, G. I. (2005). Seed Transmission of Wheat streak mosaic virus Shown Unequivocally in Wheat. *Plant Dis, 89*(10), 1048–1050. https://doi.org/10.1094/pd-89-1048

Khalaf, L., Timm, A., Chuang, W.-P., Enders, L., Hefley, T. J., & Smith, C. M. (2020). Modeling Aceria tosichella biotype distribution over geographic space and time. *PLOS ONE*, *15*(5), e0233507. https://doi.org/10.1371/journal.pone.0233507

Kuczyński, L., Rector, B. G., Kiedrowicz, A., Lewandowski, M., Szydło, W., & Skoracka, A. (2016). Thermal Niches of Two Invasive Genotypes of the Wheat Curl Mite Aceria tosichella: Congruence between Physiological and Geographical Distribution Data. *PLOS ONE, 11*(4), e0154600. https://doi.org/10.1371/journal.pone.0154600

Kumssa, T. T., Rupp, J. S., Fellers, M. C., Fellers, J. P., & Zhang, G. (2019). An isolate of Wheat streak mosaic virus from foxtail overcomes Wsm2 resistance in wheat. *Plant Pathol*. https://doi.org/10.1111/ppa.12989

Lister, R.M. & Ranieri, R. (1995). Distribution and economic importance of barley yellow dwarf. In J.C. D'Arcy & P.A. Burnett, eds. Barley yellow dwarf, 40 years of progress, p. 29-53. St Paul, MN, USA, APS Press.

Liu, W., Seifers, D. L., Qi, L. L., Friebe, B., & Gill, B. S. (2011). A Compensating Wheat–Robertsonian Translocation Conferring Resistance to and Mosaic Virus. *Crop Sci*, *51*(6), 2382. https://doi.org/10.2135/cropsci2011.03.0118

Lu, H., Price, J., Devkota, R., Rush, C., & Rudd, J. (2011). A Dominant Gene for Resistance to in Winter Wheat Line CO960293-2. *Crop Sci*, *51*(1), 5. https://doi.org/10.2135/cropsci2010.01.0038

Malik, R. (2001). Molecular genetic characterization of wheat curl mite, Aceria tosichella Keifer (Acari: Eriophyidae), and wheat genes conferring wheat curl mite resistance. PhD. Dissertation.

Mar, T. B., Lau, D., Schons, J., Pereira, P. R. V. D. S., & Carminatti, A. J. (2013). Identification and Characterization of Wheat Streak Mosaic Virus Isolates in Wheat-Growing Areas in Brazil. *Int J Agron*, *2013*, 1–6. https://doi.org/10.1155/2013/983414

Mbega, E., Ndakidemi, P., Mamiro, D., Mushongi, A., Kitenge, K., & Ndomba, O. (2016). Role of potyviruses in synergistic interaction leading to maize lethal necrotic disease on maize. *Int J Curr Microbiol App Sci* *5*(6): 85-96. http://dx.doi.org/10.20546/ijcmas.2016.506.011

McKinney, H.H. (1937). Mosaic diseases of wheat and related cereals. *USDA, 4*, 1–22. https://archive.org/stream/mosaicdiseasesof442mcki/mosaicdiseasesof442mcki_djvu.txt

Mcmechan, A. J., Tatineni, S., French, R., & Hein, G. L. (2014). Differential Transmission of Triticum mosaic virus by Wheat Curl Mite Populations Collected in the Great Plains. *Plant Dis*, *98*(6), 806–810. https://doi.org/10.1094/pdis-06-13-0582-re

Miller And, W. A., & Rasochová, L. (1997). Barley Yellow Dwarf Viruses. *Annu Rev Phytopathol*, *35*(1), 167–190. https://doi.org/10.1146/annurev.phyto.35.1.167

Miranda, G. J., Azzam, O., & Shirako, Y. (2000). Comparison of Nucleotide Sequences between Northern and Southern Philippine Isolates of Rice Grassy Stunt Virus Indicates Occurrence of Natural Genetic Reassortment. *Virol*, *266*(1), 26–32. https://doi.org/10.1006/viro.1999.0068

National Agricultural Statistics Service. (2020). *Kansas States 2019 Agricultural Overview*. https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=kansas

Nelson G.C., Rosegrant M.W., Palazzo A., Gray I., Ingerstoll C., Robertson R. (2010). Food security, farming, and climate change to 2050: scenarios, results, policy options. Research Monograph. International Food Policy Research Institute, Washington, DC. https://doi.org/10:2499/9780896291867.

Nigam D., LaTourrette K., Souza P.F.N., Garcia-Ruiz H. 2019. Genome-wide variation in Potyviruses. *Front Plant Sci, 12*,1-28.

Pruss, G., Ge, X., Shi, X. M., Carrington, J. C., & Bowman Vance, V. (1997). Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. *Plant Cell*, *9*(6), 859–868. https://doi.org/10.1105/tpc.9.6.859

Rabenstein, F., Schubert, J., Seifers, D. L., Stenger, D. C., & French, R. (2002). Phylogenetic relationships, strain diversity and biogeography of tritimoviruses. *J Gen Virol*, *83*(4), 895–906. https://doi.org/10.1099/0022-1317-83-4-895

Roberts, R., Zhang, J., Mayberry, L. K., Tatineni, S., Browning, K. S., & Rakotondrafara, A. M. (2015). A Unique 5′ Translation Element Discovered in Triticum Mosaic Virus. *J Virol*, *89*(24), 12427–12440. https://doi.org/10.1128/jvi.02099-15

Robinson, M. D., & Murray, T. D. (2013). Genetic Variation of Wheat streak mosaic virus in the United States Pacific Northwest. *Phytopathology*, *103*(1), 98–104. https://doi.org/10.1094/phyto-05-12-0108-r

Roossinck, M. J. (1997). Mechanisms of plant virus evolution. *Annu Rev Phytopathol*, *35*, 191–209. https://doi.org/10.1146/annurev.phyto.35.1.191

Roossinck, M. J., Martin, D. P., & Roumagnac, P. (2015). Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathol*, *105*(6), 716–727. https://doi.org/10.1094/phyto-12-14-0356-rvw

Seifers, D. L., Harvey, T. L., Louie, R., Gordon, D. T., & Martin, T. J. (2002). Differential Transmission of Isolates of the High Plains virus by Different Sources of Wheat Curl Mites. *Plant Dis*, *86*(2), 138–142. https://doi.org/10.1094/pdis.2002.86.2.138

Seifers, D. L., Harvey, T. L., Martin, T. J., & Jensen, S. G. (1997). Identification of the Wheat Curl Mite as the Vector of the High Plains Virus of Corn and Wheat. *Plant Dis*, *81*(10), 1161–1166. https://doi.org/10.1094/pdis.1997.81.10.1161

Seifers, D. L., Martin, T. J., Harvey, T. L., & Haber, S. (2007). Temperature-Sensitive Wheat streak mosaic virus Resistance Identified in KS03HW12 Wheat. *Plant Dis*, *91*(8), 1029–1033. https://doi.org/10.1094/pdis-91-8-1029

Seifers, D. L., Martin, T. J., Harvey, T. L., Fellers, J. P., & Michaud, J. P. (2009). Identification of the Wheat Curl Mite as the Vector of Triticum mosaic virus. *Plant Dis*, *93*(1), 25–29. https://doi.org/10.1094/pdis-93-1-0025

Seifers, D. L., Martin, T. J., Harvey, T. L., Fellers, J. P., Stack, J. P., Ryba-White, M., Haber, S., Krokhin, O., Spicer, V., Lovat, N., Yamchuk, A., & Standing, K. G. (2008). Triticum mosaic virus: A New Virus Isolated from Wheat in Kansas. *Plant Dis*, *92*(5), 808–817. https://doi.org/10.1094/pdis-92-5-0808

Seifers, D. L., Martin, T. J., Harvey, T. L., Haber, S., & Haley, S. D. (2006). Temperature Sensitivity and Efficacy of Wheat streak mosaic virus Resistance Derived from CO960293 Wheat. *Plant Dis*, *90*(5), 623–628. https://doi.org/10.1094/pd-90-0623

Simon-Loriere, E., & Holmes, E. C. (2011). Why do RNA viruses recombine? *Nat Rev Microbiol*, *9*(8), 617–626. https://doi.org/10.1038/nrmicro2614

Singh, K., Jarošová, J., Fousek, J., Chen, H., & Kundu, J. K. (2020). Virome identification in wheat in the Czech Republic using small RNA deep sequencing. *J Integr Agr*, *19*(7), 1825–1833. https://doi.org/10.1016/s2095-3119(19)62805-4

Singh, K., Wegulo, S. N., Skoracka, A., & Kundu, J. K. (2018). Wheat streak mosaic virus: a century old virus with rising importance worldwide. *Mol Plant Pathol*, *19*(9), 2193–2206. https://doi.org/10.1111/mpp.12683

Slykhuis J.T. (1955). Aceria tulipae Keifer (Acarina: Erophyidae) in relation to the spread of wheat streak mosaic. *Phytopathol, 45*, 116–128.

Stenger, D. C., & French, R. (2009). Wheat streak mosaic virus genotypes introduced to Argentina are closely related to isolates from the American Pacific Northwest and Australia. *Arch Virol*, *154*(2), 331–336. https://doi.org/10.1007/s00705-008-0297-1

Stenger, D. C., Hall, J. S., Choi, I.-R., & French, R. (1998). Phylogenetic Relationships Within the Family Potyviridae: Wheat Streak Mosaic Virus and Brome Streak Mosaic Virus Are Not Members of the Genus Rymovirus. *Phytopathol*, *88*(8), 782–787. https://doi.org/10.1094/phyto.1998.88.8.782

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. *J Virol*, *79*(14), 9054–9061. https://doi.org/10.1128/jvi.79.14.9054-9061.2005

Stenger, D. C., Seifers, D. L., & French, R. (2002). Patterns of Polymorphism in Wheat streak mosaic virus: Sequence Space Explored by a Clade of Closely Related Viral Genotypes Rivals That between the Most Divergent Strains. *Virol*, *302*(1), 58–70. https://doi.org/10.1006/viro.2001.1569

Stewart, L. R. (2016). Sequence diversity of wheat mosaic virus isolates. *Virus Res*, *213*, 299–303. https://doi.org/10.1016/j.virusres.2015.11.013

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. *Mol Plant Microbe In*, *27*(2), 150–162. https://doi.org/10.1094/mpmi-09-13-0272-r

Tatineni, S., & French, R. (2016). The Coat Protein and NIa Protease of Two Potyviridae Family Members Independently Confer Superinfection Exclusion. *J Virol*, *90*(23), 10886–10905. https://doi.org/10.1128/jvi.01697-16

Tatineni, S., Kovacs, F., & French, R. (2014a). Wheat streak mosaic virus Infects Systemically despite Extensive Coat Protein Deletions: Identification of Virion Assembly and Cell-to-Cell Movement Determinants. *J Virol*, *88*(2), 1366–1380. https://doi.org/10.1128/jvi.02737-13

Tatineni, S., Mcmechan, A. J., Wosula, E. N., Wegulo, S. N., Graybosch, R. A., French, R., & Hein, G. L. (2014b). An Eriophyid Mite-Transmitted Plant Virus Contains Eight Genomic RNA Segments with Unusual Heterogeneity in the Nucleocapsid Protein. *J Virol*, *88*(20), 11834–11845. https://doi.org/10.1128/jvi.01901-14

Tatineni, S., Qu, F., Li, R., Jack Morris, T., & French, R. (2012). Triticum mosaic poacevirus enlists P1 rather than HC-Pro to suppress RNA silencing-mediated host defense. *Virol*, *433*(1), 104–115. https://doi.org/10.1016/j.virol.2012.07.016

Tatineni, S., Van Winkle, D. H., & French, R. (2011). The N-Terminal Region of Wheat Streak Mosaic Virus Coat Protein Is a Host- and Strain-Specific Long-Distance Transport Factor. *J Virol*, *85*(4), 1718–1731. https://doi.org/10.1128/jvi.02044-10

Tatineni, S., Ziems, A. D., Wegulo, S. N., & French, R. (2009). Triticum mosaic virus: A Distinct Member of the Family Potyviridae with an Unusually Long Leader Sequence. *Phytopathol*, *99*(8), 943–950. https://doi.org/10.1094/phyto-99-8-0943

Urcuqui-Inchima, S., Haenni, A.-L., & Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Res, 74(1-2), 157–175*. doi:10.1016/s0168-1702(01)00220-9

Vance, V.B. (1991). Replication of potato virus X RNA is altered in coinfections with potato virus Y. *Virol, 182*, 486-494.

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 Prot, 29*, 699-703. https://doi.org/10.1016/j.cropro.2010.02.005

Webb, C.A. (2018). Epidemiological factors impacting the development of Wheat streak mosaic virus outbreaks. MSc. Thesis.

Worobey, M., & Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J Gen Virol*, *80*(10), 2535–2543. https://doi.org/10.1099/0022-1317-80-10-2535

Wosula, E. N., Mcmechan, A. J., Oliveira-Hofman, C., Wegulo, S. N., & Hein, G. L. (2016). Differential Transmission of Two Isolates of Wheat streak mosaic virus by Five Wheat Curl Mite Populations. *Plant Dis*, *100*(1), 154–158. https://doi.org/10.1094/pdis-03-15-0342-re

Wu, B., Blanchard-Letort, A., Liu, Y., Zhou, G., Wang, X., & Elena, S. F. (2011). Dynamics of Molecular Evolution and Phylogeography of Barley yellow dwarf virus-PAV. *PLOS ONE*, *6*(2), e16896. https://doi.org/10.1371/journal.pone.0016896

Young, B. A., Stenger, D. C., Qu, F., Morris, T. J., Tatineni, S., & French, R. (2012). Tritimovirus P1 functions as a suppressor of RNA silencing and an enhancer of disease symptoms. *Virus Res*, *163*(2), 672–677. https://doi.org/10.1016/j.virusres.2011.12.019

Chapter 2 - The genetic variation of field wheat streak mosaic-associated viruses

Abstract

Wheat streak mosaic (WSM), a viral disease affecting cereals and grasses, costs substantial losses in crop yields. *Wheat streak mosaic virus* (WSMV) is documented as the main causal agent of the complex, but mixed infections with *Triticum mosaic virus (*TriMV) and *High plains wheat mosaic emaravirus* (HPWMoV) have reported as well. Although current resistant varieties are effective for the disease control, WSMV resistant-breaking isolates have been reported, suggesting that viral populations are genetically diverse. Previous phylogenetic studies of WSMV were mostly focused only on the virus coat protein sequence; while, there is no such study for either TriMV or HPWMoV. Here, we studied the genetic variation and evolutionary mechanisms of WSM-associated viruses in Kansas fields using the complete genome sequences. In total, 29 historic and field samples were used for total RNA deep sequencing (RNA-seq) to obtain the full genome sequences of WSM-associated viruses. Results of the survey showed WSMV as the predominant virus followed by mixed infections of WSMV+TriMV in fields. Phylogenetic analyses based on the obtained full genome sequences demonstrated that, unlike other isolates from the United States, Kansas isolates are widely distributed in sub-clades suggesting the presence of enough genetic variation among KS isolates. In contrast, the same phylogenetic analyses but for TriMV isolates did not show significant diversity probably due to the lack of enough available virus full reference genome sequences in GenBank. Moreover, the phylogenetic studies suggested that TriMV field isolates may be under selection pressure to introduce genetic variations due to the use of resistant varieties in the fields. In addition, recombination was identified as the major evolutionary force of WSMV in KS fields but not for TriMV.

Furthermore, the full genome sequence of a new Kansas HPWMoV isolate was reported here. The findings of this study will provide more insights into the genetic structure of WSM-associated viruses and in turn, help in improving strategies for disease management.

Key words: Wheat streak mosaic disease, WSMV, TriMV, HPWMoV, RNA-seq, Phylogeny

Introduction

Wheat (*Triticum aestivum L.*) is one of leading staple crops in the world. In 2019, the wheat production in Kansas estimated by the United State Department of Agriculture's National Agricultural Statistics Service (USDA NASS) was $1.37 billion which amounted to almost 350 million total bushels produced (NASS 2020). Kansas is the second national leading producer of wheat behind North Dakota (NASS 2020). Viral diseases have a great impact on reducing the yield of wheat globally. In 2017, a viral disease called Wheat Streak Mosaic (WSM) has caused a 5.6% in total yield loss which is equal to $76 million to Kansas farmers (Hollandbeck et al. 2017).

WSM is a viral disease which poses a threat to the wheat industry. The disease complex consists of three documented viruses: *Wheat streak mosaic virus* (WSMV), *Triticum mosaic virus* (TriMV), and *High plains wheat mosaic emaravirus* (HPWMoV), which are all transmitted by wheat curl mites (WCM), *Aceria tosichella* Kiefer (Tatineni and Hein 2018). The typical symptoms of WSM caused by any of the three viruses in single or mixed-infections include yellow, mosaic-like streaks on the leaves (Figure 2-1A) which lead to chlorosis and reduction in photosynthetic capabilities. Severe viral symptoms may also lead to stunted growth (Figures 2-1B to 2-1C) (Singh et al. 2018). For this reason, it is difficult to differentiate the causal virus phenotypically from symptoms, and serological and molecular techniques such as ELISA and RT-PCR are required to determine which virus or mixed-infection of viruses are present.

Compared to the other two viruses of the WSM complex, WSMV is the more widely studied and it has a longer history, with its first observation dating back to 1922 in Nebraska (McKinney 1937). WSMV belongs to the family *Potyviridae* and the genus *Tritimovirus* (Rabenstein et al. 2002). The genome of WSMV is approximately ~9.3 kb in size and encodes

a large polyprotein, which is enzymatically cleaved and forms 10 mature proteins: P1, HC-Pro (helper component protein), P3, 6K1, CI (cytoplasmic inclusion protein), 6K2, NIa-Pro (nuclear inclusion putative protease), NIa-VPg (viral protein genome-linked proteinase), and CP (coat protein) (Choi et al. 2002; Chung et al. 2008; Singh et al. 2018). The 5' and the 3' terminals contain a VPg and a Poly (A) tail, respectively (Singh et al. 2018). In comparison to other potyviruses, the P1 of WSMV acts as an enhancer of pathogenicity and a suppressor of RNA silencing (Tatineni et al. 2012; Young et al. 2012). The HC-Pro has been discovered to be required for vector transmission and one of the proteolytic processors of WSMV (Stenger et al. 2005). The P3 function is less understood, but observed to bind to P1, HC-Pro, and CI for WSMV, suggesting that this protein may play a role in the viral replication and movement (Choi et al. 2000; Choi et al. 2005). P3N-PIPO, a small overlapping protein typically found in the P3 region of potyviruses, was also discovered to be translated from the P3 region of WSMV and has been found to be essential in cell-to-cell movement and internal RNA (Chung et al. 2008). 6K1 and 6K2 functions are not well studied for WSMV, but previous studies of other potyviruses propose that these proteins are important in the formation of viral vesicles and play essential roles in the replication and pathogenicity (Nigam et al. 2019; Cui and Wang 2016). The CP is multi-functional and is utilized for host- and strain- specific long-distance transport, virus assembly, and cell-to-cell movement (Tatineni and French 2014, Tatineni et al. 2011, 2014a). The NIa-Pro is a major proteinase which aids in processing the polyprotein into 10 individual proteins (Singh et al. 2018). The NIb is presumed to be the replicase due to the presence of conserved polymerase motifs within this protein (Choi et al. 2000). The NIa-VPg binds with the eukaryotic translation initiation factor 4E (eIF4E), which initiates the translation of the viral genome and has also been found to suppress the production of the host proteins (Urcuqui-Inchima et al. 2001; Adams et al. 2005).

Previous phylogenetic study based on the coat protein sequences of WSMV isolates divided them into four different clades (clades A-D) (Stenger et al. 2002). Clade A contains an isolate from Mexico, Clade B contains isolates from Europe, Clade C contains an individual isolate from Iran, and Clade D contains isolates from USA, Argentina, and Turkey (Stenger et al. 2002). Clade D was further divided into 4 sub-clades: sub-clade D1 contains isolates from Argentina and the American Pacific Northwest, sub-clade D2 contains isolates from Kansas and Colorado, sub-clade D3 includes isolates from Kansas, Kentucky and Missouri, and sub-clade D4 consists of isolates from Kansas and Nebraska (Stenger et al. 2002). This result showed diversity among the U.S. isolates in general, and isolates from Kansas in particular. However, only the sequence of a small portion of the genome (coat protein) was compared in this study.

TriMV, a previously unknown wheat virus was first discovered in Western Kansas in 2006 and its association with WSM was reported (Seifers et al. 2008). TriMV belongs to the family *Potyviridae* like WSMV but different genus, *Poacevirus*. The genome size of TriMV is ~10.2 kb and similar to WSMV, encodes a polyprotein which is cleaved into 10 mature proteins enzymatically: P1, HC-Pro (helper component protein), P3, 6K1, CI (cytoplasmic inclusion protein), 6K2, NIa-Pro (nuclear inclusion putative protease), NIa-VPg (viral protein genome-linked proteinase), NIb (replicase), and CP (coat protein) (Fellers et al. 2009). In contrast to WSMV, TriMV has an unusually long 5' UTR spanning to 739 nt which has been implicated as a translation enhancing element (Fellers et al. 2009; Tatineni et al. 2009; Young et al. 2012). Similar to WSMV, the P1 of TriMV has been observed to suppress the RNA silencing host defense (Tatineni et al. 2012). In comparison to WSMV, the specific function of the other proteins of TriMV are not well studied and the protein functions are presumed to be similar to other potyviruses. There are currently no studies investigating the phylogeny and genetic variation of TriMV.

HPWMoV, the other documented virus associated with WSM belongs to the *Fimoviridae* family and the genus *Emaravirus*. HPWMoV is a multipartite, single-stranded, negative-sense RNA virus (Stewart 2016) first described in1993 as high plains disease infecting maize and wheat in the Great Plains (Jensen et al. 1996). However, the genome sequence and the genome organization of HPWMoV was not determined until recently (Tatineni et al. 2014). HPWMoV consists of 8 negative-sense RNA segments, designated as RNAs 1 to 8 (Tatineni et al. 2014). The encoded proteins are annotated as follows: RNA 1 is the RNA-dependent RNA polymerase (RdRp), RNA 2 is the putative glycoprotein, RNA 3 as the nucleocapsid protein, RNA 4 as the putative movement protein, and RNAs 7 and 8 act as the RNA silencing suppressor (Gupta et al. 2018). RNAs 5 and 6 currently do not have any known functions (Gupta et al. 2018). As TriMV, there are limited full genome sequences of HPWMoV available in the database and the current knowledge about the phylogenetic relationships among HPWMoV isolates is limited.

Currently, there are 3 types of resistant genes available against WSM including *Wsm2* against WSMV and *Wsm1* and *Wsm3* against WSMV and TriMV (Lu et al. 2011; Friebe et al. 1991; Liu et al. 2011). Varieties contain *Wsm3* are currently not deployed in fields, but *Wsm1* and *Wsm2* have been used by farmers. In 2019, a WSMV resistant-breaking isolate has been reported and confirmed to overcome the *Wsm2* in Kansas fields and potential resistant-breaking isolates for WSMV and TriMV have also been found to infect resistant varieties in the field (Kumssa et al. 2019; Fellers et al. 2019). These events place a greater importance in understanding the current genetic structure of viral populations associated with WSM in the field in order to assess the distribution of the associated viruses and determine the major evolutionary forces acting upon these viruses.

In this study, we determined the current distribution and prevalence of WSM-associated viruses in Kansas fields and assessed the source of genetic variation of the WSM viruses.

Additionally, and for the first time, the phylogenetic relationships among WSMV isolates was investigated based on the full genome sequences of historic and field isolates. We also showed the first phylogenetic analysis of TriMV isolates using the whole genome sequence in this study. Although there was not enough complete genome sequences in the GenBank to conduct a phylogenetic analysis for HPWMoV, complete sequences of 8 RNA segments of a new HPWMoV isolate from Kansas was extracted. With the introduction of resistant varieties in the field, evolutionary forces have been acting upon viral populations of WSM, which makes it essential to understand the genetic variation and structure of WSM-associated viruses in Kansas fields in order to determine the efficacy of these resistant varieties, and create more durable disease management strategies.

Materials and methods

Sample collection

In 2019, symptomatic and asymptomatic wheat leaf samples were collected through field surveys and sample submissions to the Kansas State University Plant Disease Diagnostics Lab. In addition to these samples, historic WSMV samples were provided by the Agricultural Research Center in Hays, KS. A few wheat samples from Nebraska and Colorado were also received and included in the study.

Screening samples for WSM-associated viruses

Total RNAs were isolated from leaf tissues using TRIzol reagent (Invitrogen, CA, USA), according to the manufacturer's instruction. The extracted RNAs from the selected samples were treated with DNase I (Zymo Research, CA, USA). The first strand cDNAs were synthesized using the SuperScript II Reverse Transcriptase (Invitrogen, CA, USA). OligodT and gene specific primers (Table S2-1) were designed for the PCR step for each virus.

PrimeSTAR GXL Premix (Takara Bio, CA, USA) was used to carry out a 25uL reaction containing: 1X PrimeSTAR GXL Buffer, nuclease-free water, and 0.5 uM each of the gene specific primers. The thermal cycler program used is as follows: 98℃ for 2 minutes, 34 cycles of 98℃ for 10 seconds, 55℃ for 15 seconds, and 68℃ for 2 mins, and 72℃ for 5 minutes.

For HPWMoV, one step RT-PCR was conducted to screen the samples. For the one step RT-PCR, the 28 uL reactions contained 1x GoTaq Flexi Buffer (Promega, WI, USA), 1 uM MgCl2, 0.1 uM dNTP, 0.4 uM of gene specific primers (Table S2-1), 1.25 U GoTaq Flexi DNA Polymerase (Promega, WI, USA), 200 U of the SuperScript IV Reverse Transcriptase (Invitrogen, CA, USA), 40 U RNaseOUT, and DEPC treated water. The thermal cycler program used is as follows: 42℃ for 10 minutes, 94℃ for 2 minutes 34 cycles of 94℃ for 10 seconds, 55℃ for 1 minute, and 72℃ for 2 minutes, and 72℃ for 5 minutes. The RT-PCR products were visualized using 2% agarose gels stained with SYBR Safe (Invitrogen, CA, USA).

Sample selection for RNA deep sequencing

22 samples were chosen for sequencing based on the results of the virus screenings and the geographic region with mixed infections being prioritized. Five historic samples and one sample each from Colorado and Nebraska were also selected for library preparation.

RNA library construction and sequencing

The RNA integrity of the DNase treated samples were measured using Qubit 4 (Invitrogen, Ca, USA) with the RNA IQ assay kit, according to the manufacturer's instruction. The quantification of the RNA was carried out using the NanoDrop Spectrophotometer (Invitrogen, CA, USA). The TruSeq Stranded Total RNA with Ribo-Zero Plant kit (Illumina Inc., CA, USA) was utilized to remove the rRNA and prepare the libraries for sequencing, following the manufacturer's instruction. Agencourt RNAClean XP (Beckman Coulter, MA,

USA) was used to purify the samples ensure the removal of all traces of rRNA. TruSeq RNA Single Indexes Sets A and B (Illumina Inc., CA, USA) were used for adapter ligation. After each step of cDNA synthesis, adapter ligation, and enrichment of the DNA fragments, the samples were purified using the Agencourt AMPure XP (Beckman Coulter, MA, USA).

The final libraries were subjected to quality control analysis using Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) and were quantified using the Qubit 4 (Invitrogen, CA, USA) with the 1X dsDNA High Sensitivity Assay (Invitrogen, CA, USA). At the Kansas State Integrated Genomics Facility, the libraries were pooled and sequenced using the NextSeq 500 (Illumina Inc., CA, USA) high-output with a read length of 1x75bp.

Bioinformatics analysis

Libraries were demultiplexed based on the index sequences. Trimmomatic was used to trim the reads for quality, length and the adapter sequences (Bolger et al. 2014). To ensure the reads no longer contained adapter sequences and were of high quality, FastQC was utilized for quality control (Andrews 2010). The reference genomes of WSMV, TriMV, and HPWMoV (Table S2-2) were retrieved from GenBank (www.ncbi.nih.gov/nuccore). The trimmed reads were mapped against the reference genomes and the consensus sequences were extracted using the CLC Genomics Workbench 20 (Qiagen, MD, USA).

Recombination analysis

Multiple nucleotide alignments of the consensus sequences from this study and the complete reference genome sequences obtained from the GenBank (Table S2-3) were conducted using the MUSCLE alignment in the Geneious Prime 2020.2.4 (Edgar 2004; www.geneious.com). The complete genome sequence alignments were then examined using 7 different algorithms integrated in the RDP5 program (Martin et al. 2015). The 7 algorithms used are as follows: RDP (Martin et al. 2000), GENECONV (Padidam et al. 1999), MaxChi

(Maynard Smith 1992), BootScan (Martin et al. 2005), Chimaera (Posada et al. 2001), 3SEQ (Lam et al. 2018), and SiScan (Gibbs et al. 2000). The recombination events which were significant ($p<0.01$) for at least four of the seven detection methods were considered potential recombinants and potential parents were determined.

The results from RDP5 were utilized to run Bootscan (Salminen et al. 1995) analysis in the SimPlot program (Lole et al. 1999), in order to verify the recombination events. The default settings for window width of 200 and step size of 20 was used. The cut off value for the percent of permuted trees to accept the sample as a potential recombinant was set at 70%. In addition to Bootscan, SimPlot analysis (Lole et al. 1999) was also utilized using the same default parameters as Bootscan in order to determine the recombination breakpoints.

Substitution models and phylogenetic analysis

The detected recombinants were removed and outgroups were added before realignment with MUSCLE (Edgar 2004). The best fitting nucleotide substitution models were determined by the jModelTest 2 (Darida et al. 2012; Guindon et al. 2003). The nucleotide substitution models selected by both the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to construct the phylogenetic trees were GTR + I + G for WSMV and GTR + I was selected for TriMV (Darida et al. 2012; Guindon et al. 2003). For the phylogenetic analysis, mrBayes plugin within the Geneious Prime 2020.2.4 program was used to construct Bayesian consensus phylogenetic trees using the default parameters (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003).

Calculation of population genetics parameters

To conduct the population genetics analyses, only the complete genome sequences of Kansas isolates were utilized. 16 isolates for WSMV and 8 isolates for TriMV were analyzed

using the DnaSP version 6 (Rozas et al. 2003) to calculate the population genetics parameters and genetic diversity (Table S2-4).

Neutrality tests

The estimation of non-synonymous substitutions (dN), synonymous substitutions (dS) and their ratio (dN/dS= ω) was calculated in MEGA 5 by using the bootstrap method with 1000 replicates under the model of the Kumar method for each encoded protein (Tamura et al. 2011; Kumar et al. 2004). Using Hyphy 2.2.4 (Kosakovsky Pond et al. 2005), the stop codons for the complete polyprotein alignments of the Kansas isolates were removed prior to neutrality tests. To evaluate the selection pressure by site of specific codons, three different methods which are implemented in the Hyphy package were used (Kosakosky-Pond et al. 2005). Fixed effects likelihood (FEL) and single likelihood ancestor counting (SLAC) utilize the maximum-likelihood (ML) methods to analyze site specific selection pressures of the polyprotein (Kosakosky-Pond & Frost 2005). In addition to the ML methods, a Bayesian approach using fast, unconstrained Bayesian approximation (FUBAR) was also applied (Murrell et al. 2013). The default cut-off P-value set by Hyphy of 0.1 for SLAC and FEL and 0.9 of Bayes Factor for FUBAR were utilized to determine the significance of the results. The codons determined to be significant by at least two methods were accepted as the sites under positive or negative selections.

Results

WSM distribution in Kansas

In total, 84 leaf samples were screened for WSM-associated viruses (Figure 2-2) by RT-PCR. Single infection of WSMV dominated Kansas fields at 52%, followed by 8% of mixed-infections of WSMV + TriMV, and a mixed-infection of WSMV + TriMV +HPWMoV

was detected in only one sample (Figure 2-3). No single TriMV or HPWMoV infections were detected.

RNASeq analysis

Over 95% of the complete genome sequences of 15 WSMV, 7 TriMV, and one HPWMoV from field collected samples and 5 historical WSMV samples were obtained by mapping clean reads into the reference genome of each virus (Table S2-3). Complete genome sequences were used for the rest of analyses (Table S2-4).

Recombination analysis

WSMV

The WSMV recombination analysis consisted of 15 field samples collected in 2019, 5 historical samples (Table S2-4), and 24 complete sequences (Table S2-3) retrieved from NCBI. In total, 7 potential recombinants were identified for WSMV: DC19, KM19, KSH294, EL17-1183, NE01-19, KstIct2017, and NS02-19 isolates detected by at least five recombination detection methods with a significant support ($p<0.05$) (Table S2-5). The detection methods also identified the potential parents, which included isolates from Nebraska, Colorado, Czech Republic, Kansas, and France (Table S2-6). The recombination hotspots were detected in the regions of WSMV genome encoding P1, HC-Pro, CI, P3, and CP proteins.

TriMV

For TriMV, 7 field samples from 2019 (Table S2-4) and 5 samples (all available complete genome sequences) obtained from the NCBI GenBank (Table S2-3) were used. The RDP5 program found no significant recombination events among TriMV isolates.

Due to the low incidence of HPWMoV in this study and the insufficient number of HPWMoV complete genome sequences in the GenBank, HPWMoV was not included in the recombination and phylogenetic analysis.

To confirm the results obtained from the RDP5 program, the potential recombinants were utilized as query sequences in the Bootscan analysis using SimPlot. To run this analysis, the major and minor parents detected by the RDP5 and 2 randomly chosen WSMV isolates were used as reference sequences. All potential recombinants demonstrated high levels of phylogenetic relatedness with the reference sequences, exceeding the 70% cut-off value (Salminen et al. 1995). The BootScan analysis using SimPlot confirmed the RDP5 results and exhibited that there is evidence of recombination in the 2019 field isolates. In addition to confirming evidence of recombination, the BootScan analysis also demonstrated that all recombinants had more than one potential parents as depicted by the reference sequences surpassing the 70% threshold value.

Phylogenetic analysis

WSMV

The potential detected recombinants were removed from the phylogenetics studies. The complete genome sequences of 11 WSMV isolates from 2019 field surveys, 4 WSMV historical samples, and 22 isolates obtained from the Genbank were used for the phylogenetic analysis (Table S2-3 to S2-4). Two outgroups were chosen for building the phylogenetic tree: *Oat necrotic mottle virus* (ONMV) and *Yellow oat-grass mosaic virus* (YOgMV). The WSMV topology consists of 4 main clades: Clade A: isolates from Mexico, Clade B: European isolates, Clade C: an isolate from Iran, and Clade D: USA, Argentina, and Turkey isolates (Figure 2-5). Clade D was further divided into four sub-clades (D1-D4) with Kansas isolates widely distribution in all sub-clades (Figure 2-5).

Sub-clade D1 contained isolates from the Pacific Northwest and the Kansas Type isolate. Sub-cade D2 consisted of isolates from KS and Ohio, whereas sub-clade D3 included only Kansas isolates. Both sub-clades D2 and D3 contained isolates from the 2019 and historical Kansas isolates. Sub-clade D4 included isolates from Nebraska, Idaho, and Kansas.

TriMV

There were 12 TriMV isolates used to construct phylogenetic trees: 7 isolates were from this study and 5 isolates were from the GenBank (Table S2-3, S2-4). *Sugarcane mosaic virus* (ScSMV) and *Caladenia virus A* (CalVA) were used as outgroups for constructing the TriMV Bayesian tree (Table S2-3). The topology of TriMV tree consisted of three clades (Figure 2-6). Clade A included a single isolate from 2019 KS isolates collection. Clade B contained an isolate from Nebraska. Clade C consisted of one sub-clade (C1) and two polytomies including two isolates from 2019 KS field collection: DC19 and NS02_19. The sub-clade C1 contained Kansas isolates from the 2019 field survey (this study), one reference isolate from 2016, and 3 potential resistant-breaking reference isolates: KSGre2017, KSHm2015, and KSIct2017 (Fellers et al. 2019). Two of these potential resistant-breaking isolates are closely related and formed their own sister taxa.

Population genetics parameters

A total of 20 and 8 Kansas isolates were used to evaluate the genetic diversity of the WSMV and TriMV, respectively (Table S2-4). The population genetic parameters including the average nucleotide diversity (π) and the mutation rate per segregating sites (Θw) were calculated for both WSMV and TriMV using DnaSP 6 (Table 2-1 and Table 2-3). The genetic diversity for both WSMV and TriMV isolates were relatively low with a mean of 0.027 and 0.0038, respectively.

Neutrality test

The ratio of dN/dS for the individual proteins was < 1 for both WSMV and TriMV suggesting purifying (negative) selection as the main selection pressure acting upon encoded proteins of both viruses (Table S2-6 and Table S2-7). To assess the selection imposed on each site (codon) of the protein, three different algorisms were used. One positive selection site was

found in site 118 (D→N) located in the P1 protein of WSMV and at site 2677 (L→I) located at the NIb protein of TriMV, which were both supported significantly by two methods; FEL ($p<0.1$) and FUBAR with the Bayes posterior probability above 0.90 (Table 2-2 and Table 2-3).

Figure 2-1. Observed symptoms of WSM in the field. A. Typical viral symptoms of WSM is yellow, mosaic-like streaks on leaves. B. Severe symptoms of mixed infections of WSMV + TriMV include stunting, which leads to the underdevelopment and total loss of the crops. C. A close-up image of the stunted wheat infected with WSM viruses which is only twice the size of the dried corn cob used for comparison. The expected height of wheat in this ripening stage is 3 times higher.

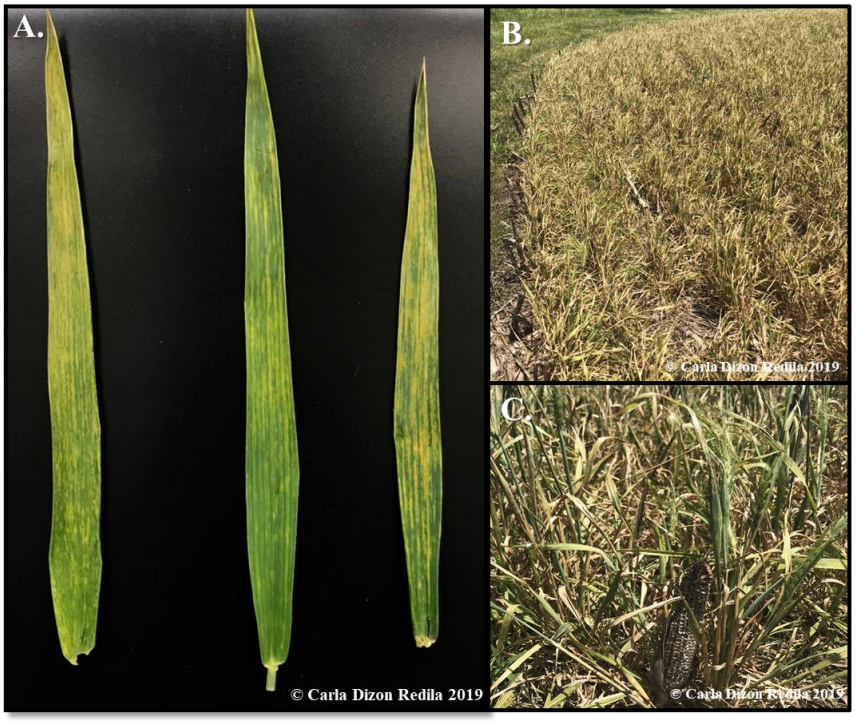


Figure 2-2.Collection sites of the wheat samples surveyed in 2019. A total of 84 symptomatic and asymptomatic leaf samples were screened for WSM-associated viruses. The survey covered 54 counties from central and western Kansas and 2 eastern counties.

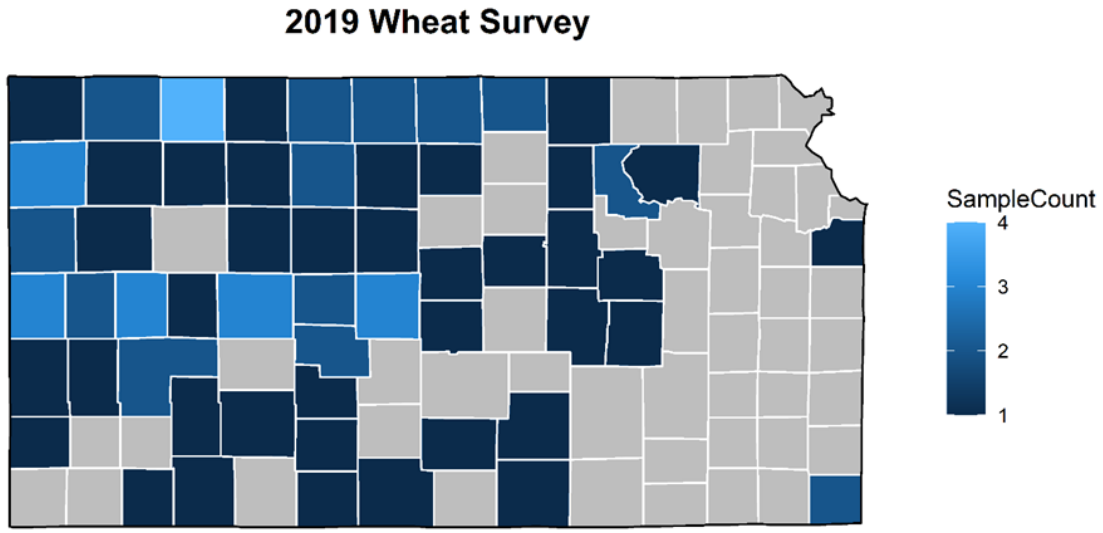


Figure 2-3. Distribution of WSM-associated viruses in Kansas. Single infections of WSMV is highlighted in different shades of blue with the lighter shades indicating more than one sample was infected with the virus. The mixed infections of WSMV+TriMV are depicted by the purple diamonds and the mixed infection of WSMV+TriMV+HPWMoV is shown in the purple star. Mixed infections were not detected in the southernmost counties.

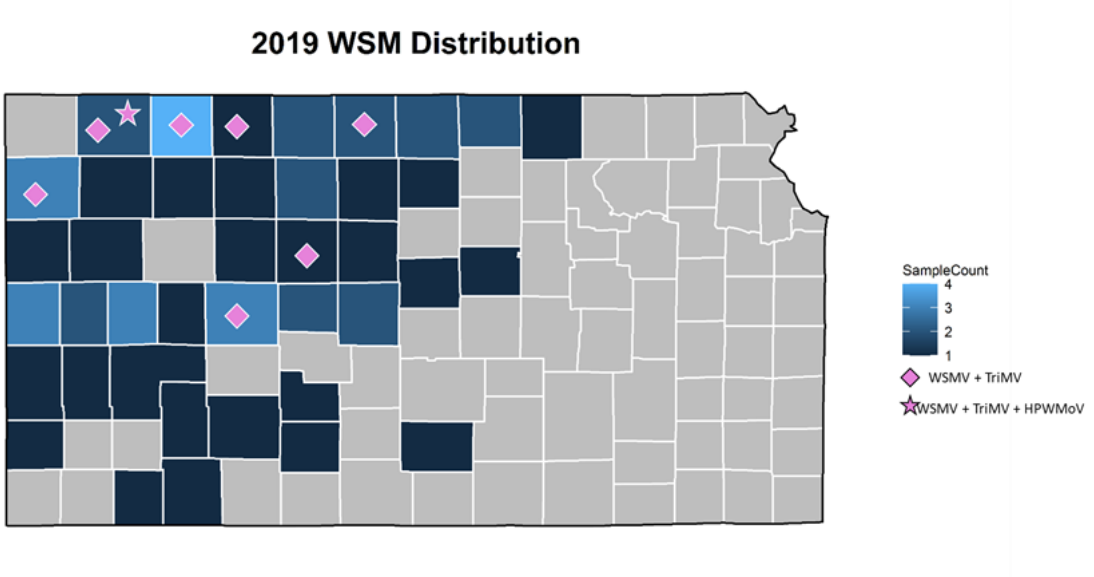
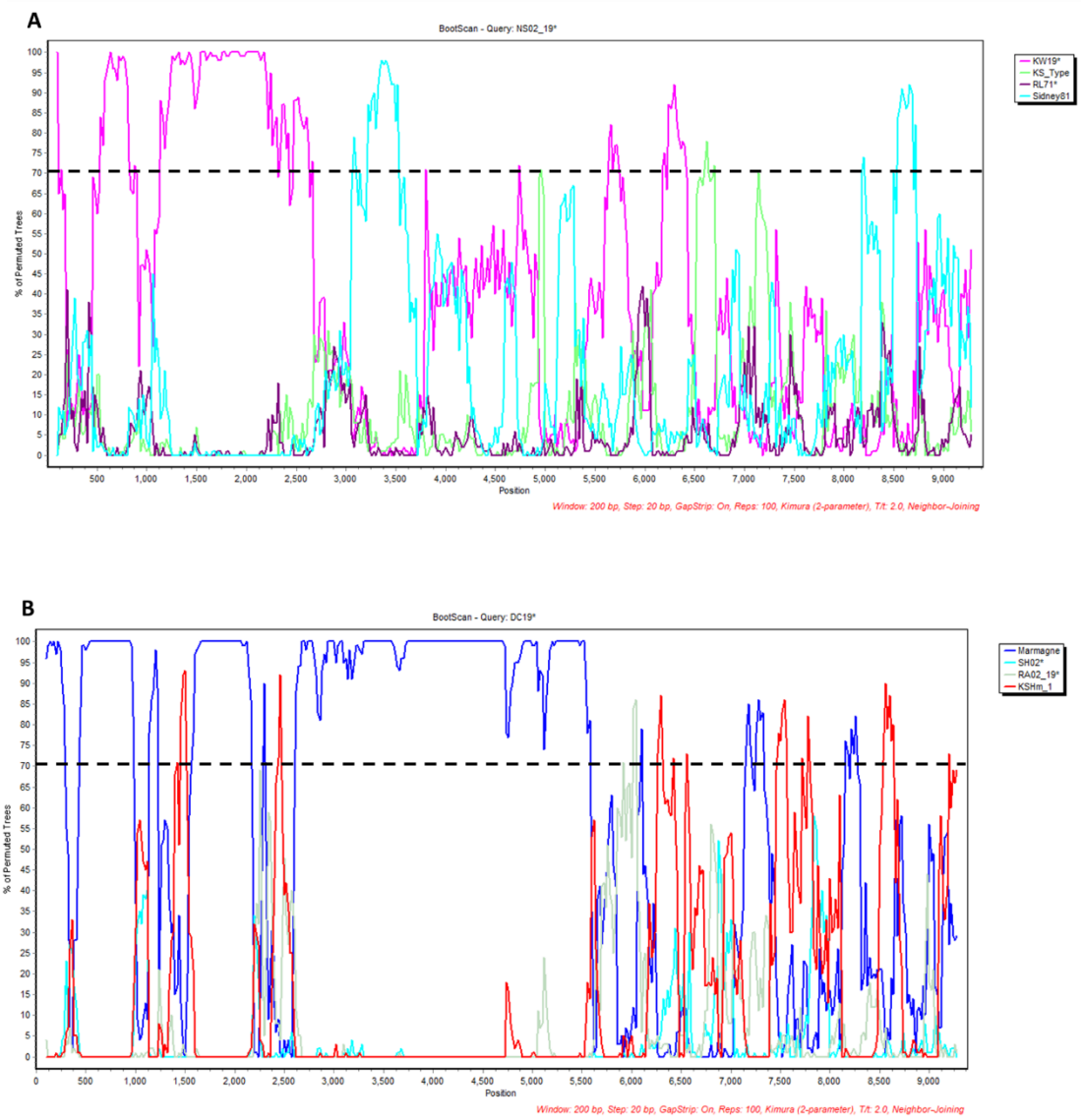
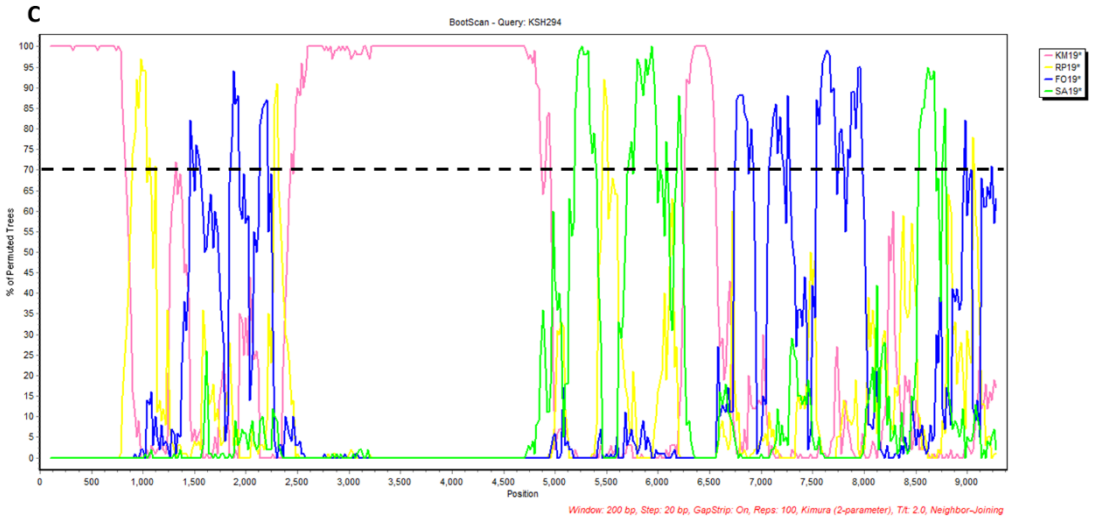
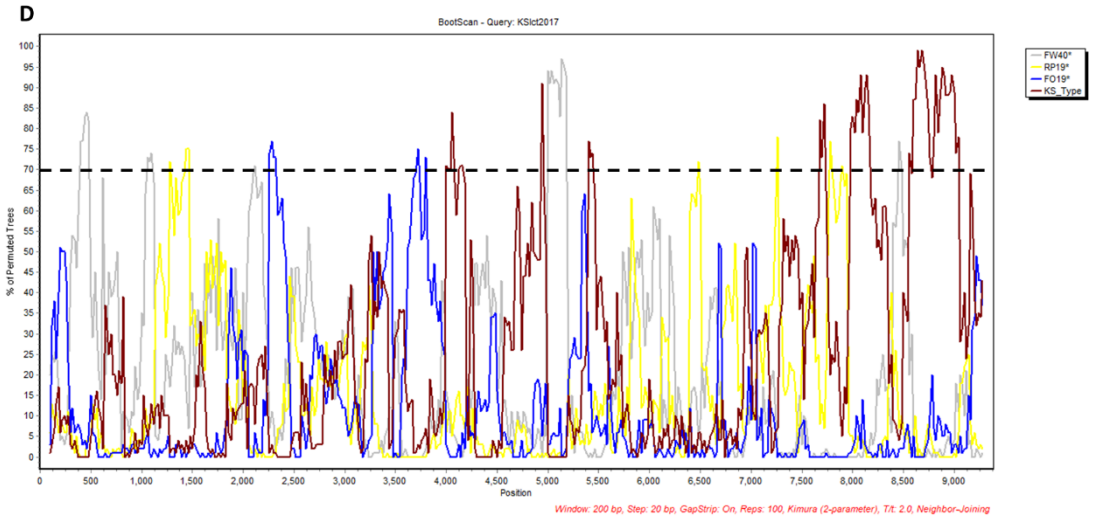


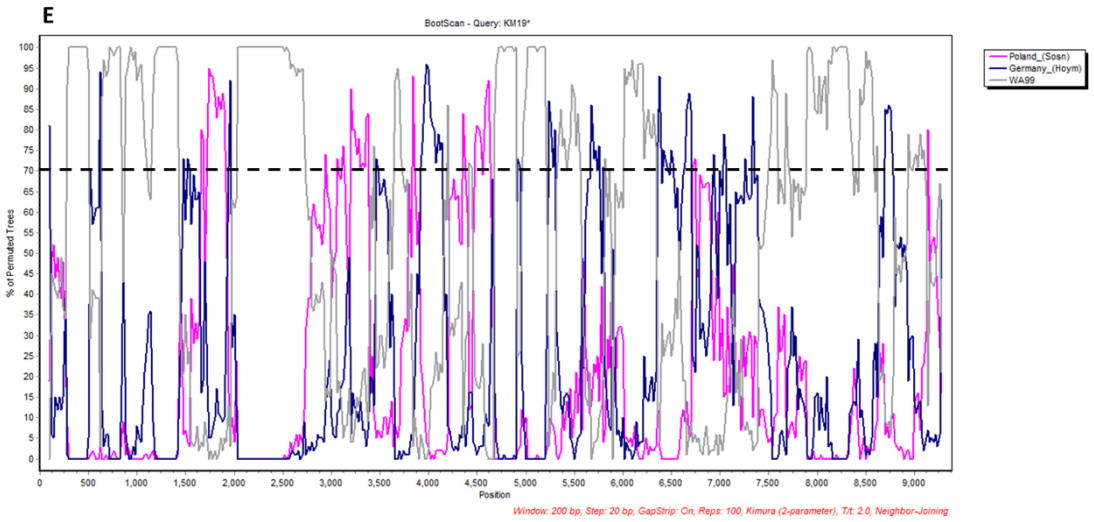
Figure 2-4. Recombination Analysis using Bootscan method. Bootscan results from Simplot using the default settings. The query sequences are NS02_19 (A), DC19 (B), KSH294 (C), KSIct2017 (D), KM19 (E), NE01_19 (F), and EL17-1183 (G). The reference sequences are found in the legend. The 70% permuted trees support is depicted by the dotted line and is the cut-off support to confirm the potential recombinants.



Figures 2-4 (Continued).

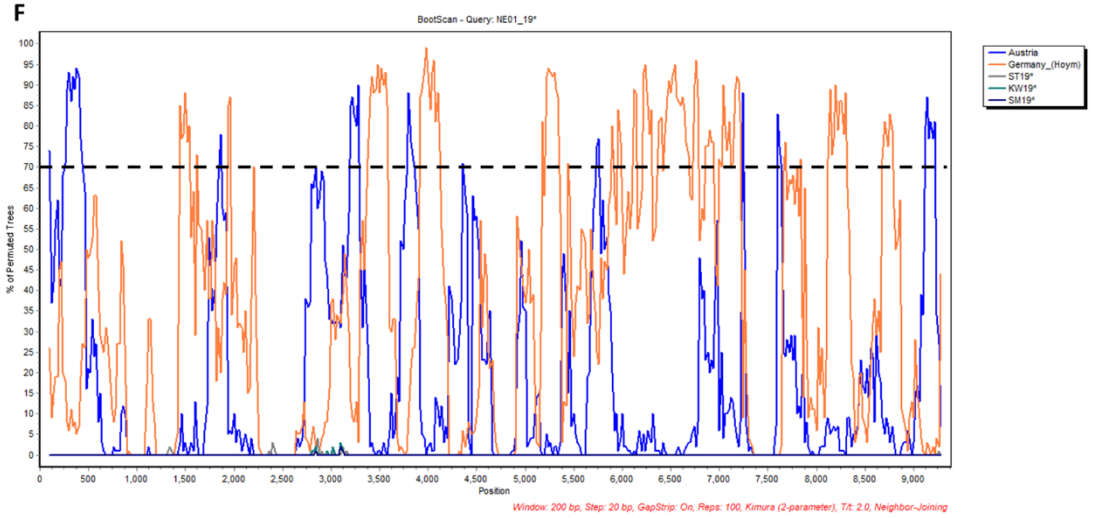






Figures 2-4 (Continued).

F



G

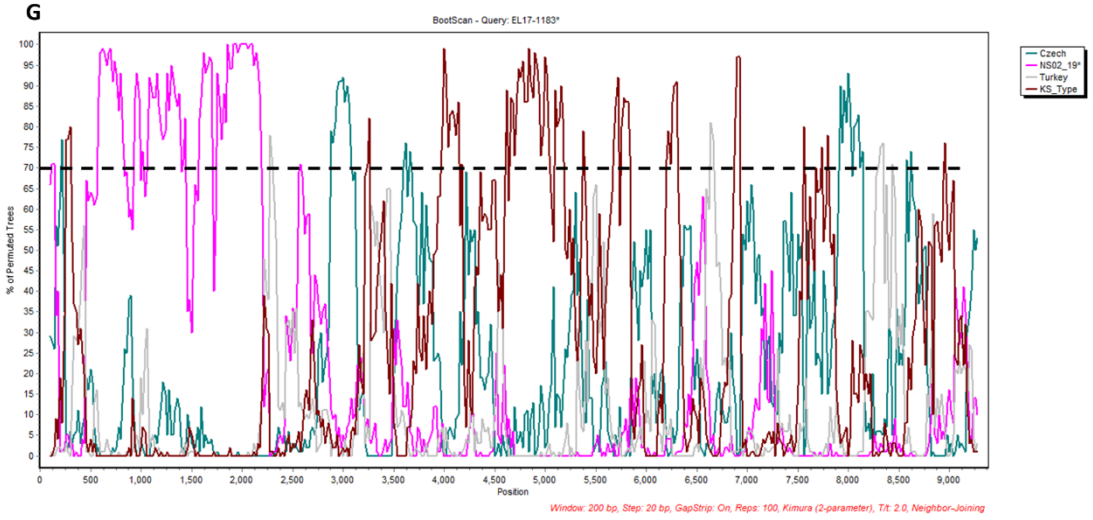


Figure 2-5. Phylogenetic tree of WSMV isolates. The Bayesian phylogenetic tree of the complete genome sequences of WSMV consisted of 15 field isolates, 5 historical isolates, and 17 reference isolates from the NCBI database. The sample IDs of Kansas isolates are in purple text and the isolates sequenced in this study are written in bold and italicized text. The posterior probability of 70% was the cut-off value and branches not supported were collapsed.

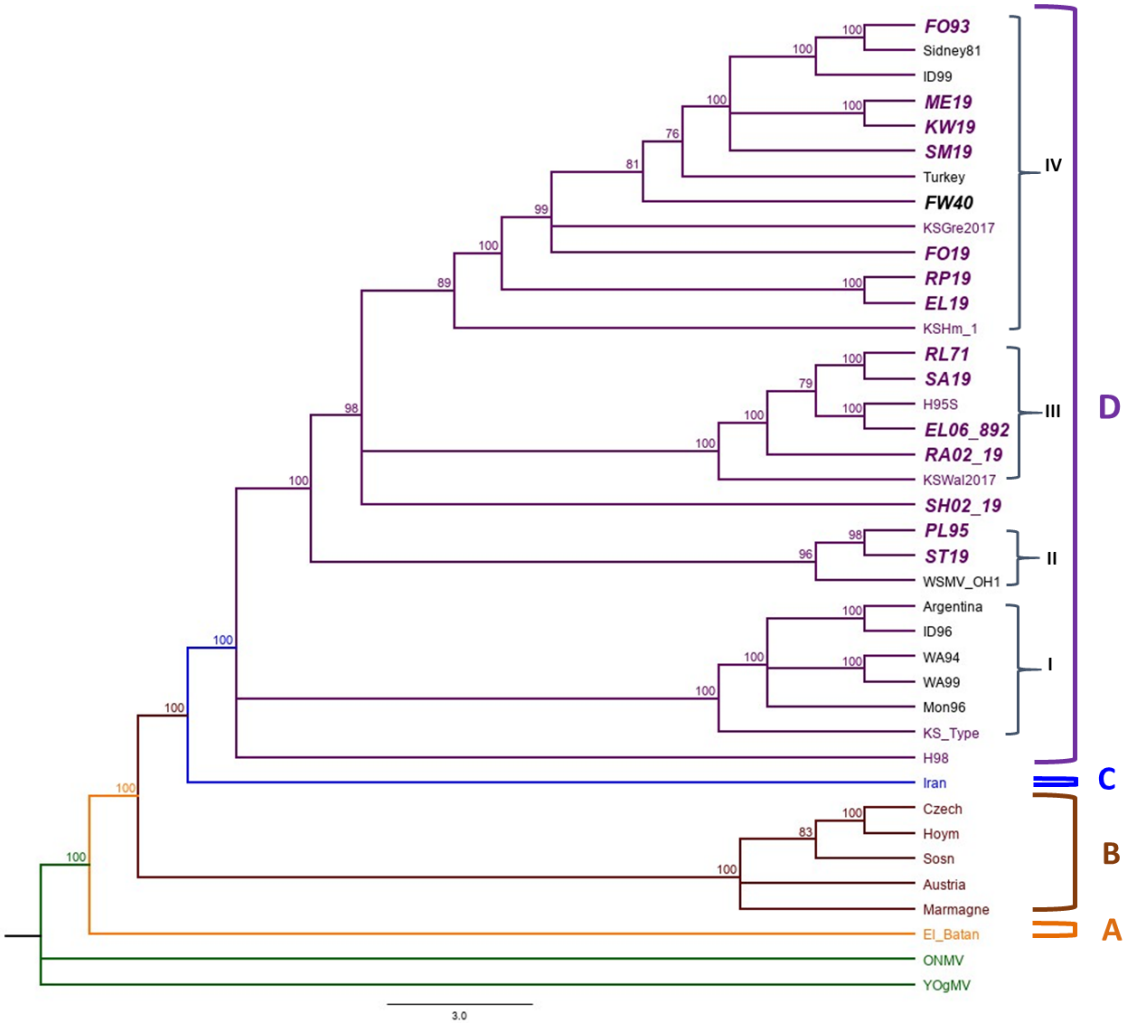


Figure 2-6. Phylogenetic tree of TriMV isolates. The Bayesian phylogenetic tree of the complete genome sequences of TriMV which consisted of 7 field isolates from 2019 and 5 isolates obtained from the GenBank. The sample IDs of Kansas isolates are written in purple text and the isolates sequenced in this study are written in bold and italicized text. The posterior probability of 70% was the cut-off value and branches not supported were collapsed.

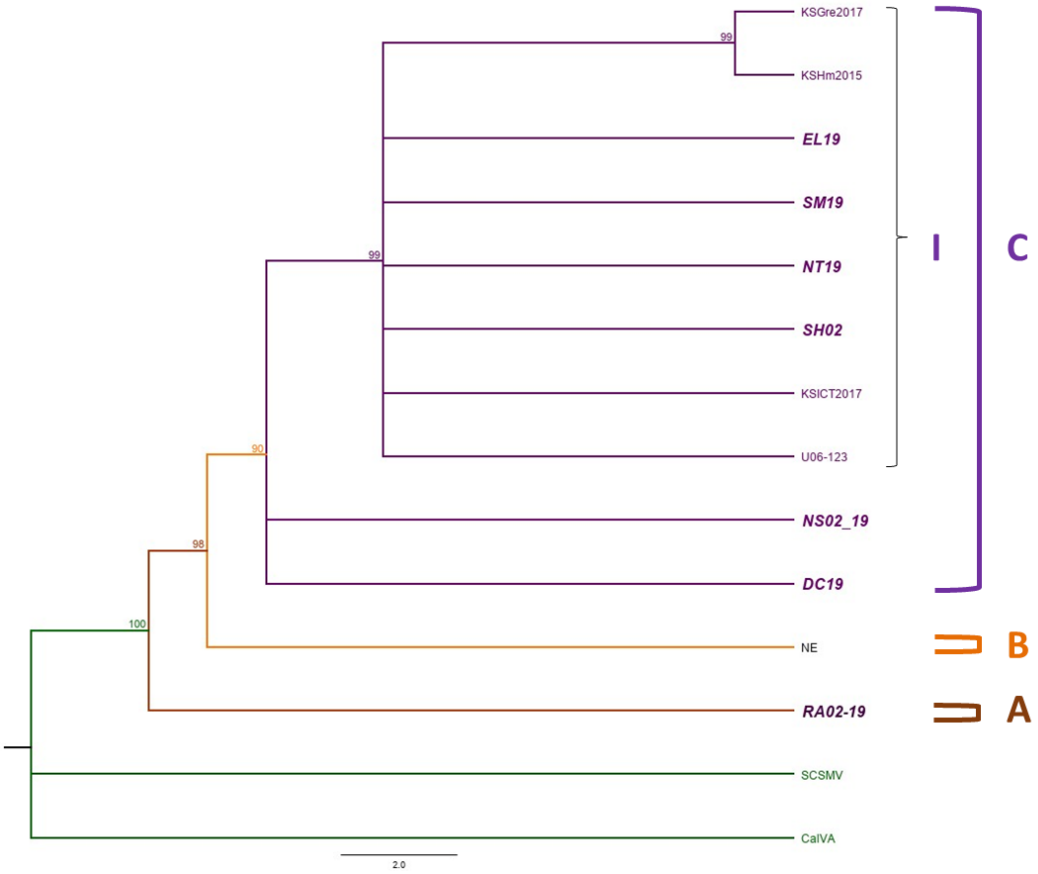


Figure 2-7. A schematic illustration of the genome organization of the Kansas HPWMoV isolate obtained in this study. The near complete genome of all 8 segments (RNA1-8) for HPWMoV was generated.

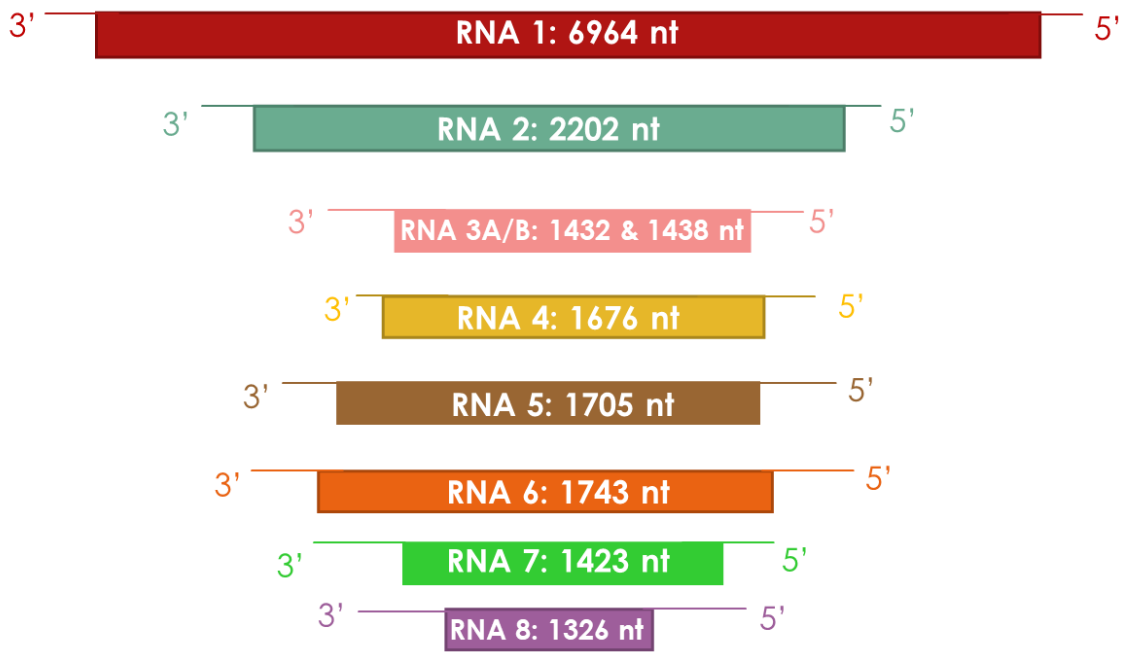


Table 2-1. Population genetics parameters calculated using DnaSP and MEGA for encoded regions of Kansas WSMV isolates.

Genomic Region	Number of Isolates	[1]S	[2]η	[3]π	[4]Θ$_W$	[5]dS	[6]dN	[7]dN/dS (GƆ)
P1	20	173	180	0.025 ± 0.002	0.046	0.070 ± 0.006	0.0064 ± 0.001	0.091
HC-Pro	20	199	218	0.031 ± 0.003	0.049	0.098 ± 0.008	0.0050 ± 0.001	0.051
P3	20	86	90	0.018 ± 0.002	0.029	0.05 ± 0.005	0.0017 ± 0.0005	0.034
6K1	20	25	27	0.028 ± 0.004	0.046	0.099 ± 0.03	0.0025 ± 0.002	0.025
CI	20	375	412	0.035 ± 0.005	0.055	0.12 ± 0.007	0.0015 ± 0.0004	0.013
6K2	20	28	31	0.030 ± 0.006	0.052	0.13 ± 0.02	0.00 ± 0.00	0.000
NIa-VPg	20	97	99	0.029 ± 0.002	0.046	0.097 ± 0.01	0.0012 ± 0.005	0.012
NIa-Pro	20	104	108	0.024 ± 0.002	0.043	0.075 ± 0.008	0.0027 ± 0.0007	0.036
NIb	20	241	252	0.028 ± 0.001	0.045	0.088 ± 0.007	0.0043 ± 0.001	0.049
CP	20	135	143	0.022 ± 0.001	0.036	0.058 ± 0.007	0.0064 ± 0.001	0.110

[1] Total number of segregating sites.
[2] Total number of mutations.
[3] Nucleotide diversity with the standard deviation calculated by DnaSP.
[4] Estimate mutation rate using segregating sites.
[5] Number of synonymous substitutions per site from the overall mean of sequence pairs.
[6] Number of nonsynonymous substitutions per site from the overall mean of sequence pairs.
[7] Ratio of dN/dS used to determine the selective pressure for coding regions.

Table 2-2. Population genetics parameters calculated using DnaSP and MEGA for encoded regions of TriMV.

Genomic Region	Number of Isolates	[1]S	[2]η	[3]π	[4]Θw	[5]dS	[6]dN	[7]dN/dS (ꞶƆ)
P1	8	27	28	0.0062 ± 0.0009	0.0091	0.012 ± 0.003	0.0028 ± 0.001	0.23
HC-Pro	8	25	25	0.0049 ± 0.0009	0.0069	0.014 ± 0.003	0.0011 ± 0.0006	0.079
P3	8	14	14	0.0039 ± 0.0009	0.006	0.0078 ± 0.002	0.0015 ± 0.0008	0.19
6K1	8	1	1	0.0015 ± 0.0001	0.0023	0.0044 ± 0.004	0.00 ± 0.0	0.00
CI	8	29	29	0.0037 ± 0.0004	0.0058	0.010 ± 0.002	0.00054 ± 0.0003	0.054
6K2	8	1	1	0.0016 ± 0.0001	0.0025	0.0051 ± .006	0.00 ± 0.0	0.00
Nia VPg	8	9	9	0.0038 ± 0.0006	0.0058	0.012 ± 0.004	0.00 ± 0.0	0.00
NIa Pro	8	13	13	0.0052 ± 0.001	0.0074	0.0015 ± 0.004	0.00053 ± 0.0005	0.035
NIb	8	41	41	0.0070 ± 0.0001	0.011	0.019 ± 0.004	0.0023 ± 0.0008	0.12
CP	8	9	9	0.0031 ± 0.0001	0.0043	0.006 ± 0.002	0.0017 ± 0.001	0.28

[1] Total number of segregating sites.
[2] Total number of mutations.
[3] Nucleotide diversity with the standard deviation calculated by DnaSP.
[4] Estimate mutation rate using segregating sites.
[5] Number of synonymous substitutions per site from the overall mean of sequence pairs.
[6] Number of nonsynonymous substitutions per site from the overall mean of sequence pairs.
[7] Ratio of dN/dS used to determine the selective pressure for coding regions.

Table 2-3. Codon positions of the coding regions in WSMV and TriMV isolates affected by positive selection.

Virus	Site	[1]FEL dN-dS	FEL P-value	SLAC dN-dS	SLAC P-value	[1]FUBAR dN-dS	Bayes Posterior Probability
WSMV	118	5.12	0.048	11.60	0.18	16.76	0.99
TriMV	2677	16.66	0.076	84.64	0.19	29.62	0.94

[1]These methods produced significant results

Discussion

The most prevalent WSM-associated virus found in Kansas field was WSMV in single infections, followed by the mixed infections of TriMV+WSMV and one sample infected with all three viruses in this study (Figure 2-2 and 2-3). No single infections of TriMV or HPWMoV were observed. This is consistent with the result of the previous reports of WSM in the Great Plains showed that WSMV in single infections was predominant in the field and over 91% of the TriMV infection occurred in mixed infection with WSMV (Byamukama et al. 2013).

The distribution of WSM in the field relies heavily on the successful transmission of the viruses by WCM. In previous studies evaluating the transmission efficiency of WCM for both single and mixed infections, WCM has been found to be efficient in transmitting WSMV alone, whereas TriMV would have to be in a mixed infection with WSMV to increase the transmission efficiency (Seifers et al. 2009). In the other study focusing on HPWMoV transmission in a single infection, WCM from Kansas was shown to vector single infections of HPWMoV poorly in comparison to WSMV (Seifers et al. 2002). In addition to the variation in vector transmission for each virus, the different biotypes of WCM also affects the transmission efficiency significantly. Kansas fields have been observed to contain two WCM biotypes: Biotype 1 and 2 and due to the varied response of each biotype to the temperature, humidity, and precipitation, the predominant biotype constantly changes (Khalaf et al. 2020; Oliveira-Hofman et al. 2015). Both biotypes are able to transmit WSMV, but Biotype 1 has a lower rate of efficiency (Oliveira-Hofman et al. 2015). Biotype 1 has been shown to be a very poor vector of HPWMoV, while other studies suggested that Biotype 1 could not transmit TriMV in single infections (McMechan et al. 2014). With the addition of WSMV in infections, WCM Biotype 1 was able to transmit HPWMoV and TriMV more efficient in comparison to the transmission rates using the source of single infections of TriMV or HPWMoV (McMechan et al. 2014).

The efficiency of Biotype 2 in transmission of HPWMoV has not been studied. The presence of both WCM biotypes in Kansas wheat fields and the differential transmission of WSM by different biotypes with the fitness advantage brought forth by mixed infections may explain the distribution of WSM in this study and the previous surveys. In addition to this, a synergistic relationship between WSMV and TriMV has been observed in which a mixed infection increases the titer of WSMV in the initial start of the infection and the titer of TriMV increases into the later stages (Tatineni et al. 2019).

Previous reports of a WSMV resistant-breaking isolate (Kumssa et al. 2019) and other potential resistant-breaking field collected isolates for WSMV and TriMV (Fellers et al. 2019) suggest that viral populations are under pressure to evolve and overcome the resistance as a consequence of the deployment of resistant varieties in Kansas fields. Our analysis determined recombination as the major evolutionary force operated upon Kansas WSMV isolates and through the study of the whole genome, the recombination hotspots were located. The identified potential parents for recombinants suggested that traces of isolates from other states such as Washington and other countries from Europe are found in Kansas fields. Interestingly, our analysis showed two isolates from our 2019 wheat survey collection, KM19 and SA19 as potential parents of the previously reported resistant-breaking isolate KSH294 in Kansas (Kumssa et al. 2019). This suggests that the parents of the recombinant- KSH294 are still found in the field and probably more recombinant isolates are present in the virus population that aid into breaking resistance. Moreover, the major and minor parents of WSMV Kansas recombinants were isolates from different states and other countries, suggesting that the viral population in Kansas fields are diverse (Table S2-2 to 2-6).

While, the previous studies which only focused on the CP of WSMV reported the 3' terminus of the CP region of WSMV genome as the recombination hotspot (Robinson MD and Murray TD 2012); our analysis detected more hotspots in other regions of WSMV genome

including the regions encoding P1, HC-Pro, CI, and P3 proteins, and there were 3 samples with a recombination hotspot in the CP. This observation suggests that due to the different functions of encoded proteins and the variability of evolutionary pressures placed upon them to increase fitness, focusing on one protein for evolutionary studies would not provide a sufficient overview of the evolutionary forces acting upon the viral population.

The phylogenetic tree constructed in this study for WSMV by using the complete genome sequences, placed the U.S. isolates in 4 sub-clades in the Clade D. This is consistent with the previous grouping of WSMV isolates based on the coat protein sequence (Rabenstein et al. 2002; Stenger et al. 2002). However, the contents of sub-clades were different. Unlike sub-clade D1 only containing American Pacific Northwest (APNW) isolates in a study done by Stenger et al. (2009), the sub-clade D1 in this study also included the Kansas Type isolate. In sub-clade D2, one 2019 and one historical isolate from Kansas were found in a sister taxa and grouped with an isolate from Ohio. This was inconsistent with the content of this sub-clade from the previous report. However, the previous study grouped an isolate from Colorado in sub-clade D2 (Stenger et al. 2002); while, the Colorado isolate from this study (FW40) was grouped in sub-clade D4. Sub-clade D3 contained mostly historical isolates from Kansas with the exception of 3 isolates: two 2019 field isolates (RA02_19 and SA19) and one previously reported potential resistant-breaking isolate (KSWal2017), which is not consistent with previous studies of the sub-clade D3 containing Kansas isolates and isolates from other states in the U.S. And finally, sub-clade D4 contained isolates from Kansas and Nebraska similar to previous studies, but had additional isolates from Idaho (ID99) and a potential resistant-breaking isolate recently reported from Kansas fields (KSGre2017). The other difference was that an isolate from Turkey was placed in the sub-clade D4 in our constructed tree based on the complete genome sequence, but was placed within a grouped of polytomies and not in any sub-clades in previous studies. These results again showed that how considering different parts of

the genome for phylogenetic analysis can change the phylogenetic relationships among isolates. As we expected, Kansas WSMV isolates are widely distributed among the 4 sub-clades of the Clade D.

The presence of both historical and field collected WSMV isolates in our collection also gave us this opportunity to study the phylogenetic relationships between these two groups of isolates, and how the genetic structure of WSMV has changed from past to present. Interestingly, we observed mixed results regarding these relationships. The Kansas isolates in both sub-clades D2 and D3 contain a mixture of field and historical isolates. In sub-clade D2, ST19 isolate forms a sister taxa with a historical isolate PL95, and in sub-clade D3, both historical isolates ranging from 1971 to 2006 and recent isolates from 2017 and 2019 were present. This close relationship of recent and historical isolates suggests that the genetic structure of WSMV populations has been conserved and little changes have been introduced throughout the years. However, the story was different in sub-clade D4 in which the separation of historical isolates from 2019 field isolates was clearly seen: the historical FO93 isolate is grouped with other historical isolates from Nebraska (Sidney81) and Idaho (ID99); while, 2019 field samples formed sister taxas separately from historical isolates in this sub-clade: ME19 and KW19 in one, and RP19 and EL19 in the other sister taxa. Hence, our results showed that some of the recent field isolates of WSMV share a close relationship with historical samples, whereas other isolates do not. These contradicting results suggest that the viral populations of WSMV is complex and determining the dominant isolate would be a difficult task. It is worth mention that this study captured a snapshot of genetic variation of WSMV field isolates in Kansas and determined the phylogenetic relationships of these isolates with historically curated isolates along with reference WSMV isolates.

We also presented the first phylogenetic study of TriMV isolates. Before the current study, the complete genome sequences of only 5 TriMV isolates were deposited in the GenBank

in which 4 of them were from Kansas (Fellers et al. 2009; Fellers et al. 2019). Our total RNA sequencing was able to generate the full genome sequences of 7 extra TriMV isolates from Kansas fields. The phylogenetic relationship of TriMV isolates presented here provided an insight into the potential evolutionary pressures which may be acting upon this virus (Figure 2-6). One of the results of this phylogenetic study that we found interesting was that some of our 2019 field isolates were grouped together with recently reported potential resistant-breaking isolates with a high support (bootstrap value of 99%). This close relationship suggests that TriMV isolates may be under pressure to evolve due to the widely use of resistant wheat varieties in the field. However, the analysis of a greater number of isolates would be needed to validate that claim.

We also reported the full genome sequences of 8 segments of an HPWMoV isolate which based on our knowledge is the second Kansas isolate to be completely sequenced, but compared to the first Kansas isolate, the sample in this study has been isolated from wheat (Figure 2-7). In previous studies, the RNA 3 of HPWMoV was found to contain two variants: RNAs 3A (1439 nt) and 3B (1441nt; Tatineni et al. 2014). RNAs 3A and 3B had differences of 12.5% in the nucleotide and 11.1% in the amino acid level (Tatineni et al. 2014). Both variants were found in a Kansas isolate, as well as Ohio and Nebraska (Tatineni et al. 2014; Stewart 2016). Similar to the findings of Stewart 2016, the Kansas HPWMoV isolate from this study also contained both variants of RNA 3. Since we could not find enough positive samples for HPWMoV in our collection, and also the presence of very limited numbers of the complete genome sequences of this virus in the GenBank, we did not perform the phylogenetic study for HPWMoV.

Overall, we observed low genetic diversity for both WSMV and TriMV here (π= 0.027 and 0.0038, respectively). The observed low genetic diversity among WSMV isolates is not surprising, as similar results have previously been reported for populations of this virus and

most plant viruses, such as *Wheat yellow mosaic virus*, *Cucumber mosaic virus*, and *Citrus psorosis virus*. (Stenger et al. 2002; Nouri et al. 2014; Sun et al. 2013; Martin et al. 2006). In fact, our finding is consistent with the concept that genetic stability is the rule in natural plant virus populations (Garcia-Arenal et al. 2001). Genetic bottleneck during the cell to cell movement and vector transmission has been suggested to be the reason for the reduction in variation of WSMV (French and Stenger 2003). In addition to genetic bottleneck, purifying selection may also be aiding in reducing diversity and maintaining a genetic stability in plant viruses. Among the encoded genomic regions of the Kansas isolates studied here, the CI gene showed the highest diversity (π=0.035) in the genome of WSMV (CI > HC-Pro > NIa-Vpg > 6K2 > 6K1 > NIb >P1 > NIa-Pro > CP > P3); while, the NIb gene possessed the greatest diversity (π=0.004) in the TriMV genome (NIb > P1 > NIa-Pro > HC-Pro > P3 > NIa-Vpg > CI > CP > 6K2 > 6K1). CI has been determined to be multifunctional, involved in the viral movement and contains a nucleotide binding motif which has been implicated to be part of the RNA helicase complex (Choi et al. 2000; Stenger et al. 1998; Lopez et al. 2009). Therefore, there is likely a high level of interaction between this protein and host components as well as other viral proteins which allow the CI to be more tolerant to nucleotide changes. In addition to this, CI was one of the recombination hotspots for WSMV in this study. On the other hand, the NIb demonstrating the highest variation in TriMV is the replicase which aids in viral replication (Tatineni et al. 2009). Although there have been no individual study focusing on other functions of NIb for TriMV, studies of other potyviruses, such as *Plum pox virus* and *Soybean mosaic virus*, have observed that the NIb is essential in interactions with viral and host proteins leading to formation of viral replication complexes (VRCs) and also post translation, targets plant defense pathway proteins to suppress immunity response, in addition to serving as the RdRp (Shen et al. 2020). Through these findings, we hypothesize that the presence of WSMV may lead into competition with TriMV in recruiting host proteins for forming VRCs

or trigger immune responses from the host, leading to the increase pressure to introduce changes in the NIb protein.

Purifying (negative) selection was found as the main selection pressure acting on the whole encoding regions of WSMV and TriMV genomes as shown in the ratio of dN/dS <1 in this study (Tables 2-1 and 2-3). Previous studies of WSMV focusing on the CP encoding genomic region also point to negative selection as the primary type of selection pressure (Robinson MD and Murray TD 2012; Stenger et al. 2002; Choi et al. 2001). Selection can be associated with various factors such as structural features of the virus, host plant and an arthropod vector. On average, the evolutionary constraints exerted on WSMV proteins P1, CP, and HC-Pro were less than CI, NIa-VPg, and 6K2. Previous studies of potyviruses observed that the proteins with the highest occurrence of positive selection and lowest negative selection were the P1, HC-Pro, NIa-VPg, and CP proteins, which correspond with the results for both WSMV and TriMV (Nigam et al. 2019). However, the NIa-VPg for WSMV does contain higher evolutionary constraints, inconsistent with previous studies. In contrast, the potyviral proteins with the highest frequency of negative selection which are considered more stable are the 6K1, 6K2, CI, and NIb proteins (Nigam et al. 2019). Interestingly, the evolutionary constraints exerted on TriMV proteins was different as the greater constraints were on 6K1, 6K2, and NIa-VPg proteins and the least were on the NIb, P1, and the CP. Although the exact function of the 6K1 and 6K2 proteins is not well known, these proteins are probably involved in virus replication and pathogenicity based on studies of other potyviruses (Nigam et al. 2019; Cui and Wang 2016).

Although the purifying selection was identified as the main pressure on the whole encoding regions of both WSMV and TriMV, our comprehensive neutrality test found a few positively selected sites (codons) for both WSMV and TriMV (Tables 2-2 and 2-4). To the best of our knowledge, this is the first deep analysis of the natural selections imposed on every

single codons of the encoded proteins of these two viruses. Through positive selection, viruses including plant viruses may introduce changes to successfully expand host range or host and vector adaptation (Nigam et al. 2019). Site 118 determined to be under positive selection for WSMV corresponds to the P1 protein. For TriMV, site 2677 located in the NIb protein was accepted as the positively selected codon. To our knowledge, there is no information about the function (s) of any of these mentioned codons. Therefore, we have insufficient information at the current time to propose specific reasons for these selections. However, we hypothesize that amino acids located in these regions of the proteins are involved in an important function.

The P1 protein is the main silencing suppressor of WSMV, and its role as the pathogenicity enhancer has also been shown (Tatineni et al. 2012; Young et al. 2012). Hence, this suggests that the site 118 of the P1 protein may be undergoing evolutionary changes to successfully counter the plant defense and increase pathogenicity (Singh et al. 2018). On the other hand, the NIb is the replicase which aids in viral replication (Tatineni et al. 2009). NIb is highly conserved among potyviruses and is often under strong purifying selection (Stenger et al. 1998; Nigam et al. 2019). However, the NIb protein is the protein with the highest diversity and is undergoing positive selection for TriMV, which suggests that this variation may be utilized by TriMV to compete with WSMV for replication during co-infection or it is essential for the interaction of the replicase (NIb) with the host factors. Hence, an improved understanding of the function of these regions for both viruses remains as an interesting aspect that warrants further investigation.

Taken together, the results obtained from this study demonstrated the importance of the whole genome sequence analyses to produce more informative data to study filed populations of the WSM-associated viruses. Gaining a better understanding of the genetic variation and evolutionary mechanisms utilized by WSMV, TriMV, and HPWMoV natural populations would aid in creating more effective and durable disease management strategies, and help

identifying key evolutionary strategies utilized by the viruses to overcome current resistance to successfully infect the wheat.

References

Adams, M. J., Antoniw, J. F., & Beaudoin, F. (2005). Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Mol Plant Pathol*, *6*(4), 471–487. https://doi.org/10.1111/j.1364-3703.2005.00296.x

Aleman-Verdaguer, M. E., Beachy, R. N., Fauquet, C., Goudou-Urbino, C., & Dubern, J. (1997). Analysis of the sequence diversity of the P1, HC, P3, NIb and CP genomic regions of several yam mosaic potyvirus isolates: implications for the intraspecies molecular diversity of potyviruses. *J Gen Virol*, *78*(6), 1253–1264. https://doi.org/10.1099/0022-1317-78-6-1253

Andrews, S. (2010). *FastQC a Quality Control Tool for High Throughput Sequence Data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Byamukama, E., Seifers, D. L., Hein, G. L., De Wolf, E., Tisserat, N. A., Langham, M. A. C., Osborne, L. E., Timmerman, A., & Wegulo, S. N. (2013). Occurrence and Distribution of Triticum mosaic virus in the Central Great Plains. *Plant Dis*, *97*(1), 21–29. https://doi.org/10.1094/pdis-06-12-0535-re

Choi, I.-R., Hall, J. S., Henry, M., Zhang, L., Hein, G. L., French, R., & Stenger, D. C. (2001). Contributions of genetic drift and negative selection on the evolution of three strains of wheat streak mosaic tritimovirus. *Arch Virol*, *146*(3), 619–628. https://doi.org/10.1007/s007050170167

Choi, I.-R., Horken, K. M., Stenger, D. C., & French, R. (2002). Mapping of the P1 proteinase cleavage site in the polyprotein of Wheat streak mosaic virus (genus Tritimovirus). *J Gen Virol*, *83*(2), 443–450. https://doi.org/10.1099/0022-1317-83-2-443

Choi, I.-R., Stenger, D. C., & French, R. (2000). Multiple Interactions among Proteins Encoded by the Mite-Transmitted Wheat Streak Mosaic Tritimovirus. *Virology*, *267*(2), 185–198. https://doi.org/10.1006/viro.1999.0117

Chung, B. Y.-W., Miller, W. A., Atkins, J. F., & Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *P Natl A Sci*, *105*(15), 5897–5902. https://doi.org/10.1073/pnas.0800468105

Cui, H., & Wang, A. (2016). Plum Pox Virus 6K1 Protein Is Required for Viral Replication and Targets the Viral Replication Complex at the Early Stage of Infection. *J Virol*, *90*(10), 5119–5131. https://doi.org/10.1128/jvi.00024-16

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*, *9*(8), 772–772. https://doi.org/10.1038/nmeth.2109

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Fellers, J. P., Seifers, D., Ryba-White, M., & Martin, T. J. (2009). The complete genome sequence of Triticum mosaic virus, a new wheat-infecting virus of the High Plains. *Arch Virol*, *154*(9), 1511–1515. https://doi.org/10.1007/s00705-009-0462-1

Fellers, J. P., Webb, C., Fellers, M. C., Shoup Rupp, J., & De Wolf, E. (2019). Wheat Virus Identification Within Infected Tissue Using Nanopore Sequencing Technology. *Plant Dis*, *103*(9), 2199–2203. https://doi.org/10.1094/pdis-09-18-1700-re

French, R., & Stenger, D. C. (2003). Evolution of Wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation. *Annu rev phytopathol*, *41*, 199–214. https://doi.org/10.1146/annurev.phyto.41.052002.095559

Friebe, B., Mukai, Y., Dhaliwal, H. S., Martin, T. J., & Gill, B. S. (1991). Identification of alien chromatin specifying resistance to wheat streak mosaic and greenbug in wheat germ

plasm by C-banding and in situ hybridization. *Theor Appl Genet*, *81*(3), 381–389. https://doi.org/10.1007/BF00228680

Gibbs, M. J., Armstrong, J. S., & Gibbs, A. J. (2000). Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinform*, *16*(7), 573–582. https://doi.org/10.1093/bioinformatics/16.7.573

Guindon, S., & Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst Biol*, *52*(5), 696–704. https://doi.org/10.1080/10635150390235520

Gupta, A. K., Hein, G. L., Graybosch, R. A., & Tatineni, S. (2018). Octapartite negative-sense RNA genome of High Plains wheat mosaic virus encodes two suppressors of RNA silencing. *Virol*, *518*, 152–162. https://doi.org/10.1016/j.virol.2018.02.013

Hollandbeck G.F., DeWolf E., and Todd T. (2017). Kansas cooperative plant disease survey report preliminary 2017 Kansas wheat disease loss estimates. https://agriculture.ks.gov/docs/default-source/pp-disease-reports-2012/2017-ks-wheat-disease-loss-estimates.pdf?sfvrsn=d49587c1_0

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinform*, *17*(8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754.

Jensen S.G., Lane L.C., and Seifers D.L. (1996). A new disease of maize and wheat in the high plains. *Plant Dis*, *80*(12), 1387-1390. https://doi.org/10.1094/PD-80-1387

Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol Biol Evol*, *22*(5), 1208–1222. https://doi.org/10.1093/molbev/msi105

Kumar, S. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*, *5*(2), 150–163. https://doi.org/10.1093/bib/5.2.150

Kumssa, T. T., Rupp, J. S., Fellers, M. C., Fellers, J. P., & Zhang, G. (2019). An isolate of Wheat streak mosaic virus from foxtail overcomes Wsm2 resistance in wheat. *Plant Pathol*, *68*(4), 783–789. https://doi.org/10.1111/ppa.12989

Lam, H. M., Ratmann, O., & Boni, M. F. (2018). Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Mol Biol Evol*, *35*(1), 247–251. https://doi.org/10.1093/molbev/msx263

Liu, W., Seifers, D. L., Qi, L. L., Friebe, B., & Gill, B. S. (2011). A Compensating Wheat-Thinopyrum intermedium Robertsonian Translocation Conferring Resistance to Wheat Streak Mosaic Virus and Triticum Mosaic Virus. *Crop Sci*, *51*(6), 2382–2390. https://doi.org/10.2135/cropsci2011.03.0118

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W., & Ray, S. C. (1999). Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *J Virol*, *73*(1), 152–160. https://doi.org/10.1128/jvi.73.1.152-160.1999

Lopez-Moya J.J., Valli A., & Garcia J.A. (2009). Potyviridae. *eLS*. https://doi.org/10.1002/9780470015902.a0000755.pub2

Lu, H., Price, J., Devkota, R., Rush, C., & Rudd, J. (2011). A Dominant Gene for Resistance to Wheat Streak Mosaic Virus in Winter Wheat Line CO960293-2. *Crop Sci*, *51*(1), 5–12. https://doi.org/10.2135/cropsci2010.01.0038

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol*, *1*(1), vev003–vev003. https://doi.org/10.1093/ve/vev003

Martin, D. P., Posada, D., Crandall, K. A., & Williamson, C. (2005). A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination

Breakpoints. *AIDS Res Hum Retrov*, *21*(1), 98–102. https://doi.org/10.1089/aid.2005.21.98

Martin, D., & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, *16*(6), 562–563. https://doi.org/10.1093/bioinformatics/16.6.562

Martin, S. (2006). Genetic variation of populations of Citrus psorosis virus. *J Gen Virol*, *87*(10), 3097–3102. https://doi.org/10.1099/vir.0.81742-0

McKinney, H.H. (1937). Mosaic diseases of wheat and related cereals. *USDA*, *4*, 1–22. https://archive.org/stream/mosaicdiseasesof442mcki/mosaicdiseasesof442mcki_djvu.txt

Mcmechan, A. J., Tatineni, S., French, R., & Hein, G. L. (2014). Differential Transmission of Triticum mosaic virus by Wheat Curl Mite Populations Collected in the Great Plains. *Plant Dis*, *98*(6), 806–810. https://doi.org/10.1094/pdis-06-13-0582-re

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol Biol Evol*, *30*(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

National Agricultural Statistics Service. (2020). Kansas States 2019 Agricultural Overview. https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=kansas

Nigam, D., Latourrette, K., Souza, P. F. N., & Garcia-Ruiz, H. (2019). Genome-Wide Variation in Potyviruses. *Front Plant Sci*, *10*. https://doi.org/10.3389/fpls.2019.01439

Nouri, S., Arevalo, R., Falk, B. W., & Groves, R. L. (2014). Genetic Structure and Molecular Variability of Cucumber mosaic virus Isolates in the United States. *PLOS ONE*, *9*(5), e96582. https://doi.org/10.1371/journal.pone.0096582

Oliveira-Hofman, C., Wegulo, S. N., Tatineni, S., & Hein, G. L. (2015). Impact of Wheat streak mosaic virus and Triticum mosaic virus Coinfection of Wheat on Transmission Rates by Wheat Curl Mites. *Plant Dis*, *99*(8), 1170–1174. https://doi.org/10.1094/pdis-08-14-0868-re

Padidam, M., Sawyer, S., & Fauquet, C. M. (1999). Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology*, *265*(2), 218–225. https://doi.org/10.1006/viro.1999.0056

Pond, S. L. K., Frost, S. D. W., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinform*, *21*(5), 676–679. https://doi.org/10.1093/bioinformatics/bti079

Posada, D., & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci*, *98*(24), 13757–13762. https://doi.org/10.1073/pnas.241370698

Rabenstein, F., Schubert, J., Seifers, D. L., Stenger, D. C., & French, R. (2002). Phylogenetic relationships, strain diversity and biogeography of tritimoviruses. *J Gen Virol*, *83*(4), 895–906. https://doi.org/10.1099/0022-1317-83-4-895

Robinson, M. D., & Murray, T. D. (2013). Genetic Variation of Wheat streak mosaic virus in the United States Pacific Northwest. *Phytopathology*, *103*(1), 98–104. https://doi.org/10.1094/phyto-05-12-0108-r

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinform*, *19*(12), 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Rozas, J., Sanchez-Delbarrio, J. C., Messeguer, X., & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinform*, *19*(18), 2496–2497. https://doi.org/10.1093/bioinformatics/btg359

Salminen, M. O., Carr, J. K., Burke, D. S., & McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retrov*, *11*(11), 1423–1425. https://doi.org/10.1089/aid.1995.11.1423

Seifers, D. L., Harvey, T. L., Louie, R., Gordon, D. T., & Martin, T. J. (2002). Differential Transmission of Isolates of the High Plains virus by Different Sources of Wheat Curl Mites. *Plant Dis*, *86*(2), 138–142. https://doi.org/10.1094/pdis.2002.86.2.138

Seifers, D. L., Martin, T. J., Harvey, T. L., Fellers, J. P., & Michaud, J. P. (2009). Identification of the Wheat Curl Mite as the Vector of Triticum mosaic virus. *Plant Dis*, *93*(1), 25–29. https://doi.org/10.1094/pdis-93-1-0025

Seifers, D. L., Martin, T. J., Harvey, T. L., Fellers, J. P., Stack, J. P., Ryba-White, M., Haber, S., Krokhin, O., Spicer, V., Lovat, N., Yamchuk, A., & Standing, K. G. (2008). Triticum mosaic virus: A New Virus Isolated from Wheat in Kansas. *Plant Dis*, *92*(5), 808–817. https://doi.org/10.1094/pdis-92-5-0808

Shen, W., Shi, Y., Dai, Z., & Wang, A. (2020). The RNA-Dependent RNA Polymerase NIb of Potyviruses Plays Multifunctional, Contrasting Roles during Viral Infection. *Viruses*, *12*(1), 77. https://doi.org/10.3390/v12010077

Singh, K., Wegulo, S. N., Skoracka, A., & Kundu, J. K. (2018). Wheat streak mosaic virus: a century old virus with rising importance worldwide. *Mol Plant Pathol*, *19*(9), 2193–2206. https://doi.org/10.1111/mpp.12683

Smith J. M. (1992). Analyzing the mosaic structure of genes. *J Mol Evol*, *34*(2), 126–129. https://doi.org/10.1007/BF00182389

Stenger, D. C., & French, R. (2009). Wheat streak mosaic virus genotypes introduced to Argentina are closely related to isolates from the American Pacific Northwest and Australia. *Arch Virol*, *154*(2), 331–336. https://doi.org/10.1007/s00705-008-0297-1

Stenger, D. C., Hall, J. S., Choi, I.-R., & French, R. (1998). Phylogenetic Relationships Within the Family Potyviridae: Wheat Streak Mosaic Virus and Brome Streak Mosaic Virus Are Not Members of the Genus Rymovirus. *Phytopathology*, *88*(8), 782–787. https://doi.org/10.1094/phyto.1998.88.8.782

Stenger, D. C., Seifers, D. L., & French, R. (2002). Patterns of Polymorphism in Wheat streak mosaic virus: Sequence Space Explored by a Clade of Closely Related Viral Genotypes Rivals That between the Most Divergent Strains. *Virology*, *302*(1), 58–70. https://doi.org/10.1006/viro.2001.1569

Stewart, L. R. (2016). Sequence diversity of wheat mosaic virus isolates. *Virus Res*, *213*, 299–303. https://doi.org/10.1016/j.virusres.2015.11.013

Sun, B.-J., Sun, L.-Y., Tugume, A. K., Adams, M. J., Yang, J., Xie, L.-H., & Chen, J. P. (2013). Selection Pressure and Founder Effects Constrain Genetic Variation in Differentiated Populations of Soilborne Bymovirus Wheat yellow mosaic virus (Potyviridae) in China. *Phytopathology*, *103*(9), 949–959. https://doi.org/10.1094/phyto-01-13-0013-r

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol*, *28*(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

Tatineni S., Kovacs F., and French, R. (2014a). Wheat streak mosaic virus infects systemically despite extensive coat protein deletions: Identification of virion assembly and cell-to-cell movement determinants. *J Virol*, *88*, 1366–1380. https://doi.org/10.1128/jvi.02737-13

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. *Mol Plant Microbe In*, *27*(2), 150–162. https://doi.org/10.1094/mpmi-09-13-0272-r

Tatineni, S., & Hein, G. L. (2018). Genetics and mechanisms underlying transmission of Wheat streak mosaic virus by the wheat curl mite. *Curr Opin Virol*, *33*, 47–54. https://doi.org/10.1016/j.coviro.2018.07.012

Tatineni, S., Alexander, J., Gupta, A. K., & French, R. (2019). Asymmetry in Synergistic Interaction Between Wheat streak mosaic virus and Triticum mosaic virus in Wheat. *Mol Plant Microbe In*, *32*(3), 336–350. https://doi.org/10.1094/mpmi-07-18-0189-r

Tatineni, S., and French, R. (2016). The coat protein and NIa protease of two Potyviridae family members independently confer superinfection exclusion. *J Virol, 90*(23),10886-10905.

Tatineni, S., Mcmechan, A. J., Wosula, E. N., Wegulo, S. N., Graybosch, R. A., French, R., & Hein, G. L. (2014b). An Eriophyid Mite-Transmitted Plant Virus Contains Eight Genomic RNA Segments with Unusual Heterogeneity in the Nucleocapsid Protein. *J Virol*, *88*(20), 11834–11845. https://doi.org/10.1128/jvi.01901-14

Tatineni, S., Qu, F., Li, R., Jack Morris, T., & French, R. (2012). Triticum mosaic poacevirus enlists P1 rather than HC-Pro to suppress RNA silencing-mediated host defense. *Virology*, *433*(1), 104–115. https://doi.org/10.1016/j.virol.2012.07.016

Tatineni, S., Van Winkle, D.H., French, R. (2011). The N-terminal region of wheat streak mosaic virus coat protein is a host- and strain-specific long-distance transport factor. *J Virol, 85*, 1718–1731.

Tatineni, S., Ziems, A. D., Wegulo, S. N., & French, R. (2009). Triticum mosaic virus: A Distinct Member of the Family Potyviridae with an Unusually Long Leader Sequence. *Phytopathology*, *99*(8), 943–950. https://doi.org/10.1094/phyto-99-8-0943

Urcuqui-Inchima, S., Haenni, A.-L., & Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Res*, *74*(1-2), 157–175. doi:10.1016/s0168-1702(01)00220-9

Young, B. A., Stenger, D. C., Qu, F., Morris, T. J., Tatineni, S., & French, R. (2012). Tritimovirus P1 functions as a suppressor of RNA silencing and an enhancer of disease symptoms. *Virus Res*, *163*(2), 672–677. https://doi.org/10.1016/j.virusres.2011.12.019

Zheng, L., Rodoni, B. C., Gibbs, M. J., & Gibbs, A. J. (2010). A novel pair of universal primers for the detection of potyviruses. *Plant Pathol*, *59*(2), 211–220. https://doi.org/10.1111/j.1365-3059.2009.02201.x

Chapter 3 - Metagenomics analysis of the wheat virome

Abstract

Wheat viruses including *Wheat streak mosaic virus* (WSMV), *Triticum mosaic virus* (TriMV), *High Plains wheat mosaic emaravirus* (HPWMoV), and *Barley yellow dwarf virus* (BYDV) cost substantial losses in crop yields every year. Although there have been extensive studies conducted on these known wheat viruses, currently, there is limited knowledge about all components of the wheat *(Triticum aestivum L.)* virome. Here, we determined the composition of the wheat virome through total RNA deep sequencing. Twenty-four symptomatic and asymptomatic field-collected leaf samples were analyzed. Sequences were *de novo* assembled after removing the host reads and a BLASTx search against NCBI viral genomes were conducted. In addition to the known wheat viruses, viral sequences sharing low (<40%) but significant similarity with potyviruses and recently discovered virga-like viruses were identified. Our analysis showed that these viral sequences belong to a novel putative virus tentatively named *Wheat-associated Vipovirus* (WaVPV). Two other novel viral sequences have also been discovered in the same libraries as WaVPV: one viral sequence has a ~40% similarity with fungal-associated viruses and another with ~40% similarity with virga-like fungal viruses and ~20% with potyviruses. The presence of viral sequences was validated in original RNAs using RT-PCR and Sanger sequencing. The discovery and characterization of novel viruses associated with wheat is important to determine if they may pose a threat to the wheat industry or have the potential to be used as new biological control agents for wheat pathogens.

Keywords: Wheat virome; Metagenomics; RNA-seq; Putative virus; Wheat virus; Fungal virus

Introduction

Wheat is one of the leading staple crops in the world; however, wheat viral diseases continue to pose a great threat to the industry (Rotenberg et al. 2016; Singh et al. 2020). Wheat viruses including *Wheat streak mosaic virus* (WSMV), *Triticum mosaic virus* (TriMV), *High Plains wheat mosaic emaravirus* (HPWMoV), *Barley yellow dwarf virus* (BYDV), and *Cereal yellow dwarf virus* (CYDV) have a significant impact in the Great Plains and other wheat growing regions around the world (Hodge et al. 2020; Rotenberg et al. 2016; Singh et al. 2020). WSMV, TriMV, and HPWMoV have been reported as causal agents of the wheat streak disease (WSM) complex; a destructive disease causing significant yield loss (Hollandbeck et al. 2017). Despite successful breeding programs, wheat viral diseases are still the major issues for farmers. Enormous studies have been conducted to characterize these wheat pathogenic viruses and determine virus-host interactions (Tatineni et al. 2014; Zhang et al. 2019; Nyhren et al. 2015). However, most of these studies have focused on the documented wheat viruses (Hodge et al. 2020; Singh et al. 2020), and there has been no metagenomics study to analyze all viral communities associated with wheat.

Viruses are the most abundant habitants of the earth, infecting both prokaryotes and eukaryotes (Monier et al. 2008; Shi et al. 2016; Shi et al. 2018). The study of virology has been spurred on by the identification of a plant virus, *Tobacco mosaic virus*, which was the first virus to be discovered in 1898 (Creager et al. 1999). Since then, multitude of virology studies have been conducted to identify and characterize pathogenic viruses causing disease in humans, animals, and plants (Jeong et al. 2014; Zakrzewski et al. 2018; Zafer et al. 2019), and determine the relationships of viral pathogens with their hosts and vectors (Murugan et al. 2011; Niu et al. 2020; Tatineni et al. 2019). With the introduction of high-throughput next-generation sequencing (NGS) and advanced bioinformatics tools, a new age of virus discovery has

commenced in the last decade and many novel pathogenic and non-pathogenic viruses have been identified (Roossinck et al. 2015 Mumo et al. 2020). Indeed, studying the virome (all viral communities associated with a particular organism, ecosystem, or holobiont) in the metagenomics era has been a significant contribution to the field of virology including plant virology (Adams et al. 2009; Villamor et al. 2019). The identification of persistent viruses which may be potential biological control candidates (Chiapello et al. 2020; Marzano et al. 2016), the discovery of novel and newly emerging viruses (Chiapello et al. 2020; Otuka A. 2013; Matsumura et al. 2017), and the determining of evolutionary relationships of economically important viral pathogens (Omeldo-Velarde et al. 2019; Ramos-Gonzalez et al. 2016) are some of these contributions.

To address the current knowledge gap about the wheat virome, we conducted a metagenomics study with symptomatic and asymptomatic wheat leaf samples collected from fields. Through high-throughput total RNA sequencing, not only known wheat viruses but novel viral sequences were identified in this study. The discovery and characterization of new viruses is crucial due to the fact that these viruses may be emerging viral pathogens and pose a threat to the wheat industry or be beneficial and have potential to serve as new biological control for wheat diseases either as wild type or recombinant viruses.

Materials and methods

Sample collection

Samples were collected in 2019 from counties in Northwest and Southwest Kansas. In addition to surveyed samples, samples sent to the Kansas State University Plant Disease Diagnostics Lab and the provided samples from Nebraska and Colorado fields were also added to the collection. A total of 24 samples were chosen for total RNA deep sequencing (Table S3-1).

RNA extraction and library preparation

Total RNAs from 100 mg of wheat leaf tissues were extracted using TRIzol reagent (Invitrogen, CA, USA), according to the manufacturer's instruction. The extracted RNAs were treated with DNase I (Zymo Research, CA, USA). The integrity and quantity of the DNase treated RNAs were measured using Qubit 4 (Invitrogen, Ca, USA) with the RNA IQ assay kit. The TruSeq Stranded Total RNA with Ribo-Zero Plant kit (Illumina Inc., CA, USA) was utilized to remove the rRNA and prepare the libraries for sequencing following the manufacturer's instruction. Agencourt RNAClean XP (Beckman Coulter, MA, USA) was used to purify the samples ensure the removal of all traces of rRNA. TruSeq RNA Single Indexes Sets A and B (Illumina Inc., CA, USA) were used for adapter ligation. After each step of cDNA synthesis, adapter ligation, and enrichment of the DNA fragments, the samples were purified using the Agencourt AMPure XP (Beckman Coulter, MA, USA). The final libraries were subjected to quality control analysis using Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) and were quantified using the Qubit 4 (Invitrogen, CA, USA) with the 1X dsDNA High Sensitivity Assay (Invitrogen, CA, USA). At the Kansas State Integrated Genomics Facility, the libraries were pooled and sequenced using the NextSeq 500 platform for high-output with a read length of 1x75bp.

Bioinformatics analysis

Trimmomatic was used to trim the reads for quality, length, and the adapter sequences (Bolger et al. 2014). To ensure the reads no longer contained adapter sequences and were of high quality, FastQC was utilized for quality control (Andrews 2010). STAR Aligner was used to index the wheat reference genome obtained from Ensemble Genomes project and align the indexed reference to the reads (Howe et al. 2020; Dobin et al. 2013). The unmapped reads were extracted and *de novo* assembled using two assemblers: CLC Genomics Workbench 20 (CLC Bio, Qiagen, MD, USA) and Trinity 2.8.0 (Haas et al. 2013). All *de novo* assembled contigs were used as query for BLASTx searches against the non-redundant protein database with the search parameter limited to the virus taxids and E-value < 0.001 (Altschul et al. 1990).

Obtaining the full length of cDNAs using RACE

The 5' and 3' untranslated regions (UTRs) of the novel viral sequences were completed through rapid amplification of cDNA ends (RACE) using the SMARTer RACE 5'/3' Kit (Takara Bio, CA, USA), following the manufacturer's instruction. Gene specific primers were designed for the 5'/3' end (Table S3-2). The PCR products were cloned using the Zero Blunt II TOPO kit (Invitrogen, CA, USA) and transformed using One Shot TOP10 Chemically Competent *E. coli* (Invitrogen, CA, USA). At least 15 clones were sent for Sanger Sequencing (Genewiz Inc., NJ, USA) and consensus sequences were extracted.

RT-PCR and small RNA deep sequencing

To validate the viral sequences, the first strand cDNAs were synthesized using SuperScript II Reverse Transcriptase (Invitrogen, CA, USA) with oligo dT and /or gene specific primers. For PCR, the 25 uL reaction contained 1x GoTaq Flexi Buffer (Promega,

WI, USA), 1 uM MgCl2, 0.1 uM dNTP, 0.4 uM of gene specific primers (Table S3-1), 1.25 U GoTaq Flexi DNA Polymerase (Promega, WI, USA), and nuclease-free water. The thermal cycle program for the GoTaq Flexi protocol is as follows: 94℃ for 2 minutes, 34 cycles of 94℃ for 10 seconds, 55℃ for 15 seconds, and 72℃ for 2 mins, and 72℃ for 5 minutes. The PCR products were cleaned up using the DNA Gel Extraction Kit (Zymo Research, CA, USA) and confirmed through Sanger Sequencing.

The total RNA of one sample confirmed to be infected with WaVPV through RT-PCR and Sanger sequencing was utilized for small RNA sequencing. The sample was treated with DNase (Zymo Research, CA, USA) and the RNA integrity was measured using the Qubit 4 RNA IQ assay (Invitrogen, CA, USA). Small RNA library preparation and sequencing was conducted by BGI (Hong Kong; MA, USA) using DNBSeq platform with a read length of 50 bp.

Phylogenetic analysis

The phylogenetic analysis was conducted only with novel viral sequences identified in this study. The amino acid sequences were aligned using MUSCLE (Edgar 2004) with the top viral reference sequences showed the highest similarities in BLASTx searches. MEGA 5 was utilize to determine the best substitution model for the alignments (Kumar et al. 2004). The maximum-likelihood tree was constructed using the PHYML plugin integrated in the Geneious Prime 2020 20.4 program (Guindon et al. 2010). The phylogenetic trees did not contain any outgroups due to the lack of sufficient information about the taxonomy of the novel viruses. The accession number for the sequences are found in Figures 3-4 and 3-5.

Results

RNA sequencing analysis

A total of 612M reads with an average of 21M reads were obtained (Table S3-3). Clean reads were mapped into the genome of the wheat (GenBank assembly accession: GCA_900519105.1) and unmapped reads were collected (Howe et al. 2020). Figure 3-1 shows all hits identified through BLASTx searches including hits for known plant and fungal-associated viruses and also novel viral sequences. The five known wheat viruses including WSMV, TriMV, HPWMoV, BYDV-PAV, and CYDV were identified and their full genome sequences were obtained. TriMV and HPWMoV did not occur in single infections, so the virus incidence of TriMV and HPWMoV in Figure 3-1 were in mixed infections. WSMV single infections (52%) followed by mixed infections of WSMV + TriMV (8%) were the most prevalent in the symptomatic leaf samples among the known wheat viruses. Other wheat viruses, such as BYDV-PAV and CYDV occurred in low incidences: 8% and 4%, respectively.

Two fungal-associated viruses, *Mitovirus* and *Totivirus*, were also found in the samples showed symptoms. The *Totivirus* showed 70-80% amino acid similarity with totiviruses infected stripe rust, *Puccinia striiformis.* In contrast, the sequences for the *Mitovirus* showed low amino acid similarities at 40% with other Mitoviruses. Along with the mixed infection of WSMV + TriMV, *Mitovirus* had the second highest incidence in the symptomatic samples (Figure 3-1). Additionally, poty-like viral sequences of the length of ~ 2kb with a BLASTx hit of 25% amino acid similarity, query coverage of 55%, and significant E-value (< 0.001) with potyviruses including *Barley mild mosaic virus* (BaMMV) were found in 6 libraries containing both symptomatic and asymptomatic samples. Other viral sequences with the length of ~ 3.2kb, a 30%

similarity, query coverage of 79%, and significant E-value (<0.001) at the amino acid level to the *Uromyces virgavirus* , a fungal-associated virga-like virus, and other insect-specific virga-like viruses were also found in the same libraries as the poty-like viral sequences. In addition to these findings, two other novel viral sequences have been discovered in the same 6 libraries in which the poty-like and virga-like novel sequences were found. One of these new viral sequences had ~48% similarity with a E value <0.001 and a coverage of 95% to the putative movement protein (MP) of unclassified virga-like, fungal-associated viruses- *Macrophomina phaseolina tobamo-like virus 1* (MpTLV1) and *Plasmopara viticola lesion associated tobamo-like virus (*PvLaT-LV1) ORF 2 but also contains a similarity of 21% with a plant virus from the *Potyiviridae* family- *Onion yellow dwarf virus.* The other novel viral sequence had a 41% similarity, 85% coverage and significant E value (<0.001) with the amino acid of putative coat protein (CP) viral sequences of virga-like, fungal-associated viruses such as MpTLV1 and PvLaT-LV1 ORF 3. These novel viral sequences will be referred to as putative MP and putative CP, respectively.

Through RT-PCR using gene-specific primers (Figure S1A-G) and Sanger sequencing, the presence of all known and novel viral sequences identified by total RNA sequencing were validated in the original RNAs.

Characterizing novel viral sequences

We completed and obtained the full genome sequences of poty-like and virga-like viral sequences using RACE. The other two novel viral sequences: putative MP and putative CP are still in the process of undergoing 3'/5' RACE and are not yet complete. Through RACE, the 5' and 3' UTRs obtained for the poty-like viral sequence were 100 and 114 nucleotides, respectively. For the virga-like sequence, the length of the 5' and 3' UTRs were 107 and 95 nucleotides,

respectively. Both poty-like and virga-like viral sequences contained a polyadenylated tail at the 3’ end.

The complete genome of the poty-like viral sequence was 2.4 kb with a single predicted open reading frame (ORF) of 2190 bp. The genome of the virga-like viral sequence was 3.7 kb with a larger predicted ORF of 3507 bp and a smaller ORF of the size of 252 bp. The 5’ and 3’ UTRs of the two viral sequences were aligned resulted in an 85% and 45% similarity, respectively (Figure 3-2). This result strongly suggested that these are two viral segments of one putative virus tentatively named *Wheat-associated vipovirus* (WaVPV) RNA 1 (virga-like) and RNA 2 (poty-like).

The predicted ORFs were used as queries to search against the Conserved Domain Database (CDD) (Marchler-Bauer et al. 2015). The RdRP conserved domain was detected for the ORF1 of WaVPV RNA1 with a significant E-value (< 0.001) which also contained the conserved GDD motif; while, no significant hit was identified for the ORF 2 of RNA 1 (Figure 3-3). The search for the WaVPV RNA2 large ORF resulted in a significant (E-value < 0.001) similarity with the DEAD-like helicase domain (Figure 3-3). Similar to WaVPV RNA2, the single predicted ORF for the viral sequence of putative MP resulted in a significant (E-value <0.001) similarity with the DEAD-like helicase domain; while, the ORF of putative CP did not have any significant hits in the CDD, much like the ORF 3 of PvLaT-LV1 (Chiapello et al. 2020).

The phylogenetic analysis of WaVPV RNA1 RdRp grouped this putative virus with a fungal-associated virus *Uromyces virgavirus* and is found in a clade of other fungal-associated and insect-specific viruses belonging in the *Virgaviridae* family (Figure 3-4). The phylogenetic tree built based on the RNA 2 dead-like helicase grouped WaVPV RNA2 with plant viruses from the *Potyviridae* family (Figure 3-5), which supports the BLASTx results.

A total of 52M small RNAs reads were obtained from small RNA deep sequencing. After quality filtering of the small RNA reads and trimming the adapters, reads were mapped to the complete genome sequences of WaVPV and the distribution of the mapped read lengths were obtained (Figure 3-6). There were 1,300 reads mapped to WaVPV RNA1 and 1,000 reads mapped to WaVPV RNA2. For both segments, the highest peak was at 21 nt, followed by 20 nt and 22 nt.

Figure 3-1. The composition of the wheat virome. Known wheat viruses (blue), known fungal-associated viruses (green), and novel viral sequences (red) were found in the RNA samples.

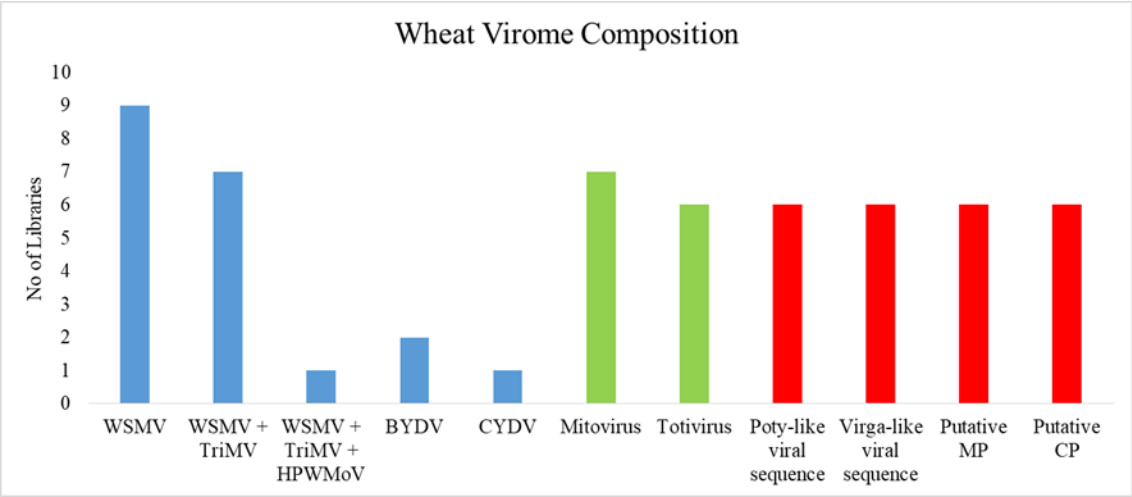


Figure 3-2. The alignment of the 5’ and 3’ UTRs of WaVPV. 85% and 45% similarity was determined for the 5’ and 3’ alignment, respectively. The shared nucleotides are highlighted in yellow.

```
5' WaVPV RNA 2 AGTTTTTGTCCTGCGTCCTTGAGACGTGACGAAAATCGTACATTAGGTACTTCCCGACAC   60
5' WaVPV RNA 1 AGTTTTTGTCCTGCGTCCTTGAGACGTGACGAAAATCGTACTATAAGTACTTCCCGAACC   60

5' WaVPV RNA 2 CACAACAAACGAACCGACCACACACGACACGACCACAAAT-------              100
5' WaVPV RNA 1 GACCACAAACCAAAACACCACACCCGACCCGACCACACAACACCGTA               107

3' WaVPV RNA 2 GTCACCAATTATATGCAATCCCTTACCAATTTATTCT-GTCCGTTGTGGACAATATGCT    58
3' WaVPV RNA 1 GTTGCT-GTCGTTTAGTTTAGTTTATTTATTTACTTTTAGATCATTGTGATCAA------   53

3' WaVPV RNA 2 ATATGTTATGCGCGTGAGTTATCACGTATTGAAATTTCCTTAAAACACGAAAGGAG---   114
3' WaVPV RNA 1 ---------ACCTTTAAGTTGT----TATT----TTTTTCCAAAACACGAGGAGAAAAG    95
```

Figure 3-3. The schematic genome organization of *Wheat-associated Vipovirus* RNAs 1 and 2 with the predicated ORFs and the conserved domains.

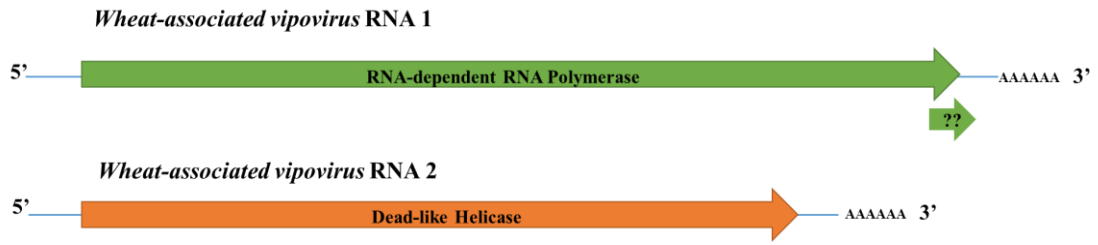


Figure 3-4. The phylogenetic analysis of the polyprotein of WaVPV RNA 1. This PHYML tree was constructed using the substitution method: rtREV + I + G with a bootstrap value of 1000. The fungal-associated viruses are in brown and the insect-specific viruses are written in green text.

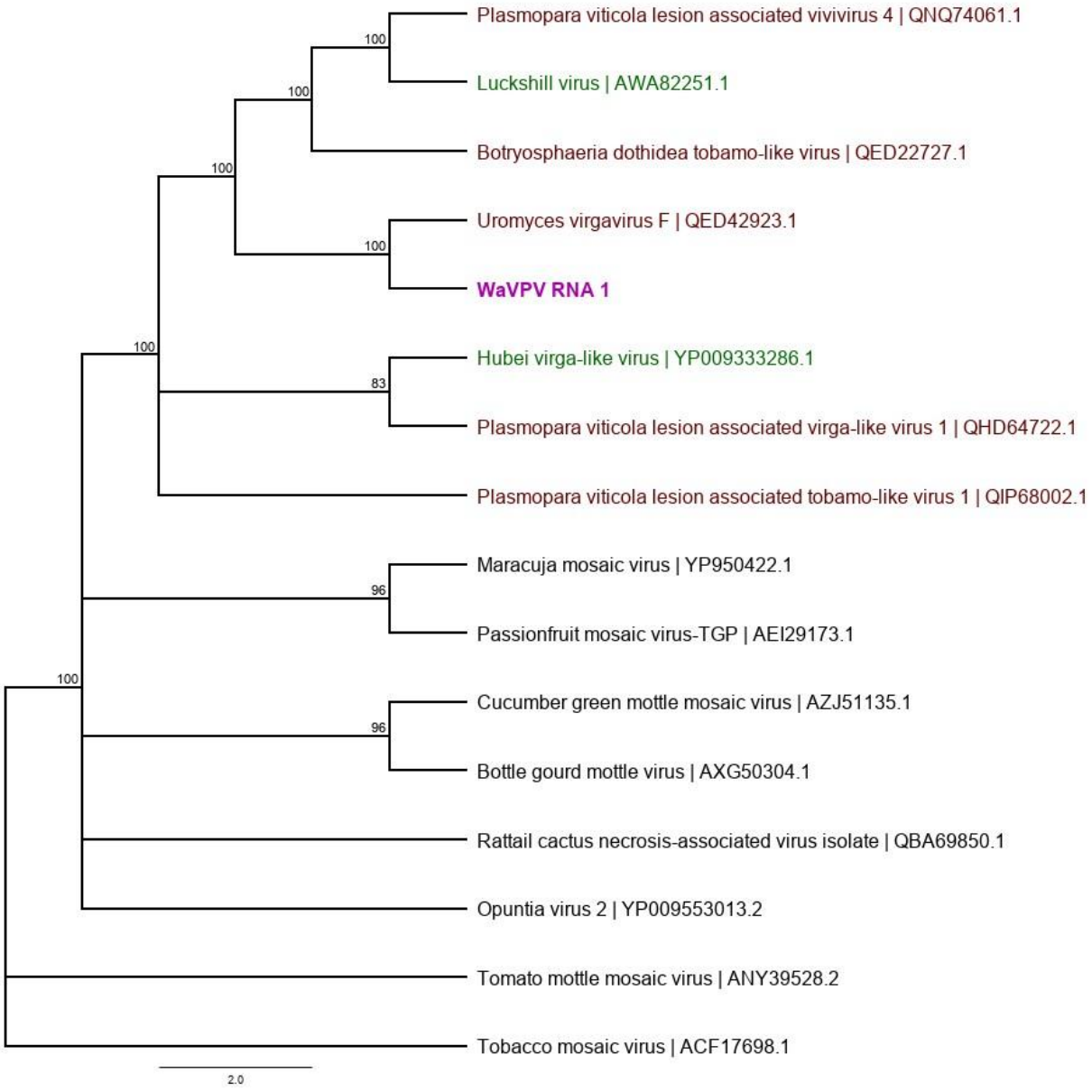


Figure 3-5. The phylogenetic analysis of the polyprotein of WaVPV RNA 2. This PHYML tree was constructed using the substitution method: WAG + G with a bootstrap value of 1000. All of the viral sequences used in this analysis were from the *Potyviridae* family showed the highest similarity with WaVPV RNA2 in BLASTx searches.

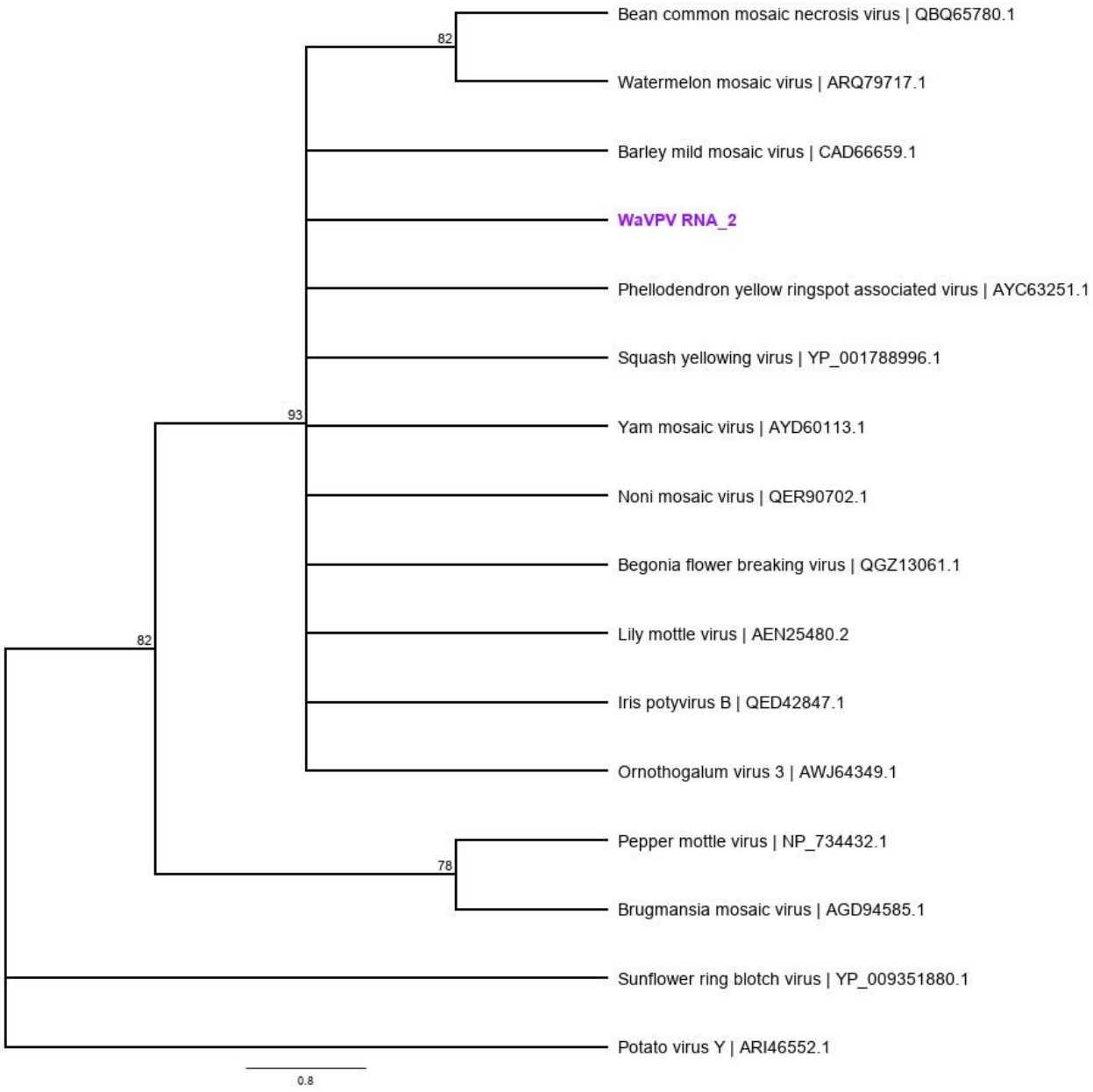
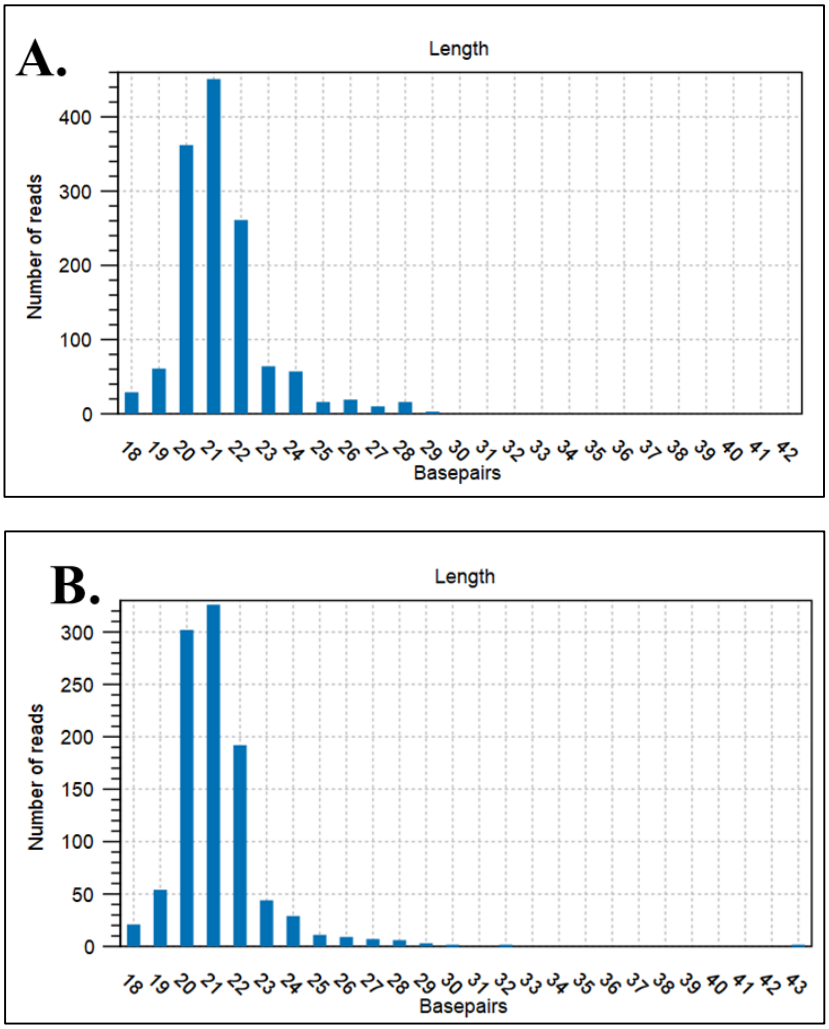


Figure 3-6. The distribution of small RNAs mapped to WaVPV. **A)** The reads mapped to WaVPV RNA1 with a strong peak at 21 nt. **B)** The reads mapped to the genome of WaVPV RNA2 with a strong peak at 21 nt.



Discussion

Wheat viruses continue to present a great threat to the wheat industry and global food security. Here, we determined the composition of the wheat virome through high throughput sequencing of the total RNA of field samples. The goal was to identify any undocumented/new viruses associated with this important crop which either may be a new threat to the wheat industry or have the potential to be used as biological agents for the control of wheat pathogens.

The study was able to uncover that among all known wheat viruses, WSMV is the most dominant. The other viruses which form a complex with WSMV, which are TriMV and HPWMoV, were only found in mixed infections with WSMV and not in single infections. The absence of single infections of TriMV and HPWMoV may be due to the types of the wheat curl mites (WCM), the natural vector of WSMV, TriMV, and HPWMoV found in the field. The presence of both WCM biotypes 1 and 2 has been documented in Kansas fields and both are efficient vectors for WSMV. However, WCM Biotype 1 has been deemed a poor vector of HPWMoV and cannot transmit TriMV, but transmission efficiency increases with WSMV in mixed infection with both viruses (McMechan et al. 2014). In addition to this, a synergistic relationship between WSMV and TriMV has been reported (Tatineni et al. 2019).

BYDV and CYDV are wheat viruses from the family *Luteoviridae*, which are causal agents of the barley yellow dwarf disease (BYD) and are transmitted mainly by the bird cherry-oat aphid (*Rhopalosiphum padi*) (Bockus et al. 2001). Historically, the 20 year average yield loss caused by BYD is 1% in Kansas; however, these losses can increase up to 49% when conditions are conducive for the proliferation of *R. padi* (Appel et al. 2014; Gaunce and Bockus 2015). Here, we reported 8% and 9% infection for BYDV and CYDV, respectively. In a 2008 survey in Kansas, the BYDV and CYDV infection were at 6% and 2%, respectively (Burrows et al. 2009). However,

in a 2011-2012 survey, 32% of the samples were infected with BYDV and 2% of CYDV (Rotenberg et al. 2016). In contrast to the Kansas surveys, a survey of wheat viruses in Ohio has seen large numbers of infection of BYDV (67%) and CYDV (69%) in 2016-2017 samples (Hodge et al. 2020). The size of the aphid (*R. padi* and other species) populations is the key in the field distributions of BYDV and CYDV.

Additionally, two known fungal-associated viruses were also found: *Mitovirus* and *Puccinia striiformis totivirus* (PsV). *Mitovirus* belong to the family *Narnaviridae*. These viruses are considered the simplest out of all mycoviruses, containing a positive-sense, single-stranded RNA genome which encodes only the RdRp and strictly replicate in the mitochondria of the host (Hillman and Cai 2013). The low amino acid similarities between the *Mitovirus* identified in this study and other Mitoviruses indicate that this may be a novel *Mitovirus* with an unknown fungal host yet.

Another mycovirus identified in this study was PsV, previously reported from stripe rust (Zheng et al. 2017). PsV is part of the family *Totiviridae* and considered cryptic due to the infection often being asymptomatic (Zheng et al. 2017). Members of the family *Totiviridae* consists dsRNA genomes ranging from 4.6-7 kb and the genome contains two ORFs which encode for the RdRp and CP (Zheng et al. 2017; Ghadbrial et al. 2015). The genome of PsV is ~5 kb long and contains an overlapping ORF of RdRp and CP, a characteristic of Totiviruses (Zheng et al. 2017). Due to the cryptic nature of PsV, it has not been widely studied. However, this virus might have the potential to serve as a biological agent to control a devastating wheat pathogen- stripe rust.

In addition to known viruses, two novel viral sequences were also discovered in this study. These two sequences were always detected together in each positive RNA sample and shared high similarity of nucleotide sequences at the 5’ UTR suggesting that these viral sequences are two

segments of a new putative virus. We tentatively named this putative virus *Wheat-associated vipovirus* RNAs 1 and 2 (WaVPV), as a combination of *Virgaviridae* and *Potyviridae*, the families with the greatest similarity hits for the novel viral sequences. RT-PCR without the reverse transcriptase enzyme was utilized to verify that the novel viral sequences were not integrated into the host genome (data not shown). During viral replication of RNA viruses, the synthesis of the double-strand RNA (dsRNA), an intermediate during viral replication, trigger the immune response of the host which leads to the degradation of dsRNAs into small RNAs (Baulcombe 2004; Agrawal et al. 2003). For plants, the small RNAS range from 21-24 nt and fungal-associated viruses remain unclear but multiple small RNA sequencing studies have observed the range of 20 to 24 nt (Kuo and Falk 2020; Lee Marzano et al. 2018; Pooggin 2018; Ozkan et al. 2017).The small RNA sequencing results indicated that peak of the mapped read lengths were found to range from 20-21 nt with the highest peak at 21 nt. Although these results do not provide enough support to determine which host defense mechanism is triggered, the results confirm that WaVPV is able to elicit a response from the host's RNA silencing machinery and in turn, is able to replicate within its host.

The schematic genome organization of WaVPV is similar to the recently reported unclassified tobamo-like, fungal-associated viruses: *Plasmopara viticola lesion associated vivirus* (PvLaVivivirus1-4) (Chiapello et al. 2020). In comparison to the bipartite genome organization of RNA1 encoding RdRp and RNA2 encoding helicase of PvLaVivivirus1-4, WaVPV RNA1 and RNA2 also encode RdRp and helicase, respectively. However, WaVPV does not contain a conserved domain for methylase, which has been observed to be encoded by either RNA 1 or RNA 2 of the PvLaVivivirus 1-4. Unlike the previous studies of unclassified fungal-associated tobamo-like viruses, WaVPV RNA2 contains the dead-like helicase conserved domain sharing more

similarities to potyviruses infecting plants and had no significant similarity hits for tobamo-like viruses. Hence, and from an evolutionary standpoint, WaVPV is an interesting novel putative virus showing similarity into both plant and fungal viruses. Our evidence is not enough at this time to conclude that wheat is the primary or alternative host for WaVPV, and still its association with fungal hosts infecting wheat cannot be ruled out.

The evolutionary relationship between fungal and plant viruses has yet to be fully understood and presents complex and conflicting standpoints. Although, the close relationship and interaction of plant and fungal-associated viruses are still being studied, some breakthroughs contain evidence to support the possibility of cross-kingdom infections. Previous studies have observed that some fungal-associated viruses can replicate in plant protoplasts without the introduction of changes to the viral genome in order to adapt to the plant host (Nerva et al. 2017). On the other hand, a plant virus, *Cucumber mosaic virus,* was observed to be able to infect a fungal host, *Rhizoctonia solani* (Andika et al. 2017). Interestingly, a recent study has shown that a mixed-infection of a fungal-associated virus with a plant virus can aid in the replication of the plant virus in the fungal host and for the fungal-associated virus, the mixed infection aids in systemic infection of the plant host and infection of vegetative incompatible fungi of a different fungal species (Bian et al. 2020).

Cross-kingdom infections of plant and fungal-associated viruses bring more questions about the origins of plant and fungal viruses. Metagenomics and phylogenetic analyses have introduced different concepts to understand the evolutionary history of plant and fungal-associated viruses. In this study, the phylogenetic analysis of the polyprotein of WaVPV RNA 1 placed the segment in the sub-clade containing both recently discovered fungal-associated and insect-specific viruses from the *Virgaviridae* family, whereas WaVPV RNA 2 did not have any fungal-associated

virus similarity hits and was found to be in a close relationship with *Barley mild mosaic virus (*BaMMV) and other plant viruses belonging to the *Potyviridae* family. These results suggest that each segment was obtained from viruses infecting different kingdoms and may have been a result of genetic exchange between distinct plant and fungal-associated viruses.

Reassortment is a type of genetic exchange and one of the main evolutionary mechanisms utilized by segmented RNA viruses (Marshall et al. 2013). During mixed-infections of multiple viruses, RNA segments may be exchanged between different viruses to produce novel viruses (Vijaykrishna et al. 2015). The limitations of reassortment has been accepted to be between only viruses with segmented RNAs and these segmented RNA viruses must be closely related enough to be of the same viral species; however, there have been reports of inter-kingdom genetic exchange and reassortment between different viral families leading to the creation of novel viruses (Rastgou et al. 2014; Chiapello et al. 2019). The first report of a novel virus origin through inter-kingdom genetic exchange was in the discovery and characterization of a plant virus, *Ourmia melon virus* (OuMV) (Rastgou et al 2014). OuMV contained three viral segments: the RdRp containing homologues of fungal viruses from *Narnaviridae,* the movement protein (MP) containing similarities with plant viruses from the *Tombusviridae,* and the coat protein (CP) showing close relationships with viruses infecting plants and animals (Rastgou et al. 2014). Another study conducted to characterize plant viruses in grapevine samples have also discovered a novel virus, *Grapevine associated jivivirus*, which has been suggested to be result from the reassortment of virga-like and flavi-like viruses (Chiapello et al. 2019). One viral segment from this study showed high similarities with insect, flavi-like viruses, suggesting that this may not only be reassortment from different viral families but also viruses infecting hosts of two different kingdoms, much like the novel viral segments found in this study.

Due to the similarities of the putative MP and putative CP viral sequences identified in this study with the virga-like and potyviruses, and also because these two viral sequences being found in the same libraries as WaVPV RNAs 1 and 2, there is a possibility that these novel viral sequences are other segments of the WaVPV genome (Figure S3-2). However, our current information is not sufficient to confirm this because obtaining the full sequence of both viral sequences through RACE is still ongoing. Interestingly, the ORF 3 of PvLaT-LV1 which had ~40% similarity with putative MP also had a significant hit for a dead-like helicase in CDD. If the putative MP is a segment of WaVPV, this would raise another evolutionary inquiry as to why a virus would need to encode two helicases.

Taken together, the findings of this study emphasize the importance of identifying and characterizing all viral communities associated with wheat. There are still many open questions and further experiments are needed to determine the natural host of WaVPV and whether or not this putative virus is a serious threat to the wheat industry. Our study also showed the high incidence of a new *Mitovirus* in Kansas wheat fields.

References

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., & Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol Plant Pathol*, *10*(4), 537–545. https://doi.org/10.1111/j.1364-3703.2009.00545.x

Agrawal, N., Dasaradhi, P. V. N., Mohmmed, A., Malhotra, P., Bhatnagar, R. K., & Mukherjee, S. K. (2003). RNA Interference: Biology, Mechanism, and Applications. *Microbiol Mol Biol Rev*, *67*(4), 657–685. https://doi.org/10.1128/mmbr.67.4.657-685.2003

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, *215*(3), 403–410. https://doi.org/10.1016/s0022-2836(05)80360-2

Andika, I. B., Wei, S., Cao, C., Salaipeth, L., Kondo, H., & Sun, L. (2017). Phytopathogenic fungus hosts a plant virus: A naturally occurring cross-kingdom viral infection. *Proc Natl Acad Sci*, *114*(46), 12267–12272. https://doi.org/10.1073/pnas.1714916114

Andrews, S. (2010). *FastQC a Quality Control Tool for High Throughput Sequence Data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Appel, J., DeWolf, E., Bockus, W., and Todd, T. (2014). *Preliminary 2014 Kansaswheat disease loss estimates. Kansas Cooperative Plant Disease Survey Re-port.* http://agriculture.ks.gov/docs/default-source/PP-Disease-Reports-2014/2014-ks-wheat-disease-loss-estimates.pdf

Baulcombe, D. (2004). RNA silencing in plants. *Nature*, *431*(7006), 356–363. https://doi.org/10.1038/nature02874

Bian, R., Andika, I. B., Pang, T., Lian, Z., Wei, S., Niu, E., Wu, Y., Kondo, H., Liu, X., & Sun, L. (2020). Facilitative and synergistic interactions between fungal and plant viruses. *Proc Natl Acad Sci*, *117*(7), 3779–3788. https://doi.org/10.1073/pnas.1915996117

Bockus, W. W., Appel, J. A., Bowden, R. L., Fritz, A. K., Gill, B. S., Martin,T. J., Sears, R. G., Seifers, D. L., Brown-Guedira, G. L., and Eversmeyer,M. G. (2001). Success stories: Breeding for wheat disease resistance in Kansas. *Plant Dis, 85*,453-461.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Burrows, M., Franc, G., Rush, C., Blunt, T., Ito, D., Kinzer, K., Olson,J., O'Mara, J., Price, J., Tande, C., Ziems, A., and Stack, J. (2009).Occurrence of viruses in wheat in the Great Plains region, 2008. Online. *Plant Health Prog* doi:10.1094/PHP-2009-0706-01-RS

Byamukama, E., Seifers, D. L., Hein, G. L., De Wolf, E., Tisserat, N. A., Langham, M. A. C., Osborne, L. E., Timmerman, A., & Wegulo, S. N. (2013). Occurrence and Distribution of Triticum mosaic virus in the Central Great Plains. *Plant Dis*, *97*(1), 21–29. https://doi.org/10.1094/pdis-06-12-0535-re

Chiapello M., Rodriguez-Romero J., Nerva L., Forgia M., Chitarra W., Ayllon M.A., Turina M. (2019). Putative new plant viruses associated with Plasmopara viticola infected grapevine samples. *Ann Appl Biol, 176,* 180-191.

Chiapello, M., Rodríguez-Romero, J., Ayllón, M. A., & Turina, M. (2020). Analysis of the virome associated to grapevine downy mildew lesions reveals new mycovirus lineages. *Virus Evol*, *6*(2). https://doi.org/10.1093/ve/veaa058

Creager, A. N. H., Scholthof, K.-B. G., Citovsky, V., & Scholthof, H. B. (1999). Tobacco Mosaic Virus: Pioneering Research for a Century. *The Plant Cell*, *11*(3), 301–308. https://doi.org/10.1105/tpc.11.3.301

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Ghabrial, S. A., Castón, J. R., Jiang, D., Nibert, M. L., & Suzuki, N. (2015). 50-plus years of fungal viruses. *Virology*, *479-480*, 356–368. https://doi.org/10.1016/j.virol.2015.02.034

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hillman, B. I., & Cai, G. (2013). The family narnaviridae: simplest of RNA viruses. *Adv Virus Res*, *86*, 149–176. https://doi.org/10.1016/B978-0-12-394315-6.00006-4

Hodge, B. A., Paul, P. A., & Stewart, L. R. (2020). Occurrence and High-Throughput Sequencing of Viruses in Ohio Wheat. *Plant dis*, *104*(6), 1789–1800. https://doi.org/10.1094/PDIS-08-19-1724-RE

Hollandbeck GF, DeWolf E, and Todd T. (2017). *Kansas cooperative plant disease survey report preliminary 2017 Kansas wheat disease loss estimates.* https://agriculture.ks.gov/docs/default-source/pp-disease-reports-2012/2017-ks-wheat-disease-loss-estimates.pdf?sfvrsn=d49587c1_0

Howe, K. L., Contreras-Moreira, B., Nishadi, Maslen, G., Akanni, W., Allen, J., Alvarez-Jarreta, J., Barba, M., Bolser, D. M., Cambell, L., Carbajo, M., Chakiachvili, M., Christensen, M., Cummins, C., Cuzick, A., Davis, P., Fexova, S., Gall, A., George, N., … Flicek, P. (2020). Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res*, *48*(D1), D689–D695. https://doi.org/10.1093/nar/gkz890

Jeong, J.-J., Ju, H.-J., & Noh, J. (2014). A Review of Detection Methods for the Plant Viruses. *Res Plant Dis*, *20*(3), 173–181. https://doi.org/10.5423/rpd.2014.20.3.173

Jo, Y., & Cho, W. K. (2020). *Identification of viruses belonging to the family Partitiviridae from plant transcriptomes*. https://doi.org/10.1101/2020.03.11.988063

Käfer, S., Paraskevopoulou, S., Zirkel, F., Wieseke, N., Donath, A., Petersen, M., Jones, T. C., Liu, S., Zhou, X., Middendorf, M., Junglen, S., Misof, B., & Drosten, C. (2019). Re-assessing the diversity of negative strand RNA viruses in insects. *PLOS Pathog*, *15*(12), e1008224. https://doi.org/10.1371/journal.ppat.1008224

Kumar, S. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*, *5*(2), 150–163. https://doi.org/10.1093/bib/5.2.150

Kuo, Y.-W., & Falk, B. W. (2020). RNA interference approaches for plant disease control. *Biotechniques, 69*(6), 469–477. https://doi.org/10.2144/btn-2020-0098

Lee Marzano, S.-Y., Neupane, A., & Domier, L. (2018). Transcriptional and Small RNA Responses of the White Mold Fungus Sclerotinia sclerotiorum to Infection by a Virulence-Attenuating Hypovirus. *Viruses, 10*(12), 713. https://doi.org/10.3390/v10120713

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., & Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res*, *43*(D1), D222–D226. https://doi.org/10.1093/nar/gku1221

Marshall, N., Priyamvada, L., Ende, Z., Steel, J., & Lowen, A. C. (2013). Influenza Virus Reassortment Occurs with High Frequency in the Absence of Segment Mismatch. *PLoS Pathog*, *9*(6), e1003421. https://doi.org/10.1371/journal.ppat.1003421

Marzano, S.-Y. L., Nelson, B. D., Ajayi-Oyetunde, O., Bradley, C. A., Hughes, T. J., Hartman, G. L., Eastburn, D. M., & Domier, L. L. (2016). Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the Viromes of Five Major Fungal Plant Pathogens. *J Virol*, *90*(15), 6846–6863. https://doi.org/10.1128/jvi.00357-16

Matsumura, E., Coletta-Filho, H., Nouri, S., Falk, B., Nerva, L., Oliveira, T., Dorta, S., & Machado, M. (2017). Deep Sequencing Analysis of RNAs from Citrus Plants Grown in a Citrus Sudden Death-Affected Area Reveals Diverse Known and Putative Novel Viruses. *Viruses*, *9*(4), 92. https://doi.org/10.3390/v9040092

Mcmechan, A. J., Tatineni, S., French, R., & Hein, G. L. (2014). Differential Transmission of Triticum mosaic virus by Wheat Curl Mite Populations Collected in the Great Plains. *Plant Dis*, *98*(6), 806–810. https://doi.org/10.1094/pdis-06-13-0582-re

Monier, A., Claverie, J.-M., & Ogata, H. (2008). Taxonomic distribution of large DNA viruses in the sea. *Genome Biol*, *9*(7), R106. https://doi.org/10.1186/gb-2008-9-7-r106

Mumo, N. N., Mamati, G. E., Ateka, E. M., Rimberia, F. K., Asudi, G. O., Boykin, L. M., Machuka, E. M., Njuguna, J. N., Pelle, R., & Stomeo, F.. (2020). Metagenomic Analysis of Plant Viruses Associated With Papaya Ringspot Disease in Carica papaya L. in Kenya. *Front Microbiol, 11.* https://doi.org/10.3389/fmicb.2020.00205

Murugan, M., Sotelo Cardona, P., Duraimurugan, P., Whitfield, A. E., Schneweis, D., Starkey, S., & Smith, C. M. (2011). Wheat curl mite resistance: interactions of mite feeding with wheat streak mosaic virus infection. *J Econ Entomol*, *104*(4), 1406–1414. https://doi.org/10.1603/ec11112

Nerva, L., Varese, G. C., Falk, B. W., & Turina, M. (2017). Mycoviruses of an endophytic fungus can replicate in plant cells: evolutionary implications. *Sci Rep*, *7*(1). https://doi.org/10.1038/s41598-017-02017-3

Niu, J., Li, X-L., Wu, Y-L., Sun, Q-Z., Zhang, W., Cao, M., & Wang, J-J. (2020). RNA virome screening in diverse but ecologically related citrus pests reveals potential virus-host interactions. *J Invertebr Pathol, 170*, 1-8.

Nygren, J., Shad, N., Kvarnheden, A., & Westerbergh, A.. (2015). Variation in Susceptibility to Wheat dwarf virus among Wild and Domesticated Wheat. *PLOS ONE, 10*(4), e0121580. https://doi.org/10.1371/journal.pone.0121580

Olmedo-Velarde, A., Park, A. C., Sugano, J., Uchida, J. Y., Kawate, M., Borth, W. B., Hu, J. S., & Melzer, M. J. (2019). Characterization of Ti Ringspot-Associated Virus, a Novel Emaravirus Associated with an Emerging Ringspot Disease of Cordyline fruticosa. *Plant Dis*, *103*(9), 2345–2352. https://doi.org/10.1094/pdis-09-18-1513-re

Otuka, A. (2013). Migration of rice planthoppers and their vectored re-emerging and novel rice viruses in East Asia. *Front Microbiol*, *4*. https://doi.org/10.3389/fmicb.2013.00309

Özkan, S., Mohorianu, I., Xu, P., Dalmay, T., & Coutts, R. H. A. (2017). Profile and functional analysis of small RNAs derived from Aspergillus fumigatus infected with double-stranded RNA mycoviruses. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-3773-8

Pooggin, M. M. (2018). Small RNA-Omics for Plant Virus Identification, Virome Reconstruction, and Antiviral Defense Characterization. *Front Microbiol*, *9*. https://doi.org/10.3389/fmicb.2018.02779

Rambaut, A. (2014). *Figtree, a graphical viewer of phylogenetic trees*. http://tree.bio.ed.ac.uk/software/figtree

Ramos-González, P., Chabi-Jesus, C., Guerra-Peraza, O., Breton, M., Arena, G., Nunes, M., Kitajima, E., Machado, M., & Freitas-Astúa, J. (2016). Phylogenetic and Molecular Variability Studies Reveal a New Genetic Clade of Citrus leprosis virus C. *Viruses*, *8*(6), 153. https://doi.org/10.3390/v8060153

Rastgou, M., Habibi, M. K., Izadpanah, K., Masenga, V., Milne, R. G., Wolf, Y. I., Koonin, E. V., & Turina, M. (2009). Molecular characterization of the plant virus genus Ourmiavirus and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *J Gen Virol*, *90*(10), 2525–2535. https://doi.org/10.1099/vir.0.013086-0

Roossinck, M. J., Martin, D. P., and Roumagnac, P. (2015). Plant virus metagenomics: advances in virus discovery. *Phytopathol, 105*, 716–727. doi: 10.1094/phyto-12-14-0356-rvw

Rotenberg, D., Bockus, W. W., Whitfield, A. E., Hervey, K., Baker, K. D., Ou, Z., Laney, A. G., De Wolf, E. D., & Appel, J. A. (2016). Occurrence of Viruses and Associated Grain Yields of Paired Symptomatic and Nonsymptomatic Tillers in Kansas Winter Wheat Fields. *Phytopathol*, *106*(2), 202–210. https://doi.org/10.1094/phyto-04-15-0089-r

Shi, M., Lin, X. D., Chen, X., Tian, J. H., Chen, L. J., Li, K., Wang, W., Eden, J. S., Shen, J. J., Liu, L., Holmes, E. C., & Zhang, Y. Z. (2018). The evolutionary history of vertebrate RNA viruses. *Nature*, *556*(7700), 197–202. https://doi.org/10.1038/s41586-018-0012-7

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C., & Zhang, Y. Z. (2016). Redefining the invertebrate RNA virosphere. *Nature*, *540*(7634), 539–543. https://doi.org/10.1038/nature20167

Singh, K., Jarošová, J., Fousek, J., Chen, H., & Kundu, J. K. (2020). Virome identification in wheat in the Czech Republic using small RNA deep sequencing. *J Integr Agr*, *19*(7), 1825–1833. https://doi.org/10.1016/s2095-3119(19)62805-4

Zheng, L., Lu, X., Liang, X., Jiang, S., Zhao, J., Zhan, G., Liu, P., Wu, J., & Kang, Z. (2017). Molecular Characterization of Novel Totivirus-Like Double-Stranded RNAs from Puccinia striiformis f. sp. tritici, the Causal Agent of Wheat Stripe Rust. *Front Microbiol*, *8*. https://doi.org/10.3389/fmicb.2017.01960

Chapter 4 - Summary and conclusions

WSM is a devastating, global wheat viral disease with three documented viruses associated with the disease including WSMV, TriMV and HPWMoV. To date, resistant varieties are deployed in Kansas fields in order to manage the disease; however, resistant-breaking isolates have already been discovered which makes it very important to understand the current status of WSM in Kansas fields. On the other hand, the current knowledge about other viral communities associated with wheat is limited as the wheat virome studies have only focused on known wheat viruses.

Through this study, we provided a current survey of WSM, which have been consistent with previous surveys of WSM in Kansas and the Great Plains. WSMV is still the dominant virus found in the field, followed by the mixed infection of WSMV + TriMV, and only one occurrence of the mixed infection of all three viruses. There have been no single infections of TriMV or HPWMoV detected, which supports the declining occurrence of the single infections of these viruses observed from previous surveys. The relationships between virus-vector, virus-virus, and the plant-virus may have contributed to the distribution of WSM across Kansas. In addition to an updated survey, the first evolutionary study of WSMV and TriMV using the whole genome was also obtained in this study. The genetic variation was low for both viruses which is consistent with the concept that genetic stability is the rule in natural plant virus populations (Garcia-Arenal et al. 2001). Through the use of complete genome sequences, we observed that recombination is a major evolutionary force in WSMV natural populations but not in TriMV. The previously identified recombination sites in CP was found in only 3 recombinants in this study, which is less than the other proteins deemed as recombination hotspots (P1, HC-Pro, CP, and CI). The phylogenetic analysis of WSMV also showed that Kansas isolates are widely distributed among clade D. The

mixed results of the clustering of some historical isolates with 2019 isolates and separate groupings of other 2019 isolates from historical isolates further support the inference that WSMV Kansas isolates are diverse in the fields. The population is genetically stable with a purifying selection acting upon all parts of the genome; however, one positive selection was found in the RNA silencing suppressor protein- P1.

This study also provided the first account of the analysis of the evolutionary mechanisms of TriMV. The separate clustering of two potential resistant breaking isolates in the phylogenetic analysis suggests that genetic changes may be placed upon TriMV to overcome resistant varieties. However, the sample size in this study was low so further studies with a greater sample size is needed to support that hypothesis. Although the major selection pressure acting upon TriMV is negative, there has been one site of positive selection found in the NIb protein. The NIb encoding region also showed the greatest nucleotide diversity among all TriMV genome encoding regions. This result was interesting due to the fact that a consensus of previous studies has presumed that NIb is highly conserved and is one of the potyviral proteins which should contain little to no positive selection sites. We hypothesize that the co-infection of TriMV with WSMV may lead to the increase pressure to introduce changes in the NIb protein which has a strong interaction with host factors and plays an important role in the virus replication.

The complete genome of 8 RNA segments of HPWMoV was obtained in this study, which is the second Kansas isolate to be completely sequenced. Although no evolutionary studies were conducted due to the low available samples of HPWMoV in the NCBI GenBank, the whole genome sequences obtained in this study will be submitted in the NCBI database to be used in future studies. Through conducting evolutionary studies on WSMV and TriMV using the whole

genome sequence, this study achieved to gain an updated and more informative snapshot of the status of the WSM viral populations in Kansas.

The results of the metagenomics analysis of the wheat virome showed that a potential novel *Mitovirus* is found to be the second more prevalent after WSMV. A totivirus, PsV, infecting stripe rust was also found in samples. In addition to the presence of known wheat and fungal-associated viruses, a putative new segmented virus named WaVPV was also found. This novel putative virus brings about interesting questions about its origin and evolution. One segment is closely related to fungal-associated and insect-specific viruses, whereas the second segment is closely related to plant viruses. To date, this may be the third observation of a segmented virus potentially originating from the genetic exchange of two viral segments infecting hosts of two different kingdoms. However, further studies are necessary to validate this claim. Whether wheat is the primary or the alternative host of this novel putative virus is not known at this time and it remains for further investigations.

Appendix A - Supplementary Data for Chapter 2

Figure S2-1. The SimPlot Analysis results using the SimPlot program. The y-axis shows the nucleotide percent similarity between the query sequence (recombinant) and the reference sequence (major and minor parents shown in the box on the top right). The x-axis depicts the nucleotide position and above this, the schematic of the WSMV genome organization is shown. A crossover between the two references illustrates a recombination breakpoint, shown by the red vertical lines.

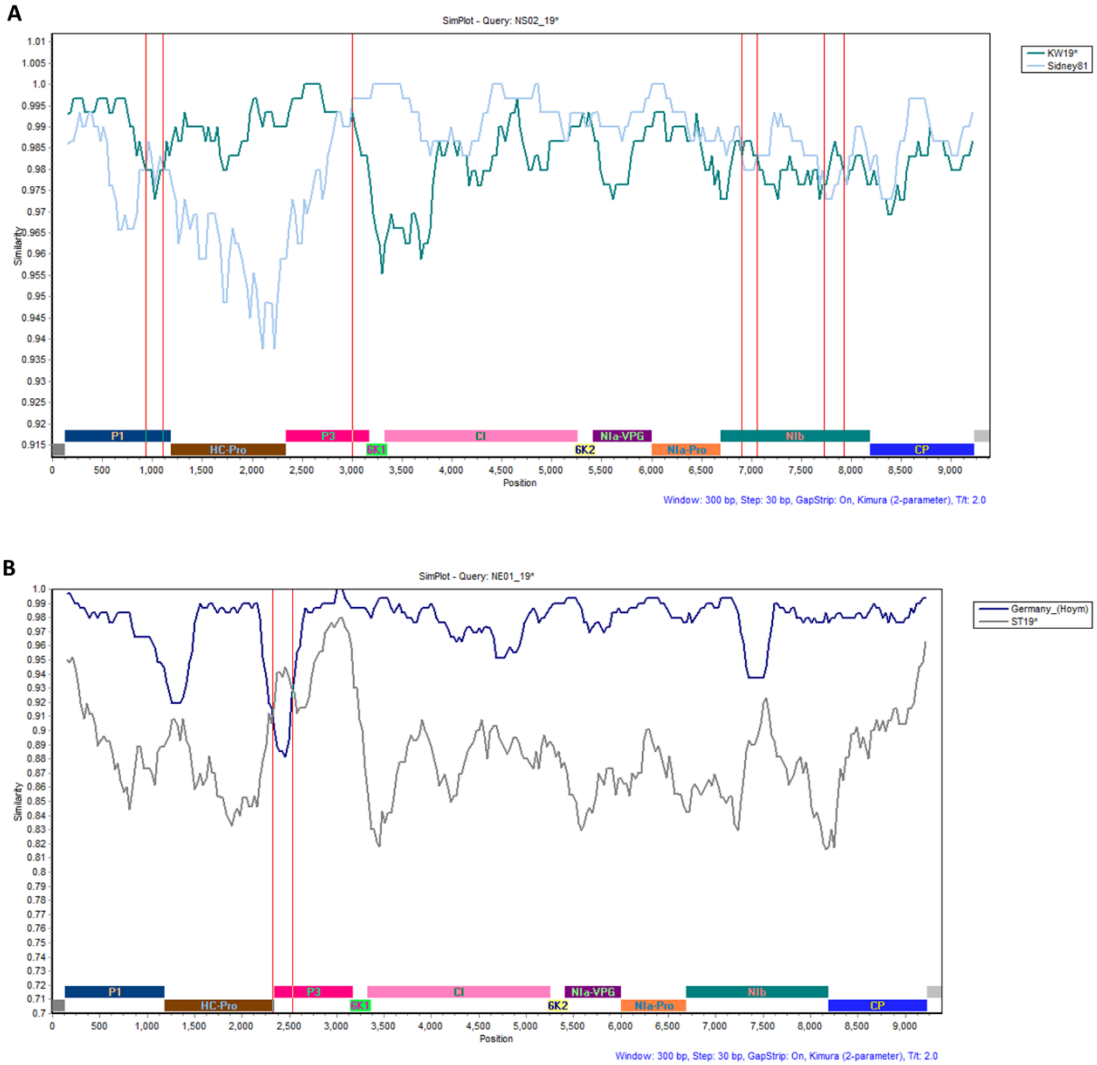
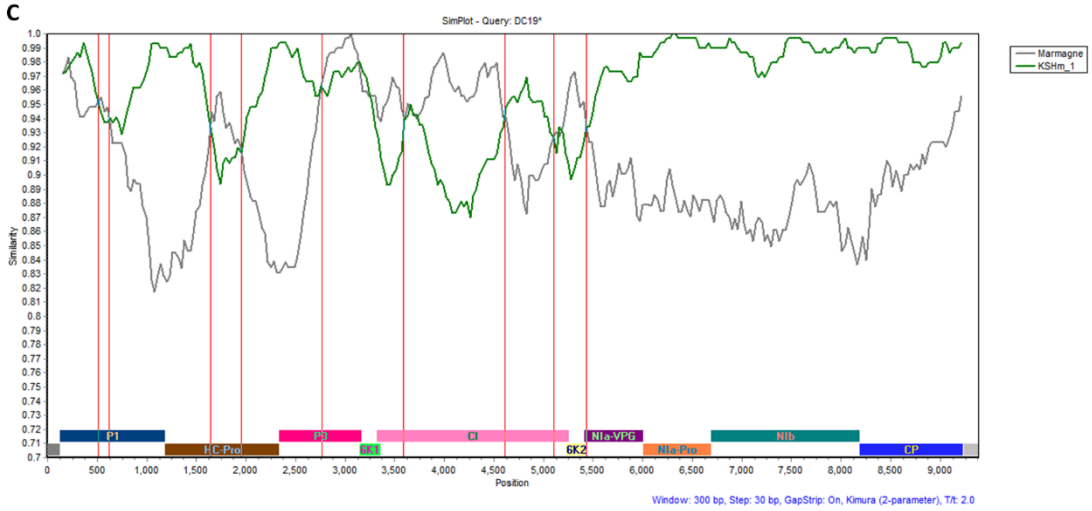
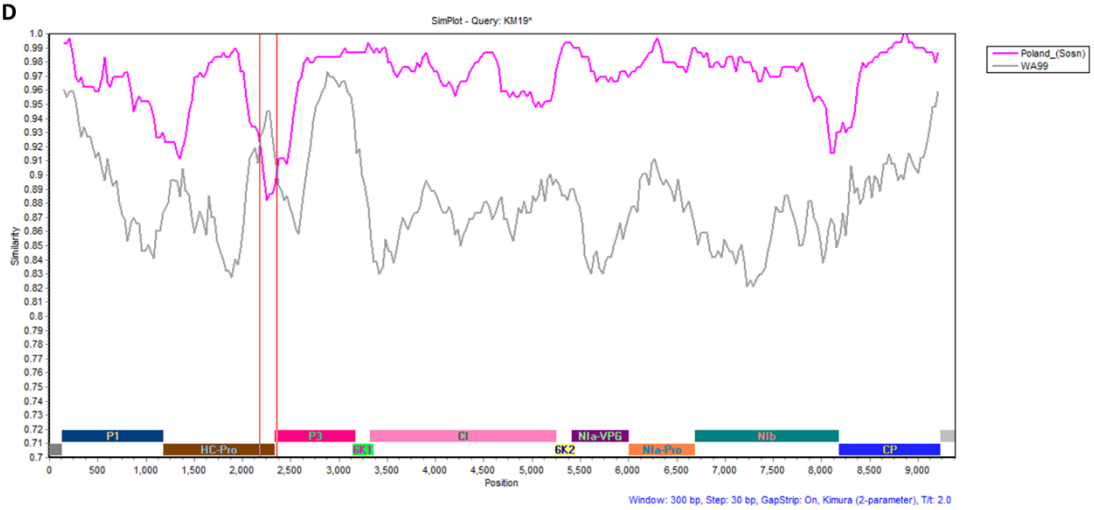


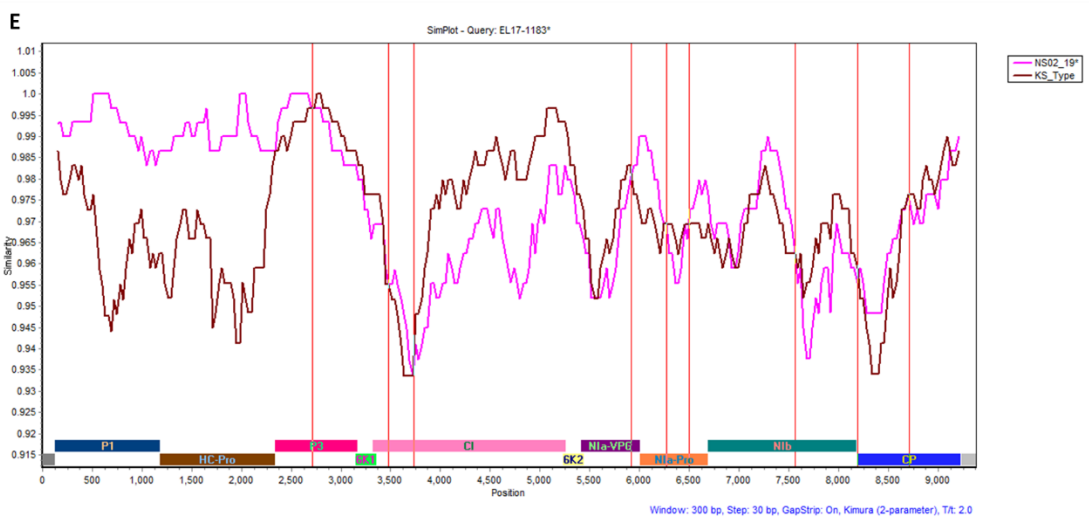
Figure S2- 1. (Continued).

C

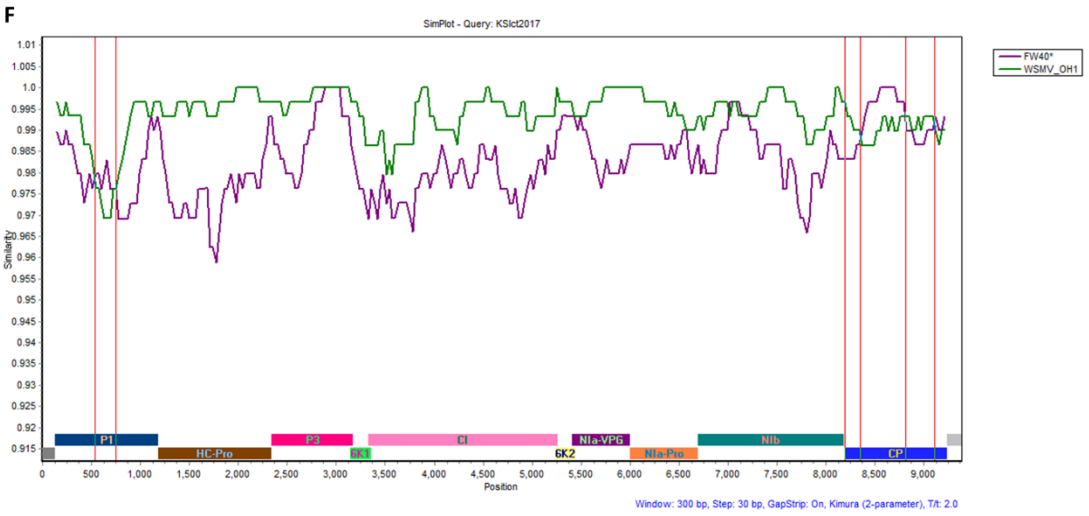


D





Figures S2-1. (Continued).



G

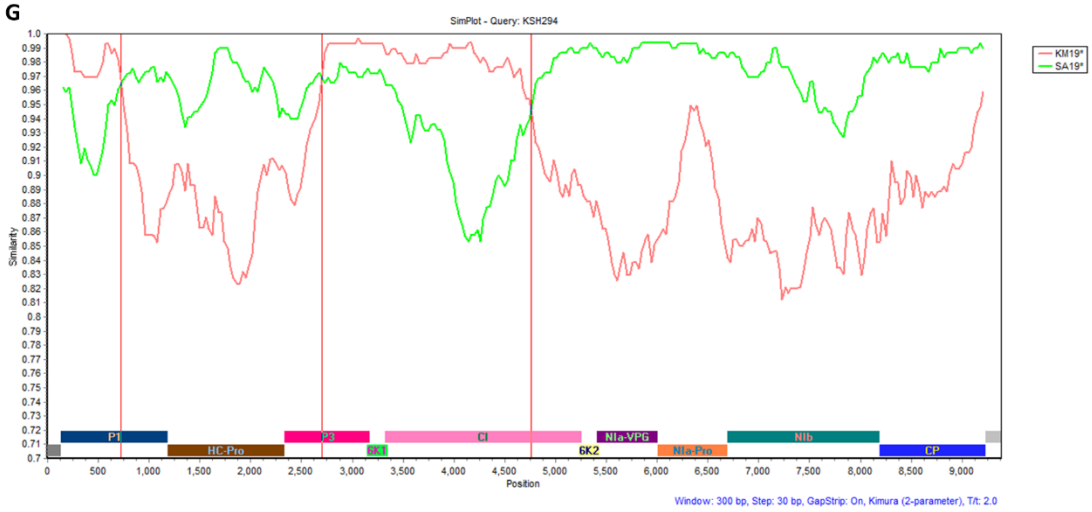

SimPlot - Query: KSH294
KM19*
SA19*
Similarity
Position
P1
HC-Pro
P3
6K1
CI
6K2
NIa-VPG
NIa-Pro
NIb
CP
Window: 300 bp, Step: 30 bp, GapStrip: On, Kimura (2-parameter), T/t: 2.0

Table S2-1. List of all primers used in this study.

Target	Primer Name	Product Size (bp)	Sequences
WSMV	WSMV_Primer3_F	1947	TGAGAAACCACCGAGTTATGACGGC
WSMV	WSMV_Primer3_R		TCCTGGTGCACTGAGAAGTTTGCTG
WSMV	WSMV_Primer5_F	1698	CGAGTGGTTCATTGACGCTGATGGTTC
WSMV	WSMV_Primer5_R		GCCTCTCGTGGAGAAGTACACACCTAG
TriMV	TriMV_Primer3_F	1994	TTGGGAAGCTTCTCAACGAAGGAGAAG
TriMV	TriMV_Primer3_R		TCCGGTAACACTTCCTGAGCCTCGA
TriMV	TriMV_Primer5_F	1953	AGCGAGGGCCATTTATTTGACTGGT
TriMV	TriMV_Primer5_R		AACGTCAGTTTGAAATGCTTGCATGCT
HPMV	HPMV_RNA1_F	2006	CAGTCAAACAGGTCCAATGAATCAGACTCT
HPMV	HPMV_RNA1_R		AGGGGATATGAGCTAAACAGCACAAAGCT

Table S2-2. List of all reference genomes used for mapping.

Sample ID:	**Reference Genomes of:**	**Accession Number:**
WSMV	*Wheat streak mosaic virus*	NC_001886.1
TriMV	*Triticum mosaic virus*	NC_012799.1
HPMV_RNA1	*High plains mosaic virus*	NC_029570.1
HPMV_RNA2	*High plains mosaic virus*	NC_029549.1
HPMV_RNA3	*High plains mosaic virus*	NC_029551.1
HPMV_RNA4	*High plains mosaic virus*	NC_029551.1
HPMV_RNA5	*High plains mosaic virus*	NC_029552.1
HPMV_RNA6	*High plains mosaic virus*	NC_029553.1
HPMV_RNA7	*High plains mosaic virus*	NC_029554.1
HPMV_RNA8	*High plains mosaic virus*	NC_029555.1

Table S2-3. List of sequences retrieved from the GenBank.

Sample ID	Complete Genome of	Country of Origin	Accession Number
El Batan	*Wheat streak mosaic virus*	Mexico	AF285170.1
Austria	*Wheat streak mosaic virus*	Austria	LN624217.1
Marmagne	*Wheat streak mosaic virus*	France	HG810953.1
Sosn	*Wheat streak mosaic virus*	Poland	MH939146.1
Czech	*Wheat streak mosaic virus*	Czech	AF454454.1
Hoym	*Wheat streak mosaic virus*	Germany	HG810954.1
Iran	*Wheat streak mosaic virus*	Iran	EU914917.1
[3]H98	*Wheat streak mosaic virus*	KS, USA	AF511615.2
Mon96	*Wheat streak mosaic virus*	MT, USA	AF511630.2
WA99	*Wheat streak mosaic virus*	WA, USA	AF511643.2
WA94	*Wheat streak mosaic virus*	WA, USA	FJ348358.1
ID96	*Wheat streak mosaic virus*	ID, USA	AF511618.2
Argentina	*Wheat streak mosaic virus*	Argentina	FJ348359.1
Turkey	*Wheat streak mosaic virus*	Turkey	AF454455.1
[3]KS_Type	*Wheat streak mosaic virus*	KS, USA	AF285169.1
KSGre2017	*Wheat streak mosaic virus*	KS, USA	MK318275.1
ID99	*Wheat streak mosaic virus*	ID, USA	AF511619.2
Sidney 81	*Wheat streak mosaic virus*	NE, USA	AF057533.1
H95S	*Wheat streak mosaic virus*	KS, USA	AF511614.2
KSHm1	*Wheat streak mosaic virus*	KS, USA	MK318276.1
WSMV_OH1	*Wheat streak mosaic virus*	OH, USA	MK975887.1
[2]KSIct2017	*Wheat streak mosaic virus*	KS, USA	MK318279.1
[2]KSH294	*Wheat streak mosaic virus*	KS, USA	MF459661.1
[1]ONMV	*Oat necrotic mottle virus*	-	NC_005136.1
[1]YOgMV	*Yellow oat-grass mosaic virus*	-	NC_024471.1
KSGre2017	*Triticum mosaic virus*	KS, USA	MK318272.1
KSHm2015	*Triticum mosaic virus*	KS, USA	MK318273.1
KSICT2017	*Triticum mosaic virus*	KS, USA	MK318274.1
[3]U06-123	*Triticum mosaic virus*	KS, USA	FJ263671.1
NE	*Triticum mosaic virus*	NE, USA	FJ669487.1
[1]SCSMV	*Sugarcane streak mosaic virus*	-	NC_014037.1
[1]CalVA	*Caladenia virus A*	-	NC_018572.1

[1]These virus isolates were chosen as outgroups for the phylogenetic analysis.
[2] These recombinant isolates were removed from the phylogenetic studies.
[3]These isolates were included in the neutrality test and population genetics analyses.

Table S2-4. List of the complete genomes obtained from this study.

Sample Name	Complete Genome of	County, State
[3]RA02_19	*Triticum mosaic virus*	Rawlins, Kansas
[3]NS02	*Triticum mosaic virus*	Ness, Kansas
[3]EL19	*Triticum mosaic virus*	Ellis, Kansas
[3]NT19	*Triticum mosaic virus*	Norton, Kansas
[3]DC19	*Triticum mosaic virus*	Decatur, Kansas
[3]SH02_19	*Triticum mosaic virus*	Sherman, Kansas
[3]SM19	*Triticum mosaic virus*	Smith, Kansas
[3] RA02_19	*Wheat streak mosaic virus*	Rawlins, Kansas
[2]NS02_19	*Wheat streak mosaic virus*	Ness, Kansas
[3]EL19	*Wheat streak mosaic virus*	Ellis, Kansas
[3]RP19	*Wheat streak mosaic virus*	Republic, Kansas
[3]FO19	*Wheat streak mosaic virus*	Ford, Kansas
[2]NE01_19	*Wheat streak mosaic virus*	Cheyenne, Nebraska
[2]DC19	*Wheat streak mosaic virus*	Decatur, Kansas
[2]KM19	*Wheat streak mosaic virus*	Kingman, Kansas
[3]KW19	*Wheat streak mosaic virus*	Kiowa, Kansas
[3]SM19	*Wheat streak mosaic virus*	Smith, Kansas
[3]SH02	*Wheat streak mosaic virus*	Sherman, Kansas
[3]ME19	*Wheat streak mosaic virus*	Meade, Kansas
[3]ST19	*Wheat streak mosaic virus*	Stanton, Kansas
[3]SA19	*Wheat streak mosaic virus*	Saline, Kansas
FW40	*Wheat streak mosaic virus*	Colorado
[1,3]PL95	*Wheat streak mosaic virus*	Phillips, Kansas (1995)
[1,3]RL71	*Wheat streak mosaic virus*	Riley, Kansas (1971)
[1,2]EL17-1183	*Wheat streak mosaic virus*	Ellis, Kansas (2017)
[1,3]EL06-892	*Wheat streak mosaic virus*	Ellis, Kansas (2006)
[1,3]FO93	*Wheat streak mosaic virus*	Ford, Kansas (1993)

[1]These isolates are the historical WSMV samples.
[2]These recombinant isolates were removed from the phylogenetic studies.
[3]These isolates were included in the neutrality test and population genetics analyses

Table S2-5. Results from the recombination analysis using the RDP5 program.

	RDP5 Methods						
Recombinants	**RDP**	**GENECONV**	**BootScan**	**MaxChi**	**Chimaera**	**SiScan**	**3Seq**
NS02	2.1×10^{-8}	5.8×10^{-4}	9.2×10^{-8}	4.4×10^{-9}	4.4×10^{-9}	3.1×10^{-11}	1.2×10^{-10}
NE01	6.9×10^{-19}	1.4×10^{-14}	5.2×10^{-19}	4.3×10^{-6}	7.7×10^{-6}	1.5×10^{-3}	6.9×10^{-12}
DC19	3.3×10^{-7}	3.6×10^{-5}	3.6×10^{-5}	9.6×10^{-8}	2.8×10^{-7}	6.9×10^{-4}	1.5×10^{-6}
KM19	7.2×10^{-15}	4.5×10^{-8}	5.6×10^{-15}	9.1×10^{-6}	8.3×10^{-6}	2.9×10^{-2}	1.1×10^{-7}
EL17-1183	1.2×10^{-3}	-	1.8×10^{-4}	6.7×10^{-11}	3.9×10^{-5}	1.5×10^{-18}	1.4×10^{-22}
KSIct2017	-	-	2.5×10^{-2}	8.4×10^{-3}	3.4×10^{-2}	1.3×10^{-3}	-
KSH294	5.1×10^{-52}	2.2×10^{-43}	8.1×10^{-51}	3.4×10^{-19}	2.8×10^{-20}	7.7×10^{-23}	1.9×10^{-43}

Table S2-6. The potential major and minor parents of WSMV recombinants detected by the RDP5 program.

Recombinants	Major Parent	Minor Parent
NS02	Sydney 81	KW19
NE01	Hoym	ST19
DC19	Marmagne	KSHm1
KM19	Sosn	WA99
EL17-1183	NS02	KS_Type
KSIct2017	FW40	WSMV_OH1
KSH294	SA19	KM19

Table S2-7. Codons under negative selection for Kansas WSMV isolates obtained from the Neutrality Test.

Significant Selection Methods	Sites
FEL, FUBAR	26, 43, 48, 57, 58, 81, 89, 92, 99, 103, 104,119, 139, 173, 186, 188, 192, 200, 201, 208
FEL, FUBAR, SLAC	81, 92, 99, 119, 173, 186, 188, 192, 200, 201, 208, 245, 253 258, 261, 268,280, 297,332, 364, 385, 399, 401, 427, 450, 458, 481, 485, 489, 491, 505, 529, 547, 565, 579589, 602, 608, 670, 673, 679, 685, 690, 701, 703, 723, 738, 742 , 744, 749 , 751 , 759 , 768 , 794 , 809, 822 , 853

Table S2-8. Codons under negative selection for Kansas TriMV isolates using the Neutrality Tests.

Significant Selection Methods	Sites
FEL, FUBAR	70, 181, 206, 346, 361, 417, 432, 610, 901, 953, 1347, 1419, 1461, 1502, 1669, 1763, 1771, 1812,1930, 19371968, 1994, 2069, 2093, 2129, 2190, 2320, 2347, 2384, 2392, 2431, 2507, 2514, 2540, 2574, 2590, 2672 , 2705, 2726, 2782, 2879, 2944

Appendix B - Supplementary Data for Chapter 3

Figure S3-1. Validation of viral sequences in the original RNAs. **M)** 1Kb plus DNA ladder; **A)** WSMV and the samples are as follows (1-17): FO19, NS02, RP19, EL19, RA02, NE01, SM19, SH02, ST19, SA19, KM19, KW19, DC19, FW40, ME19, NT 19, and KE19; **B)** TriMV and and the samples are as follows (1-7): EL19, NS02, RA02, DC19, NT19, SM19, and SH02; **C)** the only positive HPWMoV sample: RA02; **D1)** BYDV RdRp: KE19 and LE01; **D2)** BYDV CP: KE19 and LE01; **E1)** CYDV RdRp: EL19; **E2)** CYDV CP: EL19; **F)** *Mitovirus* for the following samples (1-7): GL01, KM19, NT19, ME19, LE01, KW19, and SA19; **G)** PsV for the following samples (1-6): KW19, ME19, GL01, NS02, and FO19; **H**) WaVPV RNA1 for the following samples (1-6): GL01, ME19, KW19, EL19, FO19, and NS02; **I**) WaVPV RNA2 for the following samples (1-6): GL01, ME19, EL19, KW19, NS02, and FO19.

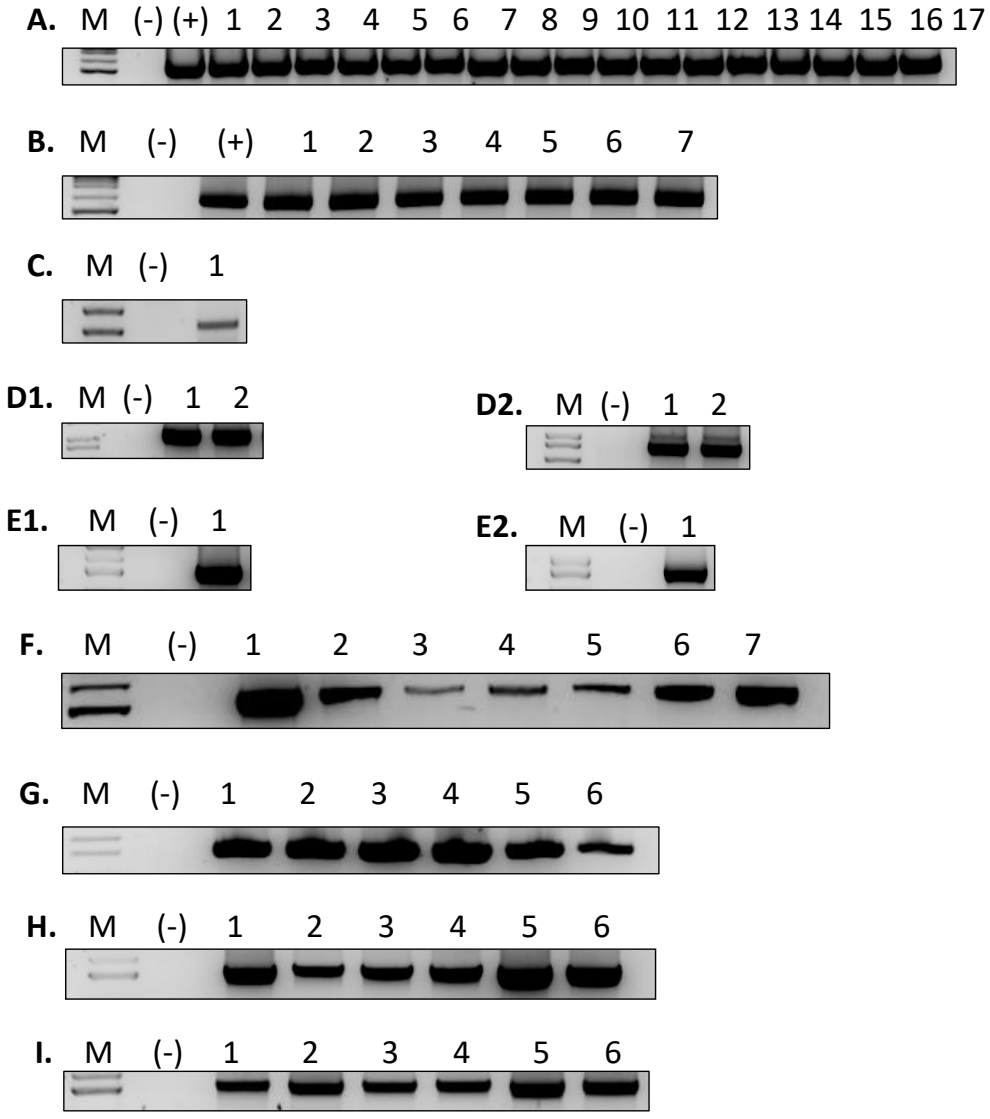


Figure S3-2. The schematic of the genome organization of the incomplete viral fragments: Putative MP and CP.

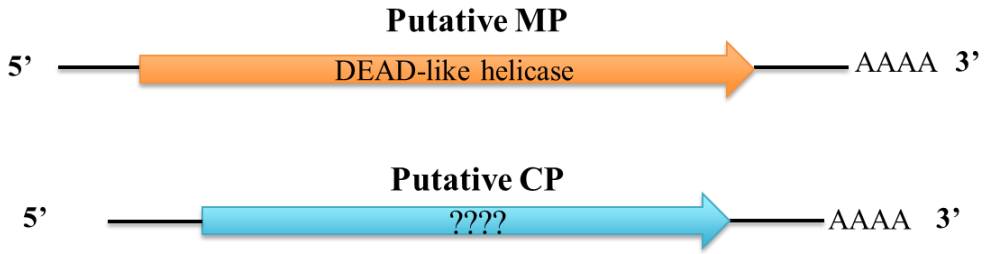


Table S3-1. List of RNA samples sequences in this study

Name	County, State
NT19	Norton, Kansas
RA02	Rawlins, Kansas
NS02	Ness, Kansas
EL19	Ellis, Kansas
RP19	Republic, Kansas
FO19	Ford, Kansas
DC19	Decatur, Kansas
KM19	Kingman, Kansas
KW19	Kiowa, Kansas
SM19	Smith, Kansas
SH02	Sherman, Kansas
ME19	Meade, Kansas
ST19	Stanton, Kansas
GL01	Greeley, Kansas
LE01	Lane, Kansas
JO19	Johnson, Kansas
CK19	Cherokee, Kansas
CM02	Comanche, Kansas
RL05	Riley, Kansas
DK03	Dickinson, Kansas
KE19	Kearney, Kansas
SA19	Saline, Kansas
FW40	Colorado
NE01	Cheyenne, Nebraska

Table S3-2. List of primers used in this study.

Target	Primer Name	Product Size (bp)	Sequences
WSMV	WSMV_Primer3_F	1947	TGAGAAACCACCGAGTTATGACGGC
WSMV	WSMV_Primer3_R		TCCTGGTGCACTGAGAAGTTTGCTG
TriMV	TriMV_Primer2_F	1944	TGCTATGTGGGTTAATGCCGTTTGC
TriMV	TriMV_Primer2_R		AAGGGGAACTTCACACATGTAATCGT
HPWMoV	HPWMoV_RNA1_F	547	CGACACTCCTCGGTGTTATGTAATC
HPWMoV	HPWMoV_RNA1_R		GGATATGAGCTAAACAGCACAAAGC
BYDV	BYDV_RdRp_F	1206	TTCTATCCCCACCGCCATCATGTTCTTC
BYDV	BYDV_RdRp_R		GTGCAGGGAGGCATTTGGTGCCTTATAT
BYDV	BYDV_CP_F	909	GTGAAGGTGACGACTCCACATCTGCAAT
BYDV	BYDV_CP_R		AGAATCTTTTGGGGACAGGTGCAGGAGT
CYDV	CYDV_RdRp_F	1749	CACAACTGGGAGCAGAACTCCTATATGC
CYDV	CYDV_RdRp_R		CACTTGCAACTTACGTTTGTCCGGCTAG
CYDV	CYDV_CP_F	905	TGAGTACGGTCGTCCTTAGATCCAATGG
CYDV	CYDV_CP_R		GGGACTGGAAAGACAAACATCGGAACTG
Mitovirus	Mito_F	1800	ACCTAACGGTCACGCCATAACATCCT
Mitovirus	Mito_R		CTCGAGACCACAGTTGATTGAGATGAC
Totivirus	Toti_F	905	CCC GAA TCC CGC TGC CGT AAT TAA ACA T
Totivirus	Toti_R		GTG ATT CTA CTG TTT CTG TCG CGG GTT G
POTY-R	WaVPV R2	1664	ATCGTGAGGTTTTGAGCGCTTCAAAGG
POTY-F	WaVPV R2		GAGTGACAACCTTTCTGAGTTCACCAC
Virga-R	WaVPV R1	1698	ACCATTGCAGGGATTACCCTCACGTGA
Virga-F	WaVPV R1		CGGATAAGCTACAACAGAGGGCTGTT

Table S3-3. Results from RNASeq Data.

	Number of Reads		
Sample name	**Raw Reads**	**Trimmed Reads**	**Unmapped Reads**
CK19	18681308	16830877	7296185
CM02_19	14899250	13477912	5891195
DK03_19	16339417	14661495	4405779
EL06_892	15269574	14045108	2984585
EL17_1183	20475459	18716341	7492151
EL19	15847609	14488245	6463206
FO19	17820060	16132868	7897039
FO93	20301134	18351596	7390188
JO19	14561026	13310425	2644781
NE01_19	17111311	15793422	9578710
NS02	21416060	19488674	10100980
PL95	21855323	19824265	8225088
RA02	15847278	13871932	5981577
RL05_19	15884242	14371057	5679442
RL71	19880128	18313625	3362382
RP19	15913757	14186259	5596479
DC19	23421520	19934286	1873823
FW40	28812361	24441297	4255230
GL01	24476910	20818023	11096006
KE19	21808767	18454922	4953301
KM19	23649954	20024188	8267987
KW19	26088658	21875135	7339108
LE01	30093691	25462991	12079643
ME19	21667049	18499913	9266606
NT19	23291475	19826852	7324039
SA19	23530569	19858459	7266210
SH02	24860180	21126771	10295076
SM19	30139534	25437254	8297632
ST19	28329930	23865519	12023448