PLANT VIRUSES, VOLUME I: DETECTION METHODS, GENETIC DIVERSITY AND EVOLUTION

EDITED BY: Akhtar Ali, Bright Agindotan, Xifeng Wang, Rajarshi Kumar Gaur, Xiaofei Cheng and Kristiina Mäkinen

PUBLISHED IN: Frontiers in Microbiology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-129-8
DOI 10.3389/978-2-88974-129-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PLANT VIRUSES, VOLUME I: DETECTION METHODS, GENETIC DIVERSITY AND EVOLUTION

Topic Editors:
Akhtar Ali, University of Tulsa, United States
Bright Agindotan, USDA APHIS Veterinary Services, United States
Xifeng Wang, State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection (CAAS), China
Rajarshi Kumar Gaur, Deen Dayal Upadhyay Gorakhpur University, India
Xiaofei Cheng, Northeast Agricultural University, China
Kristiina Mäkinen, University of Helsinki, Finland

Citation: Ali, A., Agindotan, B., Wang, X., Gaur, R. K., Cheng, X., Mäkinen, K., eds. (2022). Plant Viruses, Volume I: Detection Methods, Genetic Diversity and Evolution. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-129-8

Table of Contents

frontiers in Microbiology

EDITORIAL
published: 06 December 2021
doi: 10.3389/fmicb.2021.793071

Editorial: Plant Viruses, Volume I: Detection Methods, Genetic Diversity, and Evolution

Akhtar Ali[1], Rajarshi Kumar Gaur[2], Xifeng Wang[3], Xiaofei Cheng[4] and Kristiina Mäkinen[5]*

[1] Department of Biological Sceince, The University of Tulsa, Tulsa, OK, United States, [2] Department of Biotechnology, Deen Dayal Upadhyay Gorakhpur University, Gorakhpur, India, [3] State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection (CAAS), Beijing, China, [4] Key Laboratory of Germplasm Enhancement, Physiology and Ecology of Food Crops in Cold Region of Chinese Education Ministry, College of Agriculture, Northeast Agricultural University, Harbin, China, [5] Department of Microbiology, Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland

Keywords: high throughput sequences (HTS), novel viruses, diversity, recombination, complete genome sequences

Editorial on the Research Topic

Plant Viruses, Volume I: Detection Methods, Genetic Diversity, and Evolution

Both RNA and DNA plant viruses commonly infect agricultural crops worldwide and cause tremendous losses in crop yield. Infections are also common in non-cultivated host plants and trees. Over the last 40 years, the number of plant viruses identified worldwide from various plants has increased tremendously because of the availability of new diagnostic techniques.

Plant viruses are constantly emerging and new viruses are being identified at an accelerating rate worldwide. One of the main reasons for the emergence of new viruses is their constant evolution, which allows them to re-emerge as a new strain of the same virus or as a new virus to infect new hosts. Therefore, control measures for a specific plant virus will work for a short time but may not be effective for longer periods. Despite large numbers of evolutionary studies of plant viruses over the last 20 years, plant virologists are still in the early stages of understanding how these viruses evolve, interact with their hosts, and survive naturally in the wild.

In this Research Topic, we received 11 research articles, which cover topics about the identification and evolutionary analysis of new plant viruses from a variety of cultivated crops including wheat, barley, maize, pea, pepper, tomato, soybean, garlic, and Zanthoxylum armatum (a spiny shrub).

The availability of high throughput sequences (HTS) technique has made identification of plant viruses regarding their genome from both domesticated and wild plants much faster and easier than previously described methods. A recent study by Kondo et al. was focused on yellow mosaic disease (YMD), which infects winter wheat in Asian countries, specifically Japan. YMD is usually attributed to the infection of bymovirus or furovirus. In their work, they identified a new virus associated with YMD of winter wheat in Hokkaido, Japan. This new virus was named wheat virus Q (WVQ), which belongs to the subfamily *Quinvirinae* (family *Betaflexiviridae*). WVQ is potentially transmitted through soil to wheat and rye plants. The virus infection also induces antiviral RNA silencing responses in the host.

Wheat is commonly infected by wheat streak mosaic (WSM) disease complex that contains three 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. In our Research Topic, Redila et al. presented the genetic variation based on the complete genome sequences among the isolates of three viruses associated with WSM, which were collected from wheat fields located in

OPEN ACCESS

Edited and reviewed by:
Linqi Zhang,
Tsinghua University, China

****Correspondence:***
Akhtar Ali
akhtar-ali@utulsa.edu

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *11 October 2021*
Accepted: *09 November 2021*
Published: *06 December 2021*

Citation:
Ali A, Gaur RK, Wang X, Cheng X and Mäkinen K (2021) Editorial: Plant Viruses, Volume I: Detection Methods, Genetic Diversity, and Evolution. Front. Microbiol. 12:793071. doi: 10.3389/fmicb.2021.793071

Kansas or Great Plains and analyzed by HTS. The results suggest the association of all three viruses with WSM disease and interestingly low genetic diversity in the natural populations of WSMV and TriMV.

Yellow dwarf viruses (YDVs; ssRNA viruses in the family *Luteoviridae*) are the most widespread group of cereal viruses worldwide and are transmitted by more than 25 aphid species in a persistent non-propagative manner. A study on barley/cereal yellow dwarf viruses using HTS by Somera et al. describes the complete genome sequences for 22 isolates of CYDV-RPS, BYDV, GAV, BYDV-PAS, BYDV-PAV, and BYDV-OYV from symptomatic plants collected from 47 cereal (wheat, barley, oat, rye, and triticale) fields in Estonia and meadow fescue in Sweden. In addition, they assembled a nearly complete genome of a novel species of BYDV-OYV from Swedish samples. Recombination analysis of the virus sequences revealed that BYDV-OYV is the parental virus for BYDV-PAV and BYDV-PAS, which further improved our knowledge about the complexity of viruses associated with YDVs.

Maize and sorghum are staple food crops and are infected by diverse pathogens to reduce their marketable yield. Maize strip virus (MSpV, Tenuivirus) infects maize, but the complete genome sequence has not been available so far. The HTS work presented by Bolus et al. reports the first complete reference genome sequence of MSpV and those of additional virus isolates collected from maize in the United States and Papua New Guinea. From these sequences, they determined genetic diversity.

Another metagenomics study on green peas (*Pisum sativum*) virome in Germany using rRNA-depleted total RNA extracts from pooled plant tissue in combination with HTS revealed 35 viruses and 9 associated nucleic acids (Gaafar et al.). Bioinformatics analysis of virus sequences and further confirmation by reverse transcription-polymerase chain reaction (RT-PCR) showed that 25 viruses are either novel strains or viruses. The most abundant viruses in peas were pea enation mottling virus 1 (PEMV1) and pea enation mottling virus 2 (PEMV2), followed by turnip yellow virus (TuYV). In addition, a number of viruses were reported for the first time to infect peas in Germany.

Soybeans are valuable cash money-making crops that provide protein rich food for both humans and animals worldwide. Soybean mosaic virus (SMV, genus *Potyvirus*) and cowpea mild mottle virus (CPMMV, genus *Carlavirus*) are the two important ssRNA viruses that infect soybean worldwide. Wei et al. reports the presence of SMV and CPMMV in soybean samples from different regions of China. The HTS results showed that both viruses co-infect soybeans. Phylogenetic analysis of 11 nearly complete genome sequences of CPMMV isolates formed a distinct clade from previously reported isolates from around the world, indicating the presence of a new genetic clade.

Cao et al. describes a new virus associated with flower yellowing disease (FYD) of *Zanthoxylum armatus,* which has been a main constraint on the production for the last 10 years in China. *Z. armatus* belongs to the family *Rutaceae*, which comprises deciduous, spiny shrub species and are cultivated for spices, medicine, and essential oils. It is also known as green Sichuan (Szechwan) pepper. *Z. armatus* is economically important crop and is cultivated for green fruits that can be used as a seasoning for its special aroma and numbing flavor. The results of HTS from different diseased trees showed the association of a new ilarvirus (family *Bromoviridae*) named green Sichuan pepper-nepovirus (GSPneV) and the satellite RNA (satGSPNeV). This work showed for the first time that FYD is caused by a virus/subviral infection.

Pepper and tomatoes are infected by a range of plant viruses and cause significant effects on the plant itself and fruit. In this Research Topic, we had three studies, which showed the identification, biological and genetic characterization of viruses infecting peppers and tomatoes. One study showed the complete genome sequence of a new recombinant polerovirus: pod pepper vein yellows virus (PoPeVYV) (Peng et al.) that was determined by HTS and further confirmed by RT-PCR. Another study (Li et al.) based on serological and RT-PCR techniques found 19 different viruses in both crops, while seven viruses were common between peppers and tomatoes. The highest infection (20%) was recorded for an orthotosposvirus and tomato spotted wilt virus (TSWV), followed by a polerovirus, pepper vein yellows virus (PeVYV, 13.0%). The results showed for the first time that pepper is a natural host of tobacco vein distorting virus (TVDV; genus *Polerovirus*) worldwide and tomato is a natural host of potato leafroll virus (PLRV; genus *Polerovirus*) in China. The third study (Rao et al.) reported the identification of two new isolates of chili veinal mottle virus (ChiVMV; family *Potyviridae*, genus *Potyvirus*). Phylogenetic analysis showed virus isolates clustered based on the geographical origin while recombination was a major factor affecting the diversity of ChiVMV isolates.

Allium spp (onion, green onion, garlic, leek) is a hub of viruses, particularly tospovirues, which is a major constraint for their production worldwide. Tabassum et al. presented a global evolutionary analysis of many iris yellow spot virus isolates (IYSV;genus *Tospovirus*, family *Bunyaviridae*) based on the complete N gene sequences (considered as a marker gene for tospovirus identification and classification). Phylogenetic and in *silico* restriction fragment length polymorphism (RFLP) analysis of the N gene of 142 IYSV isolates categorized IYSV isolates into two major gentoypes: IYSV Netherlands (IYSVNL; 55.63%), IYSV Brazil (IYSVBR; 38.73%) and the rest fell in neither group [IYSV other (IYSVother; 5.63%)]. These results suggest that IYSVNL is the predominant genotype on a global scale.

In conclusion, these articles provided novel studies and knowledge about viruses infecting cereals, peppers, tomatoes, peas, and trees. Most importantly, the identification of new viruses from various plant species or the availability of complete genome sequences of known viruses will further close the knowledge gap about these viruses. This knowledge can be applied in future studies for the identification of novel viruses and their characteristics to understand how we can protect our agricultural crops to ensure a sustainable food supply in the coming years.

AUTHOR CONTRIBUTIONS

AA initiated this Research Topic and invited other guest editors. AA wrote the original draft of the editorial and all other authors reviewed and approved it for publication. All authors listed above have made direct and intellectual contribution to the above topic.

ACKNOWLEDGMENTS

The aim of this Research Topic was to invite authors and provide latest insights on the detection techniques of virus diseases in various plants, genetic diversity, virus evolution, and recombination studies of viruses from diverse settings. All the editors express their sincerely gratitude to the all researchers who have contributed their work to this Research Topic. We are also thankful to all the reviewers who took their valuable time by providing constructive critics to further improving these articles under the above Research Topic.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 November 2020
doi: 10.3389/fmicb.2020.583242

Investigating the Pea Virome in Germany—Old Friends and New Players in the Field(s)

Yahya Z. A. Gaafar[1], Kerstin Herz[1], Jonas Hartrick[1], John Fletcher[2], Arnaud G. Blouin[2,3†], Robin MacDiarmid[2,3] and Heiko Ziebell[1]*

[1] Julius Kühn Institute, Institute for Epidemiology and Pathogen Diagnostics, Braunschweig, Germany, [2] The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand, [3] School of Biological Sciences, The University of Auckland, Auckland, New Zealand

OPEN ACCESS

Edited by:
Kristiina Mäkinen,
University of Helsinki, Finland

Reviewed by:
Armelle Marais,
UMR 1332 Biologie du Fruit et Pathologie, France
Jared May,
University of Missouri–Kansas City, United States

***Correspondence:**
Heiko Ziebell
heiko.ziebell@julius-kuehn.de

†Present address:
Arnaud G. Blouin,
Integrated and Urban Plant Pathology Laboratory Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: 14 July 2020
Accepted: 28 September 2020
Published: 13 November 2020

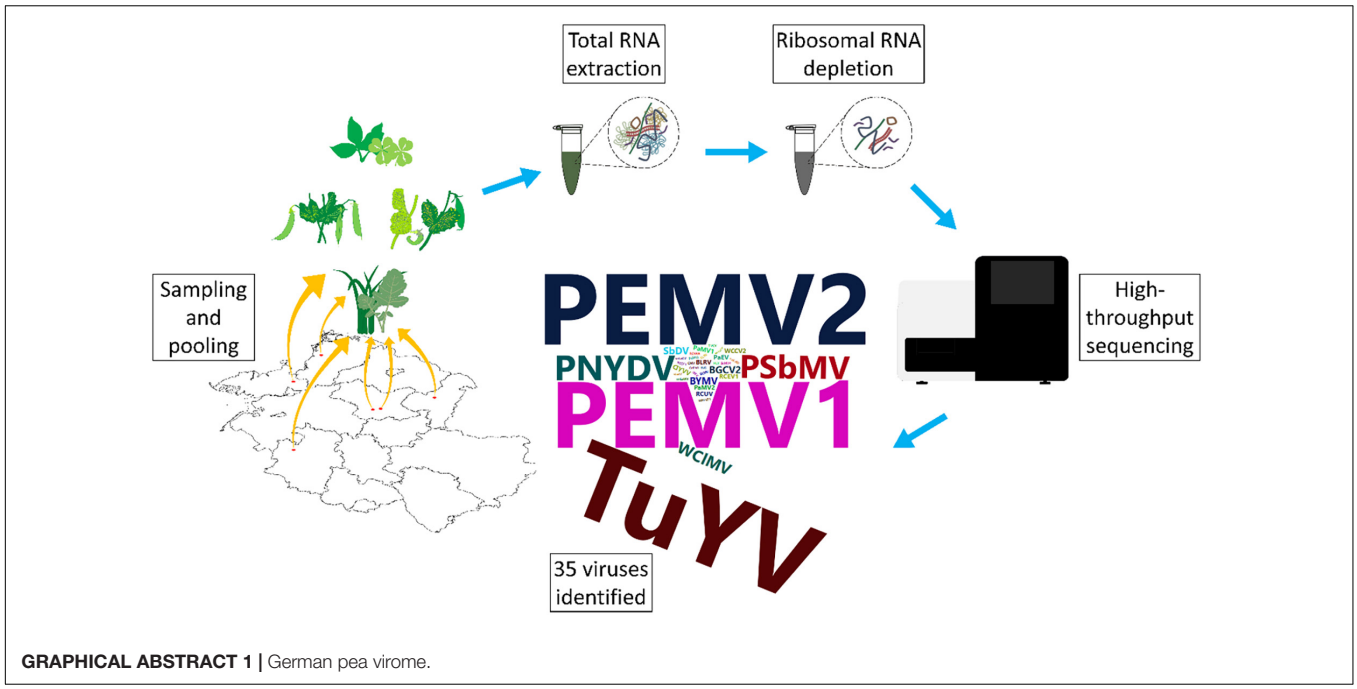
Citation:
Gaafar YZA, Herz K, Hartrick J, Fletcher J, Blouin AG, MacDiarmid R and Ziebell H (2020) Investigating the Pea Virome in Germany—Old Friends and New Players in the Field(s). Front. Microbiol. 11:583242. doi: 10.3389/fmicb.2020.583242

Peas are an important legume for human and animal consumption and are also being used as green manure or intermediate crops to sustain and improve soil condition. Pea production faces constraints from fungal, bacterial, and viral diseases. We investigated the virome of German pea crops over the course of three successive seasons in different regions of pea production to gain an overview of the existing viruses. Pools from 540 plants, randomly selected from symptomatic and asymptomatic peas, and non-crop plants surrounding the pea fields were used for ribosomal RNA-depleted total RNA extraction followed by high-throughput sequencing (HTS) and RT-PCR confirmation. Thirty-five different viruses were detected in addition to nine associated nucleic acids. From these viruses, 25 are classified as either new viruses, novel strains or viruses that have not been reported previously from Germany. Pea enation mosaic virus 1 and 2 were the most prevalent viruses detected in the pea crops, followed by pea necrotic yellow dwarf virus (PNYDV) and turnip yellows virus which was also found also in the surrounding non-legume weeds. Moreover, a new emaravirus was detected in symptomatic peas in one region for two successive seasons. Most of the identified viruses are known to be aphid transmissible. The results revealed a high virodiversity in the German pea fields that poses new challenges to diagnosticians, researchers, risk assessors and policy makers, as the impact of the new findings are currently unknown.

Keywords: *Pisum sativum*, high throughput sequencing, emaravirus, aphid transmitted viruses, PEMV, PNYDV, TuYV

INTRODUCTION

Green peas (*Pisum sativum* L.) are popular vegetables in Germany. The production of green peas increased from 4,444 ha in 2010 to 5,488 ha in 2018 (Behr, 2015, 2019). In addition, owing to the "Protein Strategy" of the Federal Government of Germany, the production areas of protein peas used as animal fodder, green manure or as intermittent crops, increased from 57,200 ha in 2010 to 70,700 ha in 2018 (BMEL, 2019). However, depending on the intended use of the crop, pea production in Germany is highly regionalized. The main green pea production areas are in Saxony, because of the nearby frozen foods processing facilities. Seed production of peas is predominantly carried out in Saxony-Anhalt. By contrast, green pea production for the fresh market or protein pea



GRAPHICAL ABSTRACT 1 | German pea virome.

production for animal fodder/green manure is scattered around the country and often associated with small-scale or organic farming.

Pea plants are known to be hosts to several viruses from different families, e.g., *Luteoviridae*, *Nanoviridae,* and *Potyviridae*. These viruses often occur in mixed infections causing a range of symptoms such as chlorosis, dwarfing, mottling, vein clearing, enations, and necrosis (Musil, 1966; Bos et al., 1988; Kraft, 2008; Gaafar et al., 2016; Gaafar and Ziebell, 2019b). Because of the recent detection of novel pea-infecting viruses, such as pea necrotic yellow dwarf virus (PNYDV), across Germany and within neighboring countries (Grigoras et al., 2010a; Gaafar et al., 2016, 2017, 2018) we were interested to know whether there are more unknown and/or undetected viruses present in this crop in Germany.

Currently, virus diagnostics generally depend on observational, serological or molecular methods that are based on prior knowledge of the target virus. These methods do not address the potential presence of other viruses that may contribute to the etiology of a disease. In recent years, high-throughput sequencing (HTS) has enabled the identification of many new viruses from domesticated and wild plants (Roossinck et al., 2015; Gaafar et al., 2019c,d). HTS allows sequencing of all the genetic material in a given sample. Therefore there is no need for prior knowledge of the infectious agent (Adams et al., 2009; Roossinck et al., 2015; Maree et al., 2018). Improvements of HTS technologies and bioinformatic tools have helped to identify the virus community or viromes of several crops (Coetzee et al., 2010; Czotter et al., 2018; Jo et al., 2018). The generated data reveal a description of plant virus biodiversity, allow the discovery of new viruses and viroids, identify genomic variants of the viral species and aid the development of specific and sensitive diagnostics (Coetzee et al., 2010; Li et al., 2012; Gaafar et al., 2019a,b). However, the vast number of new viruses identified by HTS results in challenges for diagnosticians, pest risk assessors and policymakers. The actual risks to crop plants and alternative hosts of these new viruses need to be evaluated and explored (MacDiarmid et al., 2013; Massart et al., 2017; Rott et al., 2017; Maree et al., 2018). We believe the metagenomics data revealed by HTS, along with biological studies will help us better understand viral ecology, viral evolution and disease epidemiology, particularly in economically important crops. Nevertheless, HTS studies impose a huge challenge on risk assessors and policy makers as the "real" importance of the findings (Does the presence of a novel virus cause a problem to the crop plant it was found in? What is the "real" distribution/quantity of a potential threatening virus in the whole crop?) cannot be assessed from sequence data alone.

In this study, using HTS we explored the spatio-temporal changes of the pea virome in six German regions with different production foci, including frozen produce, protein pea seed production, seed production and breeding over a period of three seasons. We also investigated alternative virus reservoirs of legume and non-legume weeds associated with these production sites. To our knowledge, this is the first metagenomics study of crop plants incorporating spatio-temporal developments.

MATERIALS AND METHODS

Sampling

Six regions in Germany were chosen for sampling and included Salzlandkreis-1: pea seed production, Salzlandkreis-2: trial site and heritage pea collection, Münster: trial site pea breeding, Kreis Stormarn: protein pea production, Landkreis Rostock: trial site

for green manure mixtures, and Landkreis Meißen: green pea frozen production for human consumption.

A typical production field was chosen and based on long-term experience, samples were selected randomly over three successive seasons (2016, 2017, and 2018 between June and July; **Figure 1**). From each field, 10 obviously symptomatic pea plants (SP) and 10 asymptomatic (no obvious symptoms) pea plants (aSP) were selected randomly; in addition to the pea plants, five surrounding legume plants (sL), and five non-legume plants (snL), were also collected at random. The metadata were recorded, including location, sample, season category, plant host, symptoms, and the average seasonal temperature for that region (**Supplementary Table S1**). Any deviations from the sampling strategy (i.e., in cases where no non-symptomatic peas or no surrounding legume plants could be detected) were noted (**Supplementary Table S1**).

Ten sample pools were prepared each season, i.e., six separate pools containing material from each of the six regions (30 plants each) and four separate pools comprising material from all SP (60 plants), aSP (60 plants), sL (30 plants), or snL (30 plants), respectively, from all regions (**Figure 1**). Fresh tissue (100 mg) from each plant was pooled. The samples were mixed and ground using a mortar and pestle under liquid nitrogen then collected into 50 mL Falcon tubes. The pools were stored at −20°C until RNA extraction. Remaining sample tissue was also stored at −20°C for further analysis.

RNA Extraction and High Throughput Sequencing

Total RNA was extracted using innuPREP Plant RNA Kit (Analytik Jena) from three subsamples of 100 mg from each pool to ensure sufficient total RNA yield for further processing thus allowing detection of low titre viruses. The extracted subsets were mixed and used for ribodepletion. Ribosomal RNA (rRNA) was depleted using a RiboMinus^{TM} Plant Kit for RNA-Seq (Invitrogen). cDNA was synthesized using ProtoScript II Reverse Transcriptase (NEB) and random octanucleotide primers (8N), followed by second strand synthesis using a NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module (NEB). The libraries were prepared from the double-stranded cDNA using Nextera XT Library Prep Kit (Illumina). The HT sequencing was performed on an Illumina MiSeq platform (301 × 2).

Bioinformatic Analysis

Bioinformatic analysis was performed using Geneious Prime software (version 2019.1.1). The reads were quality trimmed and normalized. *De novo* assembly was performed and the resulting contigs were compared against a local database of viruses and viroids sequences downloaded from NCBI using Blastn and Blastx (*E*-value = 1e-5) (downloaded 13 August 2018). The generated consensus sequences were based on the highest quality threshold. Primers for virus validation were designed using a modified version of Primer3 (2.3.7) tool in Geneious Prime (Untergasser et al., 2012). Pairwise alignments were performed using Clustal W tool (v 2.1) in Geneious (Larkin et al., 2007). Neighbor joining phylogenetic trees were constructed using MEGA X software (Kumar et al., 2018). The phylogenetic relationships were established according to the species demarcation criteria set by the International Committee on Taxonomy of Viruses (ICTV), using the nucleotide sequences or the amino acid sequences of the capsid protein (CP) or the RNA dependent RNA polymerase (RdRP) for the respective families. The isolates were named by region number and season, e.g., R1_16 stands for region one and the season 2016. The assembled virus sequences can be accessed in GenBank under accession nos. MN314973, MN399680–MN399748, MN412725–MN412751, and MN497793–MN497846.

RT-PCR Confirmations

For virus identity confirmation, total RNA was re-extracted from each pool as described above, followed by RT-PCR with the primers listed in **Supplementary Table S2** and using the OneTaq One-Step RT-PCR Kit (NEB). The products were purified using Zymoclean Gel DNA Recovery Kit (Zymo Research) and Sanger sequenced using both RT-PCR primers.

Statistical Analysis

Statistical analysis was performed using scripts written on R (version 3.5.3) (R Core Team, 2019). The virus incidence was calculated for each virus and associated nucleic acids, where a score of 1 is the lowest, for detected once, and a score of 78 is the highest, when a virus was detected in each sampling category of all regions over all seasons.

RESULTS

Metadata and HTS Raw Data

Symptoms observed on peas in the "symptomatic" pea pool included plant dwarfing, stunting of top leaves, translucent leaf spotting and leaf enations, leaf yellowing, leaf mosaic and mottling, leaf rolling and pod distortion (**Supplementary Table S1**). In 2016 leguminous weeds appeared absent in fields surrounding Salzlandkreis-1 and no asymptomatic peas were found on the site in Salzlandkreis-2. Similarly, leguminous weeds also appeared absent in Landkreis Meißen in 2018. Consequently, these pools were omitted from the analyses. Raw data generated from the HTS MiSeq platform are summarized in **Supplementary Table S3**.

The Viruses Detected in German Pea Fields

Thirty-five viruses were detected by HTS and further confirmed by RT-PCR in the sample pools over the three seasons. In addition, several other viral reads were detected that could not be confirmed by RT-PCR or that were obvious contaminations from various samples that were placed onto the same sequencing run (**Supplementary Table S6**). The detected viruses represented 14 different families and several unassigned viruses (**Figure 2**). The family *Luteoviridae* was represented by seven species members, followed by *Secoviridae* with six members, and *Potyviridae* with five members. Over the survey, the widest virus diversity was

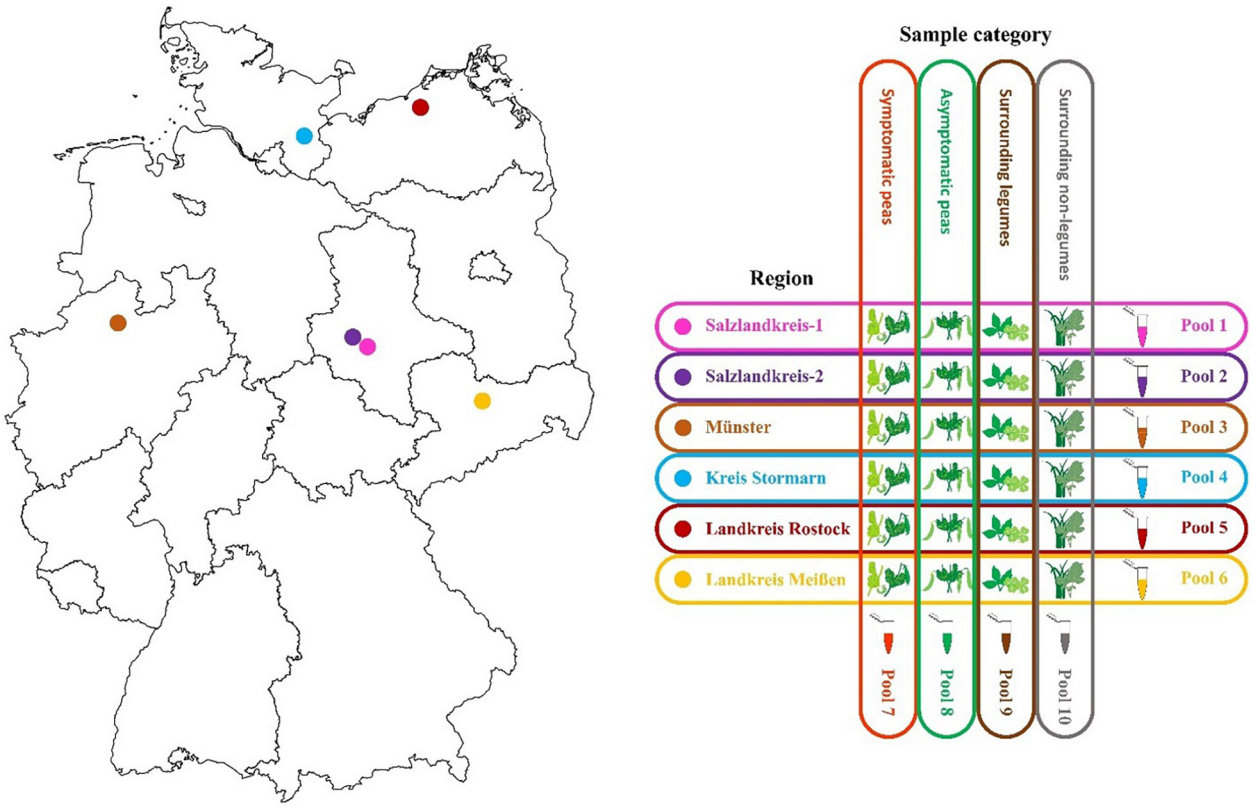


FIGURE 1 | Map of the six different pea crop sampling regions in Germany and the sample pooling strategy used for each of three successive seasons 2016, 2017, and 2018.

recorded in the SP pools, with sixteen different virus species present. The sL pool contained 12 different virus species, the snL pools included 11 virus species, and the aSP recorded seven virus species (**Figure 2**).

The identified viruses, the regions in which they were detected, and the pools from which they were discovered, are listed in **Supplementary Table S4**. The incidence of the detected viruses was calculated for each virus (**Supplementary Table S4**). Turnip yellows virus (TuYV) was the most abundant virus in this study with a score of 33, followed by pea enation mosaic virus 2 (PEMV2, score 31), pea enation mosaic virus 1 (PEMV1, score 29). Pea necrotic yellow dwarf virus (PNYDV) had a score of 11 followed by pea seed-borne mosaic virus (PSbMV) with a score of 10.

Figure 3 summarizes virus occurrence in peas (SP and aSP) and the surrounding non-crop plants (sL and snL) in each region over the three seasons. If looking at the pea pools only, PEMV2 was the most abundant virus with the score of 31 followed by PEMV1 (score 29) because they were not present in sL nor snL pools. With a score of 21, TuYV is the third most abundant virus in the pea pools.

In the sL, white clover mosaic virus (WClMV) was the most abundant virus, with a score of 5, followed by white clover cryptic virus 2 (WCCV2) and soybean dwarf virus (SbDV) with a score of 2 (**Figure 3**). All the other listed viruses were detected only once. In the snL samples, TuYV was the most abundant virus with a score of 12, while all the other viruses were detected only once.

The Most Abundant Viruses Detected in German Peas

TuYV was the most abundant virus from the study and was detected not only in pea (SP and aSP) but also in snL pools over the three seasons. Nucleotide identity (nt%) showed 95–99.2% to published TuYV sequences (**Supplementary Table S5**). Two associated RNAs, TuYV-associated RNA (TuYVaRNA) and a novel associated RNA (TuYVaRNA2), were also detected. TuYVaRNA2 isolates were closely related to cucurbit aphid borne virus-associated RNA (CABYVaRNA; KM486094, from the United States), with 89.3 to 95.1% nt identity (**Supplementary Table S5**).

Over the observation period, PEMV1 was detected almost all the time in all regions in SP and aSP pools with the exceptions of region 5 in 2017 and region 4 in 2018, where neither PEMV1 nor PEMV2 could be detected. The nucleotide (nt) sequences of PEMV1 isolates showed highest identities (96.1–98.9%) to MK948533 (**Supplementary Table S4**). In general, PEMV2 was always associated with PEMV1, with the exception of region 6 in 2018 where only PEMV2 was found. The nt sequences of PEMV2 showed 93.5–94.4% nt identity to

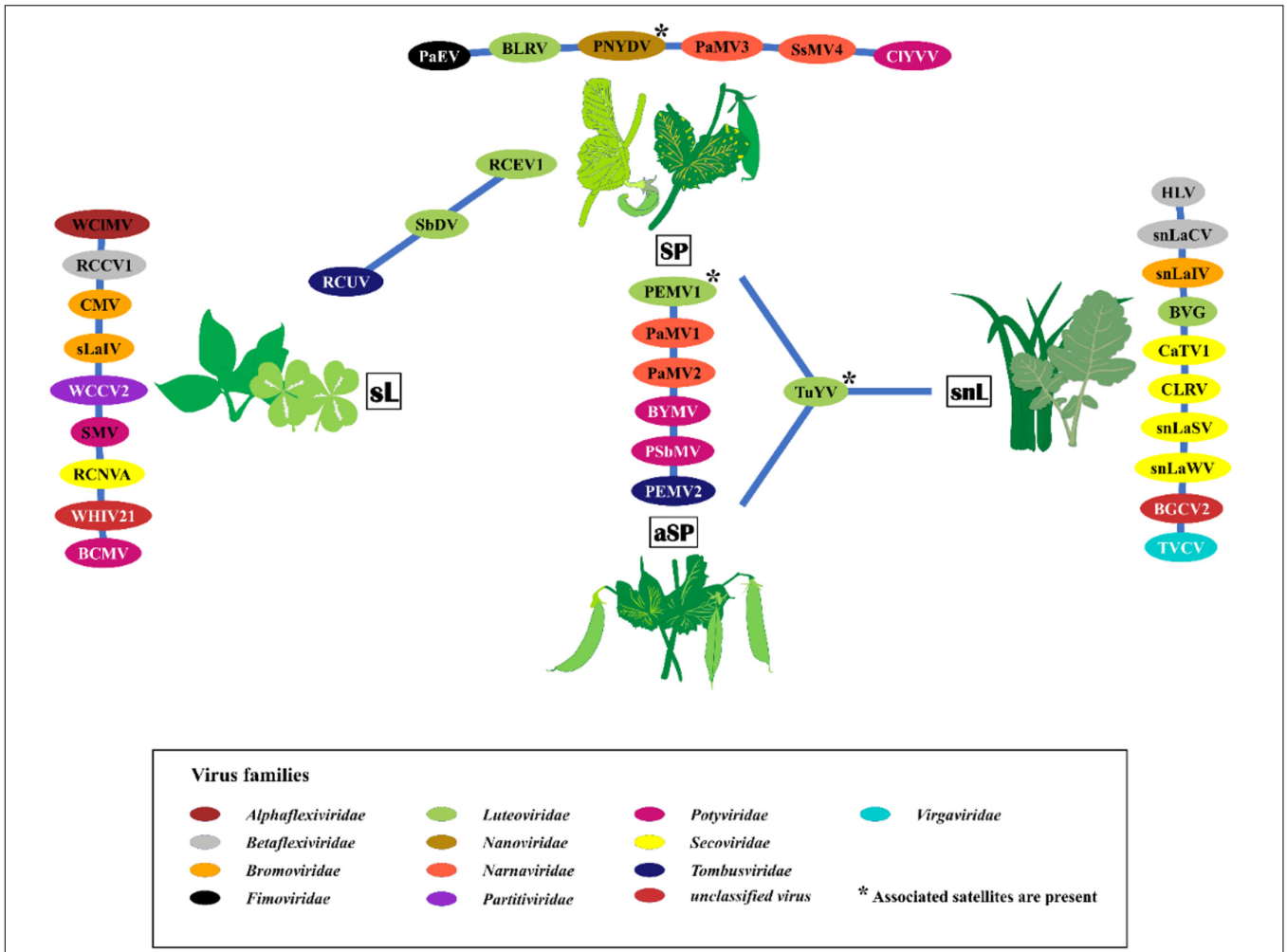


FIGURE 2 | The virus species found in the four sampling categories symptomatic peas (SP), asymptomatic peas (aSP), surrounding legumes (sL), and surrounding non-legumes (snL). Samples were collected in six pea-growing regions (Salzlandkreis-1, Salzlandkreis-2, Münster, Kreis Stormarn, Landkreis Rostock, and Landkreis Meißen) in Germany over three successive seasons 2016, 2017, and 2018. A network illustration of virus species of each category and shared viruses among them. Background colors of virus acronyms correspond to the background colors of the respective virus family. The virus names are: BVG, barley virus G; BCMV, bean common mosaic virus; BLRV, bean leafroll virus; BYMV, bean yellow mosaic virus; BGCV2, black grass cryptic virus 2; CaTV1, carrot torradovirus 1; CLRV, cherry leaf roll virus; ClYVV, clover yellow vein virus; CMV, cucumber mosaic virus; HLV, Heracleum latent virus; PaMV1, pea associated mitovirus 1; PaMV2, pea associated mitovirus 2; PaMV3, pea associated mitovirus 3; PaEV, pea associated emaravirus; PEMV1, pea enation mosaic virus 1; PEMV2, pea enation mosaic virus 2; PNYDV, pea necrotic yellow dwarf virus; PSbMV, pea seed-borne mosaic virus; RCCV1, red clover carlavirus 1; RCEV1, red clover enamovirus 1; RCNVA, red clover nepovirus a; RCUV, red clover umbravirus; SsMV4, Sclerotinia sclerotiorum mitovirus 4; sLaIV, sL associated ilarvirus; snLaCV, snL associated chordovirus; snLaIV, snL associated ilarvirus; snLaSV, snL associated secoviridae; snLaWV, snL associated waikavirus; SbDV, soybean dwarf virus; SMV, soybean mosaic virus; TVCV, turnip vein-clearing virus; TuYV, turnip yellows virus; WCCV2, white clover cryptic virus 2; WClMV, white clover mosaic virus; WHIV21, Wuhan insect virus 21.

MK948534 (**Supplementary Table S4**). In addition, we also detected a satellite RNA (PEMVSatRNA). Neighbor joining trees (NJ) based on the readthrough RdRp aa sequences of the isolates of the luteovirids and tombusvirids detected in German pea fields over three seasons are shown in **Supplementary Figures S1, S2**.

PNYDV was detected in SP pools with nt identities between 99.2 and 99.9% with PNYDV sequences available on NCBI (**Supplementary Table S5**). Six PNYDV alphasatellites (PNYDVαSat) including alphasatellites 1 and 3, and four new alphasatellites, tentatively named PNYDV-associated alphasatellites 4, 5, 6, and 7, were also detected. The isolates of PNYDVαSat4, PNYDVαSat5, PNYDVαSat6 and PNYDVαSat7 were 1,030, 991, 1,037, and 1,015 nt in length, respectively. PNYDVαSat4 showed closest nt identity to faba bean necrotic yellows virus C7 alphasatellite (FBNYC7αSat; AJ005964) from Egypt, with 88.5 and 88.6% relationship. Phylogenetic trees based on the alignments of the nt sequences and the aa sequences of the encoded proteins are shown in **Supplementary Figures S3A,B**.

PSbMV was detected only in two regions, Salzlandkreis-2 and Kreis Stormarn, but was present over all three seasons in SP and aSP pools. PSbMV isolates had 96.2–99.7% nt identity to D10930, a Danish isolate. Neighbor joining tree based on the polyprotein aa sequences of the potyvirid isolates detected in German pea fields over three seasons are shown in **Supplementary Figure S4**.

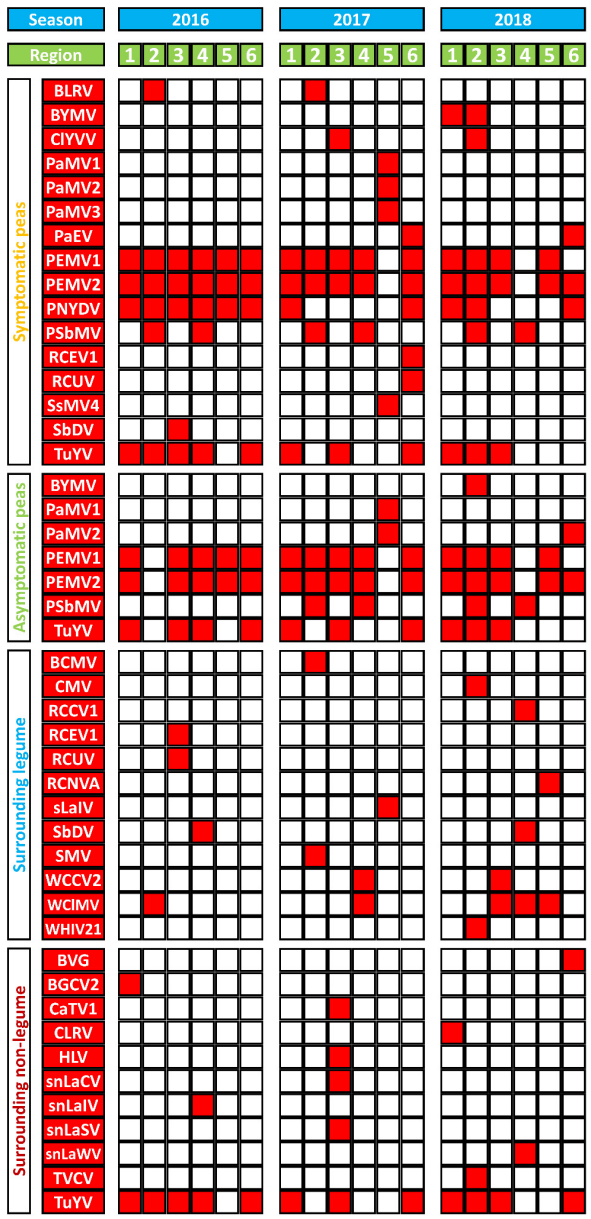


FIGURE 3 | A summary of virus species (in red) detected in and around sampled pea crops from six different German pea-growing regions (Salzlandkreis-1 [1], Salzlandkreis-2 [2], Münster [3], Kreis Stormarn [4], Landkreis Rostock [5], and Landkreis Meißen [6]) over three seasons 2016, 2017, and 2018. Virus acronyms, virus names, and their associated nucleic acids: BVG, barley virus G; BCMV, bean common mosaic virus; BLRV, bean leafroll virus; BYMV, bean yellow mosaic virus; BGCV2, black grass cryptic virus 2; CaTV1, carrot torradovirus 1; CLRV, cherry leaf roll virus; ClYVV, clover yellow vein virus; CMV, cucumber mosaic virus; HLV, Heracleum latent virus; PaMV1, pea associated mitovirus 1; PaMV2, pea associated mitovirus 2; PaMV3, pea associated mitovirus 3; PaEV, pea associated emaravirus; PEMV1, pea enation mosaic virus 1; PEMV2, pea enation mosaic virus 2; PEMVsatRNA, pea enation mosaic virus satellite RNA; PNYDVaSat1, pea necrotic yellow dwarf alphasatellite 1; PNYDVaSat3, pea necrotic yellow dwarf alphasatellite 3; PNYDVaSat4, pea necrotic yellow dwarf alphasatellite 4; PNYDVaSat5, pea necrotic yellow dwarf alphasatellite 5; PNYDVaSat6, pea necrotic yellow dwarf alphasatellite 6; PNYDVaSat7, pea necrotic yellow dwarf alphasatellite 7; PNYDV, pea necrotic yellow dwarf virus; PSbMV, pea seed-borne mosaic virus; RCCV1, red clover carlavirus 1; RCEV1, red clover enamovirus 1; RCNVA, red clover nepovirus a; RCUV, red clover umbravirus; SsMV4, Sclerotinia sclerotiorum mitovirus 4; sLalV, sL associated ilarvirus; snLaCV, snL associated chordovirus; snLalV, snL associated ilarvirus; snLaSV, snL associated secoviridae; snLaWV, snL associated waikavirus; SbDV, soybean dwarf virus; SMV, soybean mosaic virus; TVCV, turnip vein-clearing virus; TuYV, turnip yellows virus; TuYVaRNA, turnip yellows virus associated RNA; TuYVaRNA2, turnip yellows virus associated RNA 2; WCCV2, white clover cryptic virus 2; WClMV, white clover mosaic virus; WHIV21, Wuhan insect virus 21.

New Viruses and Virus Strains

A New Emaravirus in German Peas (Fimoviridae): Pea Associated Emaravirus (PaEV)

A new emaravirus was identified in SP pools from of Landkreis Meißen in 2017 and 2018. The virus showed high similarity to other members of genus *Emaravirus* including fig mosaic virus (FMV), pigeonpea sterility mosaic virus 2 (PPSMV2) and rose rosette virus (RRV). Members of the genus *Emaravirus* (family *Fimoviridae*, order *Bunyavirales*) have segmented, linear, single-stranded, negative-sense RNA genomes (Elbeaino et al., 2018). Their genomes are composed of up to eight segments. Partial sequences of segments RNA1–6 were assembled from the two Meißen isolates. An NJ tree of the nucleocapsid protein (NP) amino acid (aa) sequences of the two isolates was constructed and grouped together in a clade with PPSMV2 and FMV (**Figure 4**). The aa sequence of the NP shared the highest identity with PPSMV2 with 71.7%. According to the species demarcation of ICTV, a difference of 25% in the aa sequence of the NP indicates a new species (Elbeaino et al., 2018). Therefore, this virus represents a new emaravirus species and is tentatively called pea associated emaravirus (PaEV). To identify the pea sample in which PaEV was present, total RNA was extracted from individual samples from the positive pool of 2018 and virus presence was confirmed by RT-PCR in two samples with chlorosis symptoms. These samples were also infected with PEMV2 and PNYDV as confirmed by RT-PCR. **Supplementary Figure S5** shows the symptoms on sample R6-18-05 from Landkreis Meißen in 2018.

Pea Associated Mitoviruses

In Landkreis Rostock, four mitoviruses (family: *Narnaviridae*) were detected in the SP and aSP pools in 2017. An isolate of Sclerotinia sclerotiorum mitovirus 4 (SsMV4) was identified from the SP pool. This isolate shared 96.3% aa identity to SsMV4 from New Zealand (AGC24233). The three further mitoviruses identified are provisionally called pea associated mitovirus 1, 2 and 3 (PaMV1, PaMV2 and PaMV3). The PaMV1-CDS shared closest identity of 68.8% nt to Erysiphe necator mitovirus 3 (EnMV3; KY420040), identified from the grape powdery mildew fungus *Erysiphe necator* (Schwein.) [syn. *Uncinula necator* (Schw.)], recently described in the United States (Pandey et al., 2018). Based on the RdRp region, PaMV1 shares 65.1% aa identity with EnMV3. PaMV2-CDS shares 36% nt identity to Rhizoctonia solani mitovirus 6 (RsMV6; KP900915 from the

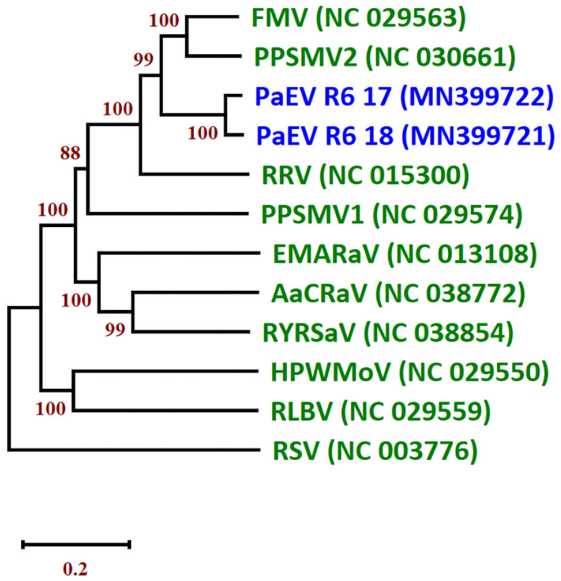


FIGURE 4 | Neighbor joining tree (NJ) of the nucleocapsid protein (NP) of the two isolates of pea associated emaravirus (PaEV) detected in Germany and representative emaraviruses. The phylogenetic trees are based on alignments of the amino acid sequences. The sequences were aligned with Clustal W and NJ trees constructed with MEGA X. The percentage of the bootstrap values above 50% (1,000 replications) are shown at the nodes. Virus acronyms and names are as follows: AcCRaV, Actinidia chlorotic ringspot-associated emaravirus; EMARaV, European mountain ash ringspot-associated emaravirus; FMV, fig mosaic emaravirus; HPWMEV, high plains wheat mosaic emaravirus; PaEV, pea associated emaravirus; PPSMV1, pigeonpea sterility mosaic emaravirus 1; PPSMV2, pigeonpea sterility mosaic emaravirus 2; RLBV, raspberry leaf blotch emaravirus; RRV, rose rosette emaravirus; RSV, rice stripe virus (*Tenuivirus; Phenuiviridae*) and RYRSaV, redbud yellow ringspot-associated emaravirus. The GenBank accession numbers are in brackets. The scale bar indicates the genetic distance.

United States). Based on the aa sequences, PaMV1 and RsMV6 share only 40.2% aa identity. PaMV3-CDS was closely related to Entomophthora muscae mitovirus 5 (EnmuMV5; MK682524), with 40% nt identity and 24.3% aa identity to the RdRp region.

Other Novel Virus Detections

Twenty-three viruses were detected from the sL and snL pools, and of these, five viruses appear to be novel. These new viruses were provisionally named according to the particular pool in which they were detected. The closest GenBank sequences of these viruses are presented in **Supplementary Table S5**. Two new ilarviruses were identified in the snL samples of Trenthorst in 2016; one appears closely related to asparagus virus 2, the other from the sL pool of Landkreis Rostock in 2018 appears to be closely related to ageratum latent virus. The putative CP aa of this snL ilarvirus shared closest identity to asparagus virus 2 (AV2; NC_011807) with 81.1% aa identity. The sL ilarvirus CP was closely related to ageratum latent virus (AgLV; NC_022129) with 60.3% aa identity. A phylogenetic tree based on the alignments of the RdRp aa sequences of the bromovirids is shown in **Supplementary Figure S6**.

In Münster 2017, a new putative member of the family *Secoviridae* with a close relationship to lettuce secovirus 1 was also identified in snL. The snLSV protease-polymerase region (Pro-Pol) shared closest aa identity to LSV1 with 67.3% while its CP region showed closest aa identity to LSV1 with 31.6%. Neighbor joining tree based on the Pro-Pol aa sequences of the secovirids detected in German pea fields over three seasons are shown in **Supplementary Figure S7**. We also detected a partial sequence of a chordovirus (subfamily: *Trivirinae*) in the snL of Münster 2017, sharing 74.4% identity with carrot chordovirus 1. Finally, in 2018, a partial waikavirus sequence was detected from the snL pool of Kreis Stormarn and appears to be related to bellflower vein chlorosis virus. A phylogenetic tree based on the alignments of the replicase aa sequences of the betaflexiviridis is shown in **Supplementary Figure S8**.

In addition to the viruses described above, new virus strains were detected including red clover enamovirus 1 (RCEV1), red clover umbravirus (RCUV), red clover nepovirus A (RCNVA), red clover carlavirus 1 (RCCV1) and Wuhan insect virus 21 (WHIV21) were identified in the sL, and divergent isolates of cherry leaf roll virus (CLRV) and carrot torradovirus 1 (CaTV1) were found in the snL. Also, the complete coding sequence of an isolate of Heracleum latent virus (HLV) (genus: *Vitivirus*; subfamily: *Trivirinae*, family: *Betaflexiviridae*), was identified in snL of Münster 2017.

In Münster 2016, a new strain of red clover enamovirus 1 (RCEV1) was identified in the sL samples. Its RdRp aa sequence shows 87.1% identity to the Czech strain (MG596229). Another isolate was detected in Landkreis Meißen in 2017, with 87.6% nt aa identity to each other and 95.8% to the Czech strain. The viruses grouped together within the *Enamovirus* genus clade (**Supplementary Figure S1**). A new strain of RCUV, a novel umbravirus found in red clover from the Czech Republic (Koloniuk, pers. comm.), was detected in the sL in Münster 2016 and the SP in Landkreis Meißen 2017. The complete coding sequence (CDS) shared 87.1% nt identity, while the aa sequence of their RdRp shared to 91.6% aa identity to the Czech isolate (MG596234). Both virus strains are grouped together within the *Umbravirus* genus (**Supplementary Figure S2**).

A divergent strain of red clover carlavirus 1 (RCCV1) (genus: *Carlavirus*; subfamily: *Quinvirinae*; family: *Betaflexiviridae*), was identified in one location (Kreis Stormarn) in 2018 only. The partial RdRp sequence shared 85.3% aa identity with RCCV1 (MG596238 and MG596239) from the Czech Republic. The CDS of an isolate of Heracleum latent virus (HLV) (genus: *Vitivirus*; subfamily: *Trivirinae*, family: *Betaflexiviridae*), was identified the snL of Münster 2017. Based on the CP sequence, this isolate shared 90.9% identity to HTV (NC_039087) on the nt level and 96.4% identity based on the aa sequence. Moreover, the RdRp shared 58.4% nt identity to grapevine virus B (GVB; MF991949) and 58.9% aa identity.

In addition, a new strain of RCNVA (genus: *Nepovirus*; subfamily: *Comovirinae*; family: *Secoviridae*), was detected in Landkreis Rostock 2017. This new strain was identified with Pro-Pol aa identity of 96.5% to MG253828 and CP aa identity of 83.2% to MG253829. A divergent cherry leaf roll virus (CLRV) (genus: *Nepovirus*) was identified in Salzlandkreis-1 in 2018. The virus

shared closest identity with CLRV isolates from New Zealand, where RNA1 has 82.4% nt identity to CLRV isolates KC937022 and RNA2 shared 80% to KC937029. The aa sequence of the Pro-Pol region is 97% identical to KC937022 while the CP region is 89.8% to KC937029. A new strain of carrot torradovirus 1 (CaTV1) with similarity to CaTV1 strain celery from Germany (MK063924 and MK063925) with 95.9% aa identity to the Pro-Pol region and 95.4% to the CP region.

Finally, a sequence with 81.3% nt identity to Wuhan insect virus 21 (WHIV21; KX883227) was detected in sL of Salzlandkreis-1 2018.

Spatial and Temporal Differences in the Pea Viruses

The spatial and temporal compositions of the virome, in the different regions over a period of three growing seasons, show similarities as well as fundamental differences. In pea crops, PEMV1, PEMV2, and PNYDV were the viruses that were found in all regions but not in every season (**Figure 3**). PEMV1, PEMV2 and their satellites were not detected in 2017 in Landkreis Rostock, or in 2018 in Kreis Stormarn. In 2018, PEMV1 was not detected in Landkreis Rostock, but PEMV2 and the satellite RNA were present. PNYDV was detected in all six regions in 2016 (**Figure 3**). In 2017, PNYDV was detected only in Salzlandkreis-1 and Landkreis Meißen. In 2018, it was detected in three regions: Salzlandkreis-1, Salzlandkreis-2, and Landkreis Meißen.

PSbMV was detected in all seasons in Salzlandkreis-2 and Kreis Stormarn. BLRV was detected only in Salzlandkreis-2 in 2016 and 2017 but it could not be found in 2018. PaEV was also detected for two seasons, in 2017 and 2018, only in Landkreis Meißen.

TuYV was detected in all three seasons in Salzlandkreis-1, Münster and Landkreis Meißen; however, after finding it present in Kreis Stormarn in 2016, but it could not be detected in 2017 and 2018. In 2017 TuYV was also not detected in Salzlandkreis-2. TuYV was never detected in the Landkreis Rostock region (**Figure 3**).

Other detected viruses included BYMV, detected only in Salzlandkreis-1 and Salzlandkreis-2 in 2018. ClYVV was found in Salzlandkreis-2 in 2018 and Münster in 2017. RCEV 1 and RCUV were detected only in Landkreis Meißen in 2017; however, both viruses were detected in the sL of Münster in 2016. SbDV was detected in SP in Münster 2016 and in the sL of Kreis Stormarn in 2016 and 2018. BLRV and PaEV were detected on SP only in Salzlandkreis-2 and Landkreis Meißen, respectively.

In summary, from the spatial distribution of the 12 pea viruses (excluding the mitoviruses), we observed that the highest virus occurrence was detected in Salzlandkreis-2 (17 occurrences over the three seasons) followed by Landkreis Meißen with 14 virus occurrences. A total of 13 pea virus occurrences were reported from Salzlandkreis-1, 12 for Münster, nine for Kreis Stormarn and the lowest number of occurrences was for Landkreis Rostock (only five pea viruses). The temporal virus occurrences appeared to be relatively stable over the successive seasons with the highest being 2016 with 27 occurrences, then 22 for 2017 and 21 for 2018 (**Figure 3**).

DISCUSSION

This is the first HTS-based study to describe the pea virome in Germany and to our knowledge, elsewhere in the world. In addition to the focus on viruses infecting pea crops, we also explored spatio-temporal aspects of virus infection across six different production regions over a period of 3 years. We also explored potential virus reservoirs by investigating legume and non-legume hosts in and around the crops. Our study is distinct from many metagenomics studies in that we explored our subject over three seasons in six regions. Other studies have focused on either just one crop plant, one production area, one season or a combination thereof. We believe that this study demonstrates the importance of incorporating spatio-temporal elements into metagenomics studies to help us to draw a more complete picture of all the viruses present and their impacts on the host crop as we would have missed many virus detections if we had focused on either just one growing season or just one region (e.g., the novel emaravirus that was not detected in 2016 nor outside Meißen).

To explore the pea virome dispassionately, we used ribosomal RNA-depleted total RNA for our sequencing approach (Pecman et al., 2017). This method has demonstrated it can successfully detect both RNA viruses with a plus and negative sense genome, as well as those with a DNA genome. Owing to the recent outbreaks of PNYDV in Germany and other European countries (Grigoras et al., 2010a; Gaafar et al., 2016, 2017, 2018), we were particularly interested to detect this DNA virus and its associated satellites, not only in pea crops but in particular in alternative hosts but unfortunately we were unable to identify an alternative natural host for PNYDV. We discovered a surprisingly high number of viruses in the different sample pools in high abundance, in particular, positive single-stranded viruses, including PEMV1, PEMV2, PEMVSatRNA, TuYV, TuYVaRNA1, and TuYVaRNA2. The recovered viral reads appeared to be pool-dependent as well as dependent on the viral genome, virus titer and frequency of the virus within the pool (**Supplementary Tables S3, S5**). It is interesting to note that also dsRNA viruses, i.e., partitivirids, and the new negative sense RNA emaravirus were detected using the ribo-depletion method, although with a lower read abundance.

Pea Virome

Detection of Well Recognized Pea Viruses in Germany

Using RT-PCR, we were able to confirm the presence of pea viruses that are commonly known to be present in Germany including PEMV1, PEMV2, TuYV, PNYDV, BLRV, SbDV, PSbMV, and CMV. PEMV1 and PEMV2 are the most common viruses found in German pea crops (Ziebell, 2017). PEMV disease is associated with mosaic and enation of leaves along with pod distortion; however, early stages of infection may remain symptomless. These symptoms lead to severe yield losses (Clement, 2006). Mixed infections of PEMV1 and PEMV2 are well documented and were also common in our pea samples (**Supplementary Table S4**; Hagedorn and Khan, 1984; Brault et al., 2010). Interestingly, we also found

PEMV2 in the absence of its helper virus PEMV1 (Landkreis Meißen 2018) and it would be interesting to know if a different helper virus would be able to facilitate PEMV2 transmission. Resistance to PEMV has been an important goal of green pea breeding for many years (Budke and Weil, pers. comm.), with many commercial varieties becoming available in the future.

The second most prevalent virus from our survey was TuYV. It was detected in pea crops and in snL pools. TuYV infects peas and may cause plant stunting and a top yellowing leaf symptom in susceptible cultivars. TuYV is also a major concern in German rapeseed crops (Graichen and Schliephake, 1999; Gaafar and Ziebell, 2019b). TuYV has a very wide host range including weeds, legume pastures and other members of the *Brassicaceae* (Stevens et al., 1994). Although we have no direct evidence that surrounding non-legumes are reservoirs for TuYV isolates that infect peas, in greenhouse experiments we demonstrated that TuYV isolates originating from peas can infect rapeseed and *vice versa* (data not shown). It is therefore very likely that rapeseed crops, other members of the *Brassicaceae* family as well as a large number of common weeds and wild species are alternative hosts for pea-infecting poleroviruses such as TuYV (Stevens et al., 1994). Resistance to luteovirids in peas is well documented. In New Zealand, it is in response to the top yellows disease complex caused by TuMV and/or SbDV, and in North America, in response to bean leafroll virus (BLRV) (Fletcher, pers. comm.). In Germany, sources of resistance have been identified in rapeseed and efforts are underway to breed TuYV-resistant rapeseed varieties (Graichen, 1994). Similarly, BRLV-resistance has been introduced into pea germplasm but it is currently not known whether it also provides resistance to TuYV. However, since BLRV was not very abundant in our survey (only found in region 2 in 2016 and 2017 in symptomatic peas), it does not seem to play a major role in the German pea virome.

In 2009, PNYDV was first described in Germany (Grigoras et al., 2010a). In the following years, PNYDV was detected in Saxony and Saxony-Anhalt, as well as in neighboring Austria (survey data, not shown). In 2016 a country-wide outbreak occurred in Germany, with PNYDV being detected in other European countries including Denmark and The Netherlands (Gaafar et al., 2016, 2017, 2018; Ziebell, 2017). Effects on infected plants are severe and can cause high yield losses (Saucke et al., 2019). PNYDV causes severe yellowing and dwarfing of infected plants that can lead eventually to plant death (Gaafar et al., 2017). PNYDV is an increasing threat to legume production in Europe as no PNYDV-resistant plant varieties have yet been identified (Ziebell, 2019). Also of concern are the high mutation rate, reassortment and recombination rates of nanoviruses such PNYDV, and these might lead to the emergence of novel strains (Grigoras et al., 2010b, 2014).

In recent years, an increasing number of nanovirus-associated single-stranded circular DNA alphasatellites have been reported in legumes, including *Sophora alopecuroides* L., *Vicia cracca* L., and *Apiaceae* members such as *Petroselinum crispum* (Mill.) Fuss. At this stage, their biological relevance is still unclear (Heydarnejad et al., 2017; Gallet et al., 2018; Vetten et al., 2019). Our survey also detected DNA alphasatellites in German peas. However, we failed to identify alternative host plants for PNYDV or its associated alphasatellites in our study. We believe by focusing on alternative plants within the pea fields or in close proximity, we may have missed potential hosts, and we need to widen the sampling radius in future surveys.

The seed-borne and aphid-transmissible potyvirus PSbMV has been reported from many countries, including Germany (Khetarpal and Maury, 1987; Latham and Jones, 2001). As the symptoms of PSbMV are often mild and transitory in peas, there is limited detection of the virus in the field. However, significant seed damage may occur in some susceptible varieties (Khetarpal and Maury, 1987). In Germany, PSbMV is not seen as a major constraint of pea production, as the provision of "clean" seed material and close surveillance of pea seed production has helped to reduce PSbMV in commercial crops to acceptable rates. One of the two sites in which we detected PSbMV is a trial site for heritage material where eradication of seed-borne virus is difficult to manage without compromising the collection. The second detection of PSbMV was found close to a protein pea breeding site where PSbMV had been previously reported (data not shown).

New Players in German Peas

Our study shows that in Germany, there are viruses present that have not been described from peas in Germany before: BCMV, BYMV, ClYVV, RCEV1, RCUV, and associated nucleic acids, i.e., PEMVSatRNA, TuYVaRNA, PNYDVαSat1, and 3. BCMV is well known to infect *Phaseolus* beans causing common mosaic or black root disease depending on the host, virus strain and the environmental conditions (Drijfhout and Bos, 1977). BCMV is a seed-borne virus, aphid transmissible and can be transmitted mechanically. It is interesting to find strain NL1 in German peas for the first time as previous studies showed that this strain could not infect peas (Drijfhout and Bos, 1977).

BYMV has a wide host range compare to other potyviruses including legumes and ornamentals (Guyatt et al., 1996; Nakazono-Nagaoka et al., 2009). Additionally, it can be transmitted by more than 20 aphid species causing symptoms including mosaic, necrosis and yellowing resulting in severe yield losses (Guyatt et al., 1996; Nakazono-Nagaoka et al., 2009). The pathogenicity and serotypes of the BYMV differs from a strain to another (Bos, 1970; Barnett, 1987). Clover yellow vein virus (ClYVV) has a host range overlapping with BYMV and often confused with it as they are closed serologically (Barnett, 1987; Nakazono-Nagaoka et al., 2009). Both viruses were reported in German legumes such as clover before. However, to our knowledge this is the first report of their natural infection to peas in German fields.

Interestingly, the red clover viruses RCEV1 and RCUV were not only detected in peas but also in the sL (but not in the same location and not in the same season), which

indicates that more information is required to determine if surrounding perennial legumes are a virus reservoir for these viruses. Mixed infections of both viruses were confirmed in red clover (*Trifolium pratense* L.) in the Czech Republic (Koloniuk et al., pers. comm.). Relationships between luteovirids and umbraviruses are important, for example, in the relationship between PEMV1 and PEMV2 where one virus provides cell-to-cell movement within hosts and the other virus' coat protein is required for genome packaging and vector transmission (Syller, 2003).

For the first time in Germany, we discovered numerous virus-associated nucleic acid sequences. PEMVSatRNA is a small linear single stranded RNA satellite that has also been extracted from peas in the United States (Demler and de Zoeten, 1989). PEMVSatRNA does not appear to influence aphid transmission, particle morphology, or symptom expression in peas but it can reduce the severity of symptoms in *Nicotiana benthamiana* (Demler and de Zoeten, 1989; Demler et al., 1994). Whether the PEMVSatRNA detected in Germany can modulate symptom expression of PEMV in its natural host, *P. sativum,* remains to be investigated.

We also discovered PNYDV-associated alphasatellites 1 and 3 in our survey for the first time in Germany. PNYDVαSat1 was previously detected in Austria, while PNYDVαSat3 was detected in both Austria and Denmark (Gaafar et al., 2018). In addition, we discovered four new PNYDV-associated alphasatellites that have not been reported before. Alphasatellites rely on their helper virus for spread as they do not encode a coat protein (Briddon et al., 2018). The presence of alphasatellites is associated with reduced infectivity of faba bean necrotic yellows virus (a nanovirus) and in the case of begomoviruses, alphasatellites reduced or intensified symptoms and/or reduced titre of the begomovirus or its associated betasatellites (Timchenko et al., 2006; Kon et al., 2009; Paprotka et al., 2010; Idris et al., 2011; Mar et al., 2017). The Rep proteins encoded by alphasatellites, associated with begomoviruses, were found to suppress transcriptional gene silencing or post-transcriptional gene silencing (Nawaz-Ul-Rehman et al., 2010; Abbas et al., 2019). The role of these nanovirus-associated alphasatellites is unknown and their association with nanoviruses and other viruses such as babuviruses and begomoviruses requires clarification.

Finally, TuYVaRNA and TuYVaRNA2 were also detected in association with TuYV in pea hosts. We recently discovered a TuYVaRNA that is associated with TuYV isolates from rapeseed in Germany (Gaafar and Ziebell, 2019b). Similarly, their role in symptom modulation, host range determination or vector transmission also remains to be investigated. The effects of these associated RNAs on TuYV transmission and infection need more investigation, as previous studies showed that beet western yellows associated RNA (strain ST9) increases the severity of beet western yellows on its host (Sanger et al., 1994). Interestingly, PEMVSatRNA was almost always detected in the same pools as either helper virus PEMV1 or PEMV2 (**Supplementary Table S4**). In contrast, the TuYVaRNAs were only detected in few pools (only in Salzlandkreis-1 and -2 as well as in Landkreis Meißen, **Supplementary Table S4**); although there was a high abundance of TuYV in Münster, no associated RNAs could be detected. The reasons for this observation is unknown and warrants further investigation. It would be interesting to study the transmission efficiency of aphids for the viruses in single infection or in mixed infection with associated RNAs.

First Report of Novel Pea Viruses

Our study in Germany identified many viruses present in peas that were not previously detected or described. PaEV, as an example, was first detected in peas over two successive seasons in Landkreis Meißen. We believe that this virus may be established and might pose a risk to pea production. We were unable to attribute clear symptoms to PaEV as when we back-tested individual plant samples from the mixed pools, PaEV was only found in mixed infection with PEMV2 and PNYDV (**Supplementary Figure S5**) and we could not recover infectious virus material from the samples to inoculated indicator plants. However, since we detected PaEV in SP pools, using our specific primers (**Supplementary Table S2**), we can specifically test peas in future surveys. Emaraviruses have mainly been reported from trees and deciduous shrubs. In the United States and Canada, the emaravirus RRV is mite-transmitted by *Phyllocoptes fructiphilus* Kiefer (Acari: Eriophyidae). RRV causes extreme damage to roses, leading to plant death within a short period of time (Olson et al., 2017). RRV and its vector were placed on the A1 alert list by the European and Mediterranean Plant Protection Organisation (EPPO, 2019). Two other emaraviruses have been reported from *Cajanus cajan* L. (*Fabaceae*): pigeonpea sterility mosaic disease (SMD), caused by pigeonpea sterility mosaic emaravirus 1, and pigeonpea sterility mosaic emaravirus 2 (Elbeaino et al., 2014, 2015). These emaraviruses are also transmitted by eriophyid mites (*Aceria cajani* Chann.) (Elbeaino et al., 2015; Patil and Kumar, 2015). One can assume that PaEV might also be a mite-transmitted virus as there are various mite species reported on pea (Annells and Ridsdill-Smith, 1994; Morishita and Takafuji, 1999; Takafuji and Morishita, 2001; Abdallah et al., 2014). The number of segments of our PaEV isolate are unknown as the virus full genome could not be recovered due to its low number of reads from the pools. In future monitoring programs, the isolation, distribution and its potential vector(s) will be evaluated to assess the risk that this virus might pose.

Pea-Associated Mitoviruses

Three new mitoviruses were found in the pea pools: PaMV1, PaMV2, and PaMV3. Mitoviruses are widespread in plants and their infection of pathogenic fungi is often associated with reduced virulence (Wu et al., 2007; Xie and Ghabrial, 2012; Bruenn et al., 2015). The fungi associated with these new mitoviruses are unknown and need more investigation. It is possible, for example, that these mitoviruses may have a role as biocontrol agents of fungal infections. For example, they may reduce the impact of powdery mildew, downy mildew, *Aphanomyces*, and/or fusarium root rot diseases in peas (Boland, 2004). The mitovirus SsMV4 infects *Sclerotinia sclerotiorum* (Lib.) de Bary, a widespread plant pathogenic fungus, which

causes white mold disease especially in peas, lentils and beans and many other hosts (McKenzie and Morrall, 1975; Purdy, 1979; Bardin and Huang, 2001; Bolton et al., 2006; Nibert et al., 2018). Research has shown that SsMV4 in combination with two other mitoviruses, *Sclerotinia sclerotiorum* mitovirus 2 (SsMV2) and *Sclerotinia sclerotiorum* mitovirus 3 (SsMV3), reduced the *in vitro* growth and virulence of *S. sclerotiorum* on cabbage, common bean, oilseed rape and tomato (Khalifa and Pearson, 2013).

Viruses in the Surrounding Plants

In our study, we analyzed leguminous and non-leguminous plants in the vicinity of the pea fields to investigate potential virus reservoirs. Not surprisingly, in the pools of sL we were able to detect common legume-infecting viruses, i.e., BCMV, SMV, WCCV2, WClMV, CMV, and SbDV. We also identified new viruses in the sL pool including RCCV1, RCNVA, WClMoV, WHIV21, RCEV1, and RCUV (discussed above) and a novel ilarvirus, sLaIV. In 2016, a survey using antibodies developed for detection of red clover vein mosaic virus (RCVMV)-like carlaviruses suggested the presence of a carlavirus in several pea samples but the exact virus species was not determined (Ziebell, 2017). Koloniuk et al. (pers. comm.) recently identified the genome of RCCV1 and found that its capsid protein sequence is closely related to its homolog from RCVMV but the RdRp sequence differs significantly. RCNVA is a new virus that was identified in red clover (*Trifolium pratense* L.) in the Czech Republic (Koloniuk et al., 2018). It was detected only once in 2018 in the sL pool of Landkreis Rostock. The host range of RCNVA is currently unknown. Although the exact host of new ilarvirus sLaIV is currently unknown, this virus will be included in future surveys to investigate the abundance and potential host plants. To our knowledge, sLaIV is only the second ilarvirus to naturally infect legumes apart from tobacco streak virus (Kaiser, 1982).

In the pools of non-legume (snL) plants, we detected viruses previously described in Germany. These are CaTV1, CLRV, and TuYV. TuYV was often detected and poses a threat to other commercial crop plants including rapeseed and sugar beet. CaTV1 was recently detected in celery plants, exhibiting chlorotic ringspots, mosaic and strong yellowing symptoms (Gaafar and Ziebell, 2019a). Viruses in the snL pool that have not previously been reported from Germany include BVG, BGCV2, HLV, and TVCV. We also detected several new viruses in these pools, i.e., snLaCV, snLaIV, snLaSV, and snLaWV.

In summary, our study represents the first comprehensive virome study of the pea crop covering pea production regions, temporal effects and alternative hosts. We believe that our strategy using HTS was successful. We not only detected well established pea viruses such as PEMV and PNYDV, but also discovered viruses not previously reported from Germany, as well as a number of new viruses. Furthermore, we also detected viruses in samples that appeared asymptomatic indicating that HTS is able to detect early stage or low titer infections and/or asymptomatic infections. The sequence data that we generated will also improve our knowledge of virus taxonomy, ecology, epidemiology and host plant resistance. The primer that we developed can be applied by diagnosticiancs to check for the presence of the novel viruses and virus-associated nucleic acids. It also demonstrates the challenges of metagenomic HTS studies in the framework of laboratory and bioinformatics, result interpretation, biological significance, pest risk analyses and data sharing as we also found viral reads that could not be attributed to "real" virus findings (**Supplementary Table S6**; Olmos et al., 2018).

While we acknowledge the strength of HTS in identifying known and unknown viruses of crops, our pooling strategy has disadvantages. Firstly, we cannot obtain detailed information on the viruses infecting a single plant without back-testing each specimen in the pool. Secondly, it was not always possible to recover the full-length viral sequence using this method. Thirdly, pooling does not allow us to link individual plants symptoms with the viruses detected.

CONCLUSION

In conclusion, our method of using rRNA-depleted total RNA extracts from pooled plant tissue in combination with HTS, bioinformatic analysis and molecular confirmation has increased the speed and breadth of virus detection in one crop species in Germany, over three seasons. This method enabled the detection of a range of viruses regardless of their genome type. After sequencing our samples, we identified thirty-five viruses, for which we obtained many nearly full genomes (**Supplementary Table S5**). As expected, well-recognized pea viruses were detected in this study, including members of the *Luteoviridae*, *Nanoviridae*, *Potyviridae,* and *Tombusviridae* families. In addition, 25 new viruses associated with pea, non-crop legumes and non-legume plants were found, some unexpected and yet unexplained. More work is needed to reveal the importance and context of these new host/virus associations.

We found PEMV1 and PEMV2 were the dominant virus species in pea, which is consistent with what has been observed in the past. The most abundant virus was TuYV, as its sequences were not only recovered from pea pools but particularly from non-leguminous alternative host plants, which is consistent with its wide host range. A new emaravirus was detected in peas over two of the survey seasons, but its method of transmission and impact are still unknown and need further investigations. Other viruses were also detected in pea or alternative plants for the first time in Germany and their impacts have yet to be determined. Interestingly, most of the viruses detected in this survey were aphid-transmitted, thus emphasizing the continued importance of aphid management to reduce virus-spread. We also found viruses with little similarity with known species and suggest they could be categorized as unclassified. Many of these viruses have yet to be included in standard monitoring programs of pea diseases and therefore the abundance and impact of these viruses on pea and other legume crops are unknown.

We believe the data from this study provides a comprehensive and improved overview of viruses present in German pea fields, and demonstrates the usefulness of HTS and metagenomics.

For the newly detected viruses, further work is needed to determine the complete host range of these viruses, their effect on hosts and their likely vectors. It is also necessary to further investigate different locations and environments to increase our understanding of the diversity of these new viruses, not only for pea crops but also for other legumes, on a global scale. Our study shows that there are many more viral players in the field(s) than known from the past. Further studies are required to address the risks that these new viruses might pose to peas or other crops. Targeted surveys could assess the distribution and abundance of these new players and the risk they might pose to crop plants.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MN314973, MN399680–MN399748, MN412725–MN412751, and MN497793–MN497846.

AUTHOR CONTRIBUTIONS

YG, JF, AB, RM, and HZ conceived and designed the experiments. YG, JF, AB, and HZ performed the sampling. YG performed the experiments, analyzed the data, and wrote the draft of the manuscript. KH and JH provided technical assistance. RM and HZ acquired the funding. All authors read and approved the final version.

FUNDING

This research was financed by a German-New Zealand cooperation grant from the German Federal Ministry of Food and Agriculture (BMEL) through the Federal Office for Agriculture and Food (BLE), Germany and The Royal Society of New Zealand awarded to JF, AB, HZ, and RM. YG was funded by the German Egyptian Research Long-Term Scholarship (GERLS). Open access publication was enabled by Julius Kühn Institute (JKI) core funding.

ACKNOWLEDGMENTS

We are grateful for Amjad Zia for his suggestions. We also thank Pam Fletcher for assistance with field sample collections. We are grateful to van Waveren Saaten GmbH (Irina Weil and Bernd Budke), IPK Gatersleben (Ulrike Lohwasser), Frosta (Bernt Schmidtgen), Thünen Institut (Herwart Böhm), Julius Kühn-Institut (Steffen Roux), and numerous farmers that allowed us access to their properties for this study. We are also grateful to all persons who kindly shared their unpublished work with us. This article was part of Dr. rer. nat. thesis by Yahya Z. A. Gaafar and assigned the following identifier: http://hdl.handle.net/21.11130/00-1735-0000-0005-12BD-2.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.583242/full#supplementary-material

Supplementary Figure 1 | Neighbour joining tree (NJ) of virus isolates from *Luteoviridae* family detected in German pea fields over three successive sampling seasons in 2016, 2017 and 2018.

Supplementary Figure 2 | Neighbour joining tree (NJ) of virus isolates from *Tombusviridae* family detected in German pea fields over three successive sampling seasons in 2016, 2017 and 2018.

Supplementary Figure 3 | Neighbour joining trees (NJ) of PNYDV alphasatellites (PNYDVαSat) detected in German pea fields over three successive sampling seasons in 2016, 2017 and 2018 and representative alphasatellites species.

Supplementary Figure 4 | Neighbour joining tree (NJ) of virus isolates from *Potyviridae* family detected in German pea fields over three successive sampling seasons in 2016, 2017 and 2018.

Supplementary Figure 5 | Photo of the chlorosis symptom observed on sample R6-18-05 from Landkreis Meißen in 2018.

Supplementary Figure 6 | Neighbour joining tree (NJ) of virus isolates from *Bromoviridae* family detected in German pea fields over three successive sampling seasons in 2016, 2017 and 2018.

Supplementary Figure 7 | Neighbour joining tree (NJ) of virus isolates from *Secoviridae* family detected in German pea fields over three successive sampling seasons in 2016, 2017 and 2018.

Supplementary Figure 8 | Neighbour joining tree (NJ) of virus isolates *Betaflexiviridae* family detected in German pea fields over three successive sampling seasons in 2016, 2017 and 2018.

Supplementary Table 1 | List of primers used to confirm each virus detected by HTS samples from six German pea-growing regions (Salzlandkreis-1 [1], Salzlandkreis-2 [2], Münster [3], Kreis Stormarn [4], Landkreis Rostock [5], and Landkreis Meißen [6]) sampled over three seasons, 2016, 2017, and 2018.

Supplementary Table 2 | Statistics of the raw data of the generated reads from each pool of samples from six German pea-growing regions (Salzlandkreis-1 [1], Salzlandkreis-2 [2], Münster [3], Kreis Stormarn [4], Landkreis Rostock [5], and Landkreis Meißen [6]) sampled over three seasons, 2016, 2017, and 2018.

Supplementary Table 3 | List of the viruses and their associated nucleic acid satellites detected by HTS and confirmed by RT-PCR in each pool of samples from six German pea-growing regions (Salzlandkreis-1 [1], Salzlandkreis-2 [2], Münster [3], Kreis Stormarn [4], Landkreis Rostock [5], and Landkreis Meißen [6]) sampled over three seasons, 2016, 2017, and 2018. The calculated incidence of each virus are indicated.

Supplementary Table 4 | List of the detected virus sequences and their associated nucleic acid satellites from six German pea-growing regions (Salzlandkreis-1 [1], Salzlandkreis-2 [2], Münster [3], Kreis Stormarn [4], Landkreis Rostock [5], and Landkreis Meißen [6]) sampled over three seasons, 2016, 2017, and 2018. Their GenBank accession numbers, their closest sequence matchess on NCBI (BLASTn), percent of coverage of the reference genomes, the number of reads and the number and size range of the contigs generated by *de novo* assembly are indicated.

Supplementary Table 5 | List of the unconfirmed virus sequences detected in the pools. Their closest sequence matches on NCBI (BLASTn), and the number and size range of the contigs generated by *de novo* assembly are indicated.

Supplementary Table 6 | The metadata records (sample, season, category, plant, symptoms, and average temperature) of samples from six German pea-growing regions (Salzlandkreis-1 [1], Salzlandkreis-2 [2], Münster [3], Kreis Stormarn [4], Landkreis Rostock [5], and Landkreis Meißen [6]) sampled over three seasons, 2016, 2017, and 2018.

REFERENCES

Abbas, Q., Amin, I., Mansoor, S., Shafiq, M., Wassenegger, M., and Briddon, R. W. (2019). The Rep proteins encoded by alphasatellites restore expression of a transcriptionally silenced green fluorescent protein transgene in *Nicotiana benthamiana*. *Virusdisease* 30, 101–105. doi: 10.1007/s13337-017-0413-5

Abdallah, A., Al-Azzazy, M., Mowafi, M., El-Saiedy, E., and Pastawy, M. (2014). Control of the two-spotted spider mite, *Tetranychus urticae* Koch on kidney bean and pea plants. *Acarines: J. Egypt. Soc. Acarol.* 8, 43–48. doi: 10.21608/ajesa.2014.4908

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

Annells, A. J., and Ridsdill-Smith, T. J. (1994). Host plant species and carbohydrate supplements affecting rate of multiplication of redlegged earth mite. *Exp. Appl. Acarol.* 18, 521–530. doi: 10.1007/BF00058935

Bardin, S. D., and Huang, H. C. (2001). Research on biology and control of *Sclerotinia* diseases in Canada 1. *Can. J. Plant Pathol.* 23, 88–98. doi: 10.1080/07060660109506914

Barnett, O. W. (1987). Relationships among Australian and North American isolates of the bean yellow mosaic potyvirus subgroup. *Phytopathology* 77:791. doi: 10.1094/Phyto-77-791

Behr, H.-C. (2015). *AMI Markt Bilanz Gemüse 2015*. Berlin: Deutschland.

Behr, H.-C. (2019). *AMI Markt Bilanz Gemüse 2019*. Berlin: Deutschland.

BMEL, (2019). *Eiweißpflanzenstrategie*. Available online at: https://www.bmel.de/DE/Landwirtschaft/Pflanzenbau/Ackerbau/_Texte/Eiweisspflanzenstrategie.html, [accessed July 04, 2019].

Boland, G. J. (2004). Fungal viruses, hypovirulence, and biological control of Sclerotinia species. *Can. J. Plant Pathol.* 26, 6–18. doi: 10.1080/07060660409507107

Bolton, M. D., Thomma, B. P. H. J., and Nelson, B. D. (2006). *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. *Mol. Plant Pathol.* 7, 1–16. doi: 10.1111/j.1364-3703.2005.00316.x

Bos, L. (1970). The identification of three new viruses isolated from *Wisteria* and *Pisum* in The Netherlands, and the problem of variation within the potato virus Y group. *Neth. J. Plant Pathol.* 76, 8–46. doi: 10.1007/BF01976763

Bos, L., Hampton, R. O., and Makkouk, K. M. (1988). "Viruses and virus diseases of pea, lentil, faba bean and chickpea," in *World crops: cool season food legumes: A global perspective of the problems and prospects for crop improvement in pea, lentil, faba bean and chickpea*, ed. R. J. Summerfield, (Dordrecht: Springer), 591–615.

Brault, V., Tanguy, S., Reinbold, C., Le Trionnaire, G., Arneodo, J., Jaubert-Possamai, S., et al. (2010). Transcriptomic analysis of intestinal genes following acquisition of pea enation mosaic virus by the pea aphid *Acyrthosiphon pisum*. *J. Gen. Virol.* 91, 802–808. doi: 10.1099/vir.0.012856-0

Briddon, R. W., Martin, D. P., Roumagnac, P., Navas-Castillo, J., Fiallo-Olivé, E., Moriones, E., et al. (2018). Alphasatellitidae: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Arch. Virol.* 163, 2587–2600. doi: 10.1007/s00705-018-3854-2

Bruenn, J. A., Warner, B. E., and Yerramsetty, P. (2015). Widespread mitovirus sequences in plant genomes. *PeerJ* 3:e876. doi: 10.7717/peerj.876

Clement, S. L. (2006). Pea aphid outbreaks and virus epidemics on peas in the US Pacific Northwest: histories, mysteries, and challenges. *Plant Health Prog.* 7:34. doi: 10.1094/PHP-2006-1018-01-RV

Coetzee, B., Freeborough, M.-J., Maree, H. J., Celton, J.-M., Rees, D. J. G., and Burger, J. T. (2010). Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology* 400, 157–163. doi: 10.1016/j.virol.2010.01.023

Czotter, N., Molnar, J., Szabó, E., Demian, E., Kontra, L., Baksa, I., et al. (2018). NGS of virus-derived small RNAs as a diagnostic method used to determine viromes of Hungarian vineyards. *Front. Microbiol.* 9:758. doi: 10.3389/fmicb.2018.00122

Demler, S. A., and de Zoeten, G. A. (1989). Characterization of a satellite RNA associated with pea enation mosaic virus. *J. Gen. Virol.* 70, 1075–1084. doi: 10.1099/0022-1317-70-5-1075

Demler, S. A., Rucker, D. G., Nooruddin, L., and de Zoeten, G. A. (1994). Replication of the satellite RNA of pea enation mosaic virus is controlled by RNA 2-encoded functions. *J. Gen. Virol.* 75(Pt 6), 1399–1406. doi: 10.1099/0022-1317-75-6-1399

Drijfhout, E., and Bos, L. (1977). The identification of two new strains of bean common mosaic virus. *Neth. J. Plant Pathol.* 83, 13–25. doi: 10.1007/BF01976508

Elbeaino, T., Digiaro, M., Mielke-Ehret, N., Muehlbach, H.-P., Martelli, G. P., and Ictv, R. C. (2018). ICTV Virus Taxonomy Profile: *Fimoviridae*. *J. Gen. Virol.* 99, 1478–1479. doi: 10.1099/jgv.0.001143

Elbeaino, T., Digiaro, M., Uppala, M., and Sudini, H. (2014). Deep sequencing of pigeonpea sterility mosaic virus discloses five RNA segments related to emaraviruses. *Virus Res.* 188, 27–31. doi: 10.1016/j.virusres.2014.03.022

Elbeaino, T., Digiaro, M., Uppala, M., and Sudini, H. (2015). Deep sequencing of dsRNAs recovered from mosaic-diseased pigeonpea reveals the presence of a novel emaravirus: pigeonpea sterility mosaic virus 2. *Arch. Virol.* 160, 2019–2029. doi: 10.1007/s00705-015-2479-y

EPPO, (2019). *EPPO A1 List*. Available online at: https://www.eppo.int/ACTIVITIES/plant_quarantine/A1_list [accessed August 30, 2019].

Gaafar, Y. Z. A., and Ziebell, H. (2019a). Complete genome sequence of a highly divergent carrot torradovirus 1 strain from *Apium graveolens*. *Arch. Virol.* 164, 1943–1947. doi: 10.1007/s00705-019-04272-3

Gaafar, Y. Z. A., and Ziebell, H. (2019b). Two divergent isolates of turnip yellows virus from pea and rapeseed and first report of turnip yellows virus-associated RNA in Germany. *Microbiol. Resour. Announc.* 8, 00214–19e. doi: 10.1128/MRA.00214-19

Gaafar, Y. Z. A., Cordsen Nielsen, G., and Ziebell, H. (2018). Molecular characterisation of the first occurrence of *Pea necrotic yellow dwarf virus* in Denmark. *N. Dis. Rep.* 37:16. doi: 10.5197/j.2044-0588.2018.037.016

Gaafar, Y. Z. A., Grausgruber-Gröger, S., and Ziebell, H. (2016). Vicia faba. V. sativa and Lens culinaris as new hosts for Pea necrotic yellow dwarf virus in Germany and Austria. *N. Dis. Rep.* 34:28. doi: 10.5197/j.2044-0588.2016.034.028

Gaafar, Y. Z. A., Richert-Pöggeler, K. R., Maaß, C., Vetten, H.-J., and Ziebell, H. (2019a). Characterisation of a novel nucleorhabdovirus infecting alfalfa (*Medicago sativa*). *Virol J.* 16:113. doi: 10.1186/s12985-019-1147-3

Gaafar, Y. Z. A., Richert-Pöggeler, K. R., Sieg-Müller, A., Lüddecke, P., Herz, K., Hartrick, J., et al. (2019c). A divergent strain of melon chlorotic spot virus isolated from black medic (*Medicago lupulina*) in Austria. *Virol J.* 16:297. doi: 10.1186/s12985-019-1195-8

Gaafar, Y. Z. A., Richert-Pöggeler, K. R., Sieg-Müller, A., Lüddecke, P., Herz, K., Hartrick, J., et al. (2019d). Caraway yellows virus, a novel nepovirus from *Carum carvi*. *Virol J.* 16:529. doi: 10.1186/s12985-019-1181-1

Gaafar, Y. Z. A., Sieg-Müller, A., Lüddecke, P., Herz, K., Hartrick, J., Maaß, C., et al. (2019b). First report of *Turnip crinkle virus* infecting garlic mustard (*Alliaria petiolata*) in Germany. *N. Dis. Rep.* 39:9. doi: 10.5197/j.2044-0588.2019.039.009

Gaafar, Y. Z. A., Timchenko, T., and Ziebell, H. (2017). First report of *Pea necrotic yellow dwarf virus* in The Netherlands. *N. Dis. Rep.* 35:23. doi: 10.5197/j.2044-0588.2017.035.023

Gallet, R., Kraberger, S., Filloux, D., Galzi, S., Fontes, H., Martin, D. P., et al. (2018). Nanovirus-alphasatellite complex identified in *Vicia cracca* in the Rhône delta region of France. *Arch. Virol.* 163, 695–700. doi: 10.1007/s00705-017-3634-4

Graichen, K. (1994). Nachweis von Resistenz gegenüber dem Turnip yellows luteovirus (TuYV) in Winterraps und verwandten Arten. *Vortr. Pflanzenzüchtung* 30, 132–143.

Graichen, K., and Schliephake, E., eds (1999). Infestation of winter oilseed rape by turnip yellows luteovirus and its effect on yield in Germany.In "*Proceedings of 10th International Rapeseed Congress—New horizons for an old crop*,". eds, N Wratten, and PA Salisbury, (Canberra, ACT: International Consultative Group for Rapeseed Research), pp. 131–136

Grigoras, I., Ginzo, A. I. D. C., Martin, D. P., Varsani, A., Romero, J., Mammadov, A. C., et al. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *J. Gen. Virol.* 95, 1178–1191. doi: 10.1099/vir.0.063115-0

Grigoras, I., Gronenborn, B., and Vetten, H. J. (2010a). First report of a nanovirus disease of pea in Germany. *Plant Dis.* 94:642. doi: 10.1094/PDIS-94-5-0642C

Grigoras, I., Timchenko, T., Grande-Pérez, A., Katul, L., Vetten, H.-J., and Gronenborn, B. (2010b). High variability and rapid evolution of a nanovirus. *J. Virol.* 84, 9105–9117. doi: 10.1128/JVI.00607-10

Guyatt, K. J., Proll, D. F., Menssen, A., and Davidson, A. D. (1996). The complete nucleotide sequence of bean yellow mosaic potyvirus RNA. *Arch. Virol.* 141, 1231–1246. doi: 10.1007/BF01718827

Hagedorn, D. J., and Khan, T. N. (1984). *Compendium of Pea Diseases*. St Paul, MN: American Phytopathological Society.

Heydarnejad, J., Kamali, M., Massumi, H., Kvarnheden, A., Male, M. F., Kraberger, S., et al. (2017). Identification of a nanovirus-alphasatellite complex in *Sophora alopecuroides*. *Virus. Res.* 235, 24–32.

Idris, A. M., Shahid, M. S., Briddon, R. W., Khan, A. J., Zhu, J.-K., and Brown, J. K. (2011). An unusual alphasatellite associated with monopartite begomoviruses attenuates symptoms and reduces betasatellite accumulation. *J. Gen. Virol.* 92, 706–717. doi: 10.1099/vir.0.025288-0

Jo, Y., Lian, S., Chu, H., Cho, J. K., Yoo, S.-H., Choi, H., et al. (2018). Peach RNA viromes in six different peach cultivars. *Sci. Rep.* 8:1844. doi: 10.1038/s41598-018-20256-w

Kaiser, W. J. (1982). Natural hosts and vectors of tobacco streak virus in Eastern Washington. *Phytopathology* 72:1508. doi: 10.1094/Phyto-72-1508

Khalifa, M. E., and Pearson, M. N. (2013). Molecular characterization of three mitoviruses co-infecting a hypovirulent isolate of *Sclerotinia sclerotiorum* fungus. *Virology* 441, 22–30. doi: 10.1016/j.virol.2013.03.002

Khetarpal, R. K., and Maury, Y. (1987). Pea seed-borne mosaic virus: a review. *Agronomie* 7, 215–224.

Koloniuk, I., Přibylová, J., and Fránová, J. (2018). Molecular characterization and complete genome of a novel nepovirus from red clover. *Arch. Virol.* 163, 1387–1389. doi: 10.1007/s00705-018-3742-9

Kon, T., Rojas, M. R., Abdourhamane, I. K., and Gilbertson, R. L. (2009). Roles and interactions of begomoviruses and satellite DNAs associated with okra leaf curl disease in Mali. *West Afr. J. Gen. Virol.* 90, 1001–1013. doi: 10.1099/vir.0.008102-0

Kraft, J. M., ed (2008). *Compendium of Pea Diseases and Pests*. St. Paul: APS Press.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Latham, L. J., and Jones, R. A. C. (2001). Alfalfa mosaic and pea seed-borne mosaic viruses in cool season crop, annual pasture, and forage legumes: susceptibility, sensitivity, and seed transmission. *Aust. J. Agric. Res.* 52:771. doi: 10.1071/AR00165

Li, R., Gao, S., Hernandez, A. G., Wechter, W. P., Fei, Z., and Ling, K.-S. (2012). Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PLoS One* 7:e37127. doi: 10.1371/journal.pone.0037127

MacDiarmid, R., Rodoni, B., Melcher, U., Ochoa-Corona, F., and Roossinck, M. (2013). Biosecurity implications of new technology and discovery in plant virus research. *PLoS Pathog.* 9:e1003337. doi: 10.1371/journal.ppat.1003337

Mar, T. B., Mendes, I. R., Lau, D., Fiallo-Olivé, E., Navas-Castillo, J., Alves, M. S., et al. (2017). Interaction between the New World begomovirus *Euphorbia yellow mosaic virus* and its associated alphasatellite: effects on infection and transmission by the whitefly *Bemisia tabaci*. *J. Gen. Virol.* 98, 1552–1562. doi: 10.1099/jgv.0.000814

Maree, H. J., Fox, A., Al Rwahnih, M., Boonham, N., and Candresse, T. (2018). Application of HTS for routine plant virus diagnostics: state of the art and challenges. *Front. Plant Sci.* 9:1082. doi: 10.3389/fpls.2018.01082

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., et al. (2017). A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. *Front. Microbiol.* 8:45 doi: 10.3389/fmicb.2017.00045

McKenzie, D. L., and Morrall, R. A. A. (1975). Diseases of specialty crops in Saskatchewan: II. Notes on field pea in 1973-74 and on lentil in 1973. *Canad. Plant Dis. Sur.* 55, 97–100.

Morishita, M., and Takafuji, A. (1999). Diapause characteristics of the Kanzawa spider mite, Tetranychus kanzawai Kishida, in pea fields of Central Wakayama Prefecture, Japan. *Jap. J. Appl. Entomol. Zoo.* 43, 185–188. doi: 10.1303/jjaez.43.185

Musil, M. (1966). Uber das Vorkommen des Virus des Blattrollens der Erbse in der Siowakei. (On the occurrence of a leaf roll virus in pea in Slovakia.). *Biol. Bratisl.* 21, 133–138.

Nakazono-Nagaoka, E., Takahashi, T., Shimizu, T., Kosaka, Y., Natsuaki, T., Omura, T., et al. (2009). Cross-protection against bean yellow mosaic virus (BYMV) and clover yellow vein virus by attenuated BYMV isolate M11. *Phytopathology* 99, 251–257. doi: 10.1094/PHYTO-99-3-0251

Nawaz-Ul-Rehman, M. S., Nahid, N., Mansoor, S., Briddon, R. W., and Fauquet, C. M. (2010). Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. *Virology* 405, 300–308. doi: 10.1016/j.virol.2010.06.024

Nibert, M. L., Vong, M., Fugate, K. K., and Debat, H. J. (2018). Evidence for contemporary plant mitoviruses. *Virology* 518, 14–24. doi: 10.1016/j.virol.2018.02.005

Olmos, A., Boonham, N., Candresse, T., Gentit, P., Giovani, B., Kutnjak, D., et al. (2018). High-throughput sequencing technologies for plant pest diagnosis: challenges and opportunities. *EPPO Bull.* 48, 219–224. doi: 10.1111/epp.12472

Olson, J., Rebek, E., and Schnelle, M. (2017). *Rose rosette disease*. Availble online at: http://factsheets.okstate.edu/documents/epp-7329-rose-rosette-disease, [accessed October 18, 2019]

Pandey, B., Naidu, R. A., and Grove, G. G. (2018). Detection and analysis of mycovirus-related RNA viruses from grape powdery mildew fungus *Erysiphe necator*. *Arch. Virol.* 163, 1019–1030. doi: 10.1007/s00705-018-3714-0

Paprotka, T., Metzler, V., and Jeske, H. (2010). The first DNA 1-like alpha satellites in association with New World begomoviruses in natural infections. *Virology* 404, 148–157. doi: 10.1016/j.virol.2010.05.003

Patil, B. L., and Kumar, P. L. (2015). Pigeonpea sterility mosaic virus: a legume-infecting Emaravirus from South Asia. *Mol. Plant Pathol.* 16, 775–786. doi: 10.1111/mpp.12238

Pecman, A., Kutnjak, D., Gutiérrez-Aguirre, I., Adams, I., Fox, A., Boonham, N., et al. (2017). Next generation sequencing for detection and discovery of plant viruses and viroids: comparison of two approaches. *Front. Microbiol.* 8:1998. doi: 10.3389/fmicb.2017.01998

Purdy, L. H. (1979). *Sclerotinia sclerotiorum*: history, diseases and symptomatology, host range, geographic distribution, and impact. *Phytopathology* 69:875. doi: 10.1094/Phyto-69-875

R Core Team, (2019). *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

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

Rott, M., Xiang, Y., Boyes, I., Belton, M., Saeed, H., Kesanakurti, P., et al. (2017). Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroids. *Plant Dis.* 101, 1489–1499. doi: 10.1094/PDIS-03-17-0306-RE

Sanger, M., Passmore, B., Falk, B. W., Bruening, G., Ding, B., and Lucas, W. J. (1994). Symptom severity of beet western yellows virus strain ST9 is conferred by the ST9-associated RNA and is not associated with virus release from the phloem. *Virology* 200, 48–55. doi: 10.1006/viro.1994.1161

Saucke, H., Uteau, D., Brinkmann, K., and Ziebell, H. (2019). Symptomology and yield impact of pea necrotic yellow dwarf virus (PNYDV) in faba bean (*Vicia faba* L. minor). *Eur. J. Plant Pathol.* 153, 1299–1315. doi: 10.1007/s10658-018-01643-5

Stevens, M., Smith, H. G., and Hallsworth, P. B. (1994). The host range of beet yellowing viruses among common arable weed species. *Plant Pathol.* 43, 579–588. doi: 10.1111/j.1365-3059.1994.tb01593.x

Syller, J. (2003). Molecular and biological features of umbraviruses, the unusual plant viruses lacking genetic information for a capsid protein. *Physiol. Mole. Plant Pathol.* 63, 35–46. doi: 10.1016/j.pmpp.2003.08.004

Takafuji, A., and Morishita, M. (2001). Overwintering ecology of two species of spider mites (Acari: Tetranychidae) on different host plants. *Appl. Entomol. Zool.* 36, 169–175. doi: 10.1303/aez.2001.169

Timchenko, T., Katul, L., Aronson, M., Vega-Arreguín, J. C., Ramirez, B. C., Vetten, H. J., et al. (2006). Infectivity of nanovirus DNAs: induction of disease

by cloned genome components of *Faba bean necrotic yellows virus*. *J. Gen. Virol.* 87, 1735–1743. doi: 10.1099/vir.0.81753-0

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucl. Acids Res.* 40:e115. doi: 10.1093/nar/gks596

Vetten, H. J., Knierim, D., Rakoski, M. S., Menzel, W., Maiss, E., Gronenborn, B., et al. (2019). Identification of a novel nanovirus in parsley. *Arch. Virol.* 164, 1883–1887. doi: 10.1007/s00705-019-04280-3

Wu, M. D., Zhang, L., Li, G. Q., Jiang, D. H., Hou, M. S., and Huang, H.-C. (2007). Hypovirulence and double-stranded RNA in *Botrytis cinerea*. *Phytopathology* 97, 1590–1599. doi: 10.1094/PHYTO-97-12-1590

Xie, J., and Ghabrial, S. A. (2012). Molecular characterization of two mitoviruses co-infecting a hypovirulent isolate of the plant pathogenic fungus *Sclerotinia sclerotiorum* corrected. *Virology* 428, 77–85. doi: 10.1016/j.virol.2012.03.015

Ziebell, H. (2017). Die Virusepidemie an Leguminosen 2016 – eine Folge des Klimawandels? *J. Kult. Pflanzen* 69, 64–68. doi: 10.1399/JFK.2017.02.09

Ziebell, H. (2019). Ertragsausfälle durch Leguminosenviren – Vektorkontrolle notwendig!. *Raps* 37, 27–28.

Conflict of Interest: JF, AB, and RM were employed by The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 December 2020
doi: 10.3389/fmicb.2020.616171

Identification of Two New Isolates of *Chilli veinal mottle virus* From Different Regions in China: Molecular Diversity, Phylogenetic and Recombination Analysis

Shaofei Rao[1], Xuwei Chen[1], Shiyou Qiu[1], Jiejun Peng[1], Hongying Zheng[1], Yuwen Lu[1], Guanwei Wu[1], Jianping Chen[1], Wen Jiang[2], Yachun Zhang[3]* and Fei Yan[1]**

[1] State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Key Laboratory of Biotechnology in Plant Protection of Ministry of Agriculture and Zhejiang Province, Institute of Plant Virology, Ningbo University, Ningbo, China, [2] Biotechnology Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China, [3] Dali Bai Autonomous Prefecture Academy of Agricultural Science and Extension, Dali, China

OPEN ACCESS

Edited by:
Xifeng Wang,
State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection (CAAS), China

Reviewed by:
Inmaculada Ferriol,
Centre for Research in Agricultural Genomics (CRAG), Spain
Subir Sarker,
La Trobe University, Australia

***Correspondence:**
Wen Jiang
jiangwengx@126.com
Yachun Zhang
zhangyachun@yeah.net
Fei Yan
yanfei@nbu.edu.cn

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 11 October 2020
Accepted: 01 December 2020
Published: 23 December 2020

Citation:
Rao S, Chen X, Qiu S, Peng J, Zheng H, Lu Y, Wu G, Chen J, Jiang W, Zhang Y and Yan F (2020) Identification of Two New Isolates of Chilli veinal mottle virus From Different Regions in China: Molecular Diversity, Phylogenetic and Recombination Analysis. Front. Microbiol. 11:616171. doi: 10.3389/fmicb.2020.616171

Chilli veinal mottle virus (ChiVMV) is an important plant pathogen with a wide host range, causing serious yield losses in pepper production all over the world. Recombination is a major evolutionary event for single-stranded RNA viruses, which helps isolates adapt to new environmental conditions and hosts. Recombination events have been identified in multiple potyviruses, but so far, there have been no reports of recombination events among the ChiVMV population. We here detected ChiVMV in pepper samples collected from Guangxi and Yunnan provinces for the first time and amplified the nearly full-length sequences. Phylogenetic and recombination analysis were performed using the new sequences and the 14 full-length and 23 capsid protein (CP) sequences available in GenBank. Isolates tend to cluster on a geographical basis, indicating that geographic-driven evolution may be an important determinant of ChiVMV genetic differences. A total of 10 recombination events were detected among the ChiVMV sequences using RDP4 with a strict algorithm, and both the Guangxi and Yunnan isolates were identified as recombinants. Recombination appears to be a significant factor affecting the diversity of ChiVMV isolates.

Keywords: ***Chilli veinal mottle virus,*** **phylogenetic analysis, recombination, genetic diversity, evolution**

INTRODUCTION

Chilli veinal mottle virus (ChiVMV) is a member of the genus *Potyvirus* in the family *Potyviridae*. It is a very common virus in chilli pepper (*Capsicum annum* L.) in east Asian countries, causing serious losses in pepper production (Ong et al., 1980; Moury et al., 2005; Wang et al., 2006; Tsai et al., 2008; Shah et al., 2009). In addition to *Capsicum annuum*, ChiVMV can also infect many other plants in the family *Solanaceae*, including *Nicotiana tabacum*, *Solanum lycopersicum*, *Solanum melongena*, and *Datura stramonium* (Tsai et al., 2008; Ding et al., 2011; Yang et al., 2013; Zhao et al., 2014; Kaur et al., 2015). Symptoms of ChiVMV infection include mosaic mottling, twisted or fallen leaves, vein banding, and reduced fruit size (Tsai et al., 2008; Gao et al., 2016). The genome of ChiVMV is a single-stranded sense RNA, about 9.7 kb excluding the poly (A) tail. It encodes a

polyprotein, which is then cleaved by virus-encoded proteases into 10 mature functional proteins (Yang et al., 2013; Gao et al., 2016). *Aphis gossypii* has been reported to transmit the virus in a non-persistent manner in solanaceous crops (Shah et al., 2008).

The public sequence databases currently contain the complete genome sequences of 14 ChiVMV isolates, all of which are from Asia. Several studies have reported genetic differences within the species based on analysis of CP sequences (Moury et al., 2005; Yang et al., 2013; Gao et al., 2016; Ahmad and Ashfaq, 2018) and there are currently 98 such sequences available. Discovering and determining the sequences of more isolates worldwide is important for our understanding of the molecular diversity and evolution of the virus. For single-stranded RNA viruses, recombination is a major evolutionary event that helps isolates adapt to new environmental conditions and hosts (Simon-Loriere and Holmes, 2011). Recombination events have been identified in many potyviruses (Revers et al., 1996; Gagarinova et al., 2008; Seo et al., 2009) but, so far, there have been no reports of recombination events among the ChiVMV population. In this study, we determined the nearly full-length sequences of ChiVMV in pepper from Guangxi and Yunnan provinces, China and used the data to analyze molecular diversity and recombination events among ChiVMV isolates.

MATERIALS AND METHODS

Whole-Genome Sequencing of Two New ChiVMV Isolates

From May to July 2020, we collected pepper samples with suspected viral disease symptoms (including dead tops, mosaic, mottling, wrinkled leaves, and chlorosis) from pepper fields in Guangxi and Yunnan provinces in China. Total RNA was extracted from infected pepper fruits using the Trizol method, and first strand cDNA was synthesized using a reverse transcription kit (Toyobo) according to the manufacturer's instructions. The complete genome sequences of the two new isolates were amplified from five over-lapping fragments using specific primers (**Supplementary Table 1**). KOD neo enzyme (Toyobo) was used for PCR amplification, and the amplified fragments were purified with an Omega gel extraction kit and cloned into the pEASY-Blunt Zero vector (TransGen). At least two clones for each fragment were picked and sent for sequencing. Sequences were assembled using Vector NTI version 10. The complete genome sequences of the two isolates have been deposited in GenBank under accession numbers MT782116 (Guangxi) and MT974520 (Yunnan). Sequence analysis and comparison of the two new isolates to the other reported sequences were done using MEGA X (Kumar et al., 2018). The complete genome sequences and CP sequences of other ChiVMV isolates were downloaded from the National Center for Biotechnology Information (NCBI) (**Supplementary Tables 2**, **3**).

Construction of Phylogenetic Trees

The whole genome sequences of 16 ChiVMV isolates and 25 CP-coding region sequences were used for phylogenetic analysis in the MEGA X software package (Kumar et al., 2018). The best-fit nucleotide substitution models for the full-length sequences of 16 isolates and the 25 CP sequences were determined using the function in MEGA X to be, respectively, GTR + G + I (General Time Reversible + Gama Distributed With Invariant Sites) and T92 + G (Tamura 3-parameter + Gama Distributed). The trees were then constructed by the maximum-likelihood (ML) method according to the corresponding model with 1,000 bootstrap replicates. The sequence of the OKP41 isolate of pepper vein mottle virus (PVMV), a closely-related member of the genus *Potyvirus*, was used as an outgroup.

Recombination and Selection Pressure Analysis

The whole genome sequences of 16 ChiVMV isolates and 25 CP-coding region sequences were used for recombination analysis. The six methods in the RDP4 software, namely RDP, GENECONV, BOOTSCAN, MaxChi, Chimera, and SISCAN were used to find possible parental isolates and recombination breakpoints with the default parameters (Martin et al., 2015). Site specific selection pressure in CP coding sequences, was determined by single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), mixed effects model of evolution (MEME) with p-value threshold of 0.1, and fast unconstrained Bayesian approximation (FUBAR) with posterior probability of 0.9 implemented in the Hyphy package[1] (Pond and Frost, 2005).

RESULTS

Sequencing and Molecular Diversity Analysis of ChiVMV Isolates From Guangxi and Yunnan

By using specific primers, we amplified overlapping fragments by RT-PCR, and assembled the nearly complete genome sequences of the ChiVMV Guangxi and Yunnan isolates. The sequences were 9,722 (Guangxi) and 9,724 (Yunnan) nucleotides long excluding the 3′-end poly A tail, and both contained a predicted open reading frame of 9,270 nt, encoding a poly-protein of 3,089 amino acids. In comparisons of their full-length genome sequences, the new isolates had 79–92.6% (Guangxi) and 80.6–89.8% (Yunnan) nt identity to the previously reported isolates while the corresponding values for comparison with the outgroup PVMV control sequence were, respectively, 67.4 and 68% (**Table 1** and **Supplementary Table 4**). The divergence of the Guangxi isolate is about 0.08–0.25 and that of the Yunnan isolate is about 0.11–0.23 (**Table 1** and **Supplementary Table 4**).

Phylogenetic Relationships of ChiVMV Isolates Worldwide

A phylogenetic tree using all available full-length ChiVMV sequences with PVMV as an outgroup divides the isolates into two major clades (**Figure 1**). The first major clade has a sub-clade containing two isolates each from Hunan and Korea, and one

[1]http://www.Datamonkey.org

TABLE 1 | Nucleotide sequence similarity and divergence among the 16 ChiVMV isolates worldwide based on their full genome sequences.

	MT782116 GX		MT974520 YN	
	Percent similarity	**Percent divergence**	**Percent similarity**	**Percent divergence**
KU987835.1 GD	92.5	8	81.9	21.1
KR296797.1 HN	92	8.5	81.5	21.5
JX088636.1 YN tobacco	79	25.1	82.6	20.3
GQ981316.1 Wenchang	92.6	7.9	81.9	21.2
AJ972878.1 Ca	92.1	8.2	81.3	21.8
KC711055.1 Yp8	81.7	21.4	89.8	11.1
KC711056.1 Pp4	81.8	21.1	89.6	11.3
GU170808.1 Ch-War	85.2	16.8	82.2	20.5
GU170807.1 Ch-Jal	85.4	16.5	82.2	20.6
AM909717.1 Korea	92.2	8.2	81.4	21.8
MN207122.1 PK	88.1	13	83.4	18.8
MK405594.1 SichuanLuzhou	81.7	21.6	89.7	11.2
NC_005778.1	85.7	16	80.6	22.8
LN832362.1 hn	92.1	8.3	81.5	21.5
MT782116 GX	–	–	81.8	21.3
MT974520 YN	81.8	21.3	–	–
LC438545.1 OKP41	67.4	42.9	68	42.2

The values were calculated by DNAStar. The four columns of values represent the percent of nucleotide sequence similarity and divergence between the Guangxi and Yunnan isolates and each of the 14 isolates previously reported. See ***Supplementary Table 4*** *for other pairwise comparison values. The geographical origin of the isolates listed are: KU987835.1 GD, Guangdong, China; KR296797.1 HN, Hunan, China; JX088636.1 YN tobacco, Yunnan, China; GQ981316.1 Wenchang, Hainan, China; AJ972878.1 Ca, Korea; KC711055.1 Yp8, Sichuan, China; KC711056.1 Pp4, Sichuan, China; GU170808.1 Ch-War, India; GU170807.1 Ch-Jal, India; AM909717.1 Korea, Korea; MN207122.1 PK, Pakistan; MK405594.1 SichuanLuzhou, Sichuan, China; NC_005778.1, India; LN832362.1 hn, Hunan, China; MT782116 GX, Guangxi, China; MT974520 YN, Yunnan, China. LC438545.1 is an isolate of pepper vein mottle virus (PVMV).*

isolate each from Guangdong, Wenchang, Guangxi (our isolate), and India. The second sub-clade is formed by two Indian and one Pakistani isolates. The second major clade contains three Sichuan isolates and one Yunnan tobacco isolate (**Figure 1**). The newly identified Yunnan pepper isolate was also included in the second clade. Thus our Guangxi and Yunnan isolates were more similar to the isolates from adjacent regions of China (such as Wenchang, Guangdong for Guangxi isolate, and Sichuan for Yunnan isolate), than to those from the more distant provinces.

The capsid protein (CP) gene is very important for the infection cycle of potyviruses. Its primary function is to encode the virus coat protein (Revers and García, 2015) and this region has often been used as the basis for comparisons to establish the taxonomy of potyviruses (Ball, 2005; Tsai et al., 2008). To better understand the genetic variability of the ChiVMV population, we selected 25 CP coding region sequences from different geographical locations to construct a phylogenetic tree. The tree had two well-defined major clades (**Figure 2**). The first clade contains isolates from Indonesia, Thailand, India, Vietnam, South Korea and China (Guangxi, Hainan and Taiwan). Isolates from Yunnan, Liaoning, Sichuan of China and Pakistan constitute the second clade (**Figure 2**).

Recombination and Selection Pressure Analysis

To analyze possible recombination signals in the ChiVMV population, we used RDP4 software to predict recombination events among the full-length sequences. Recombination events identified by at least three methods and *P* value less than 1×10^{-6} were considered credible and 10 recombination events were detected in total (**Table 2**). The Guangxi and Yunnan isolates were predicted to be recombinants. The recombination event of the Guangxi isolate occurred between nts 2,756 and 5,284. The Hunan isolate and the Wenchang (Hainan Province) isolate were predicted to be possible parents. Interestingly, Hunan and Hainan are both neighboring provinces to Guangxi. The recombination of Yunnan isolate occurred between nts 31 and 1,400, and Pakistan isolate and Yunnan tobacco isolate were predicted as possible parents (**Table 2**). The analysis indicated that geography-related recombination was likely a key factor in the evolution of ChiVMV.

We also detected two recombination events in the Yunnan tobacco isolate, at nt positions 9,316–9,739 and 27–1,696, respectively. One recombination event was detected in each of the three Indian isolates, GU170808, GU170807, and NC_005778. Three recombination events were detected in the MN207122-PK isolate, at various sites in the region 27–4,480 (**Table 2**).

Whether there were recombination events in the 25 selected CP sequences from different geographical locations was further analyzed, and 3 recombination events were found (**Supplementary Table 5**). Two Chinese isolates and one Pakistan isolate were predicted to be recombinants, indicating that the recombination contributed to the variation of the CP sequences (**Supplementary Table 5**).

We also performed selection pressure analysis on the CP coding sequences (287 aa), and found that many of the codons are subject to negative selection. A total of 87, 136, and 160 negatively selected codons were detected in the CP region by SLAC, FEL, and FUBAR methods, respectively. The codons 83,

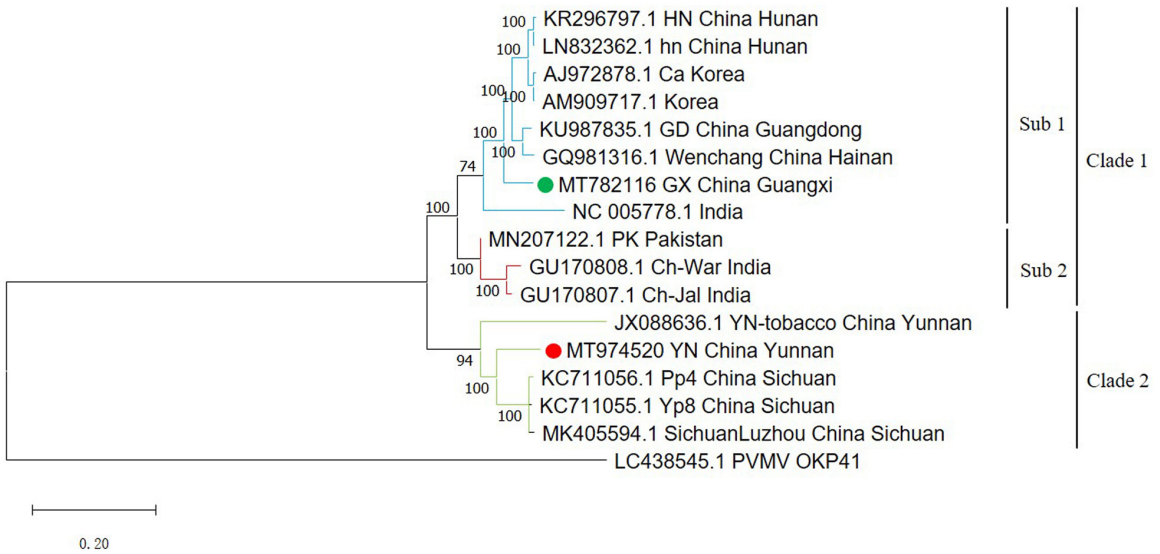


FIGURE 1 | Phylogenetic tree of 16 full-length ChiVMV isolates. The phylogenetic tree of 16 full-length ChiVMV isolates was produced using MEGA X by the maximum-likelihood (ML) method using the General Time Reversible algorithm and 1,000 bootstrap replications. The ChiVMV isolates are divided into three sub-clades which are shown in different colors. The new Guangxi and Yunnan isolates are marked by green and red circles, respectively. The number at each branch of the phylogenetic tree is the bootstrap percentage. A pepper veinal mottle virus isolate sequence (LC438545) was used as outgroup.

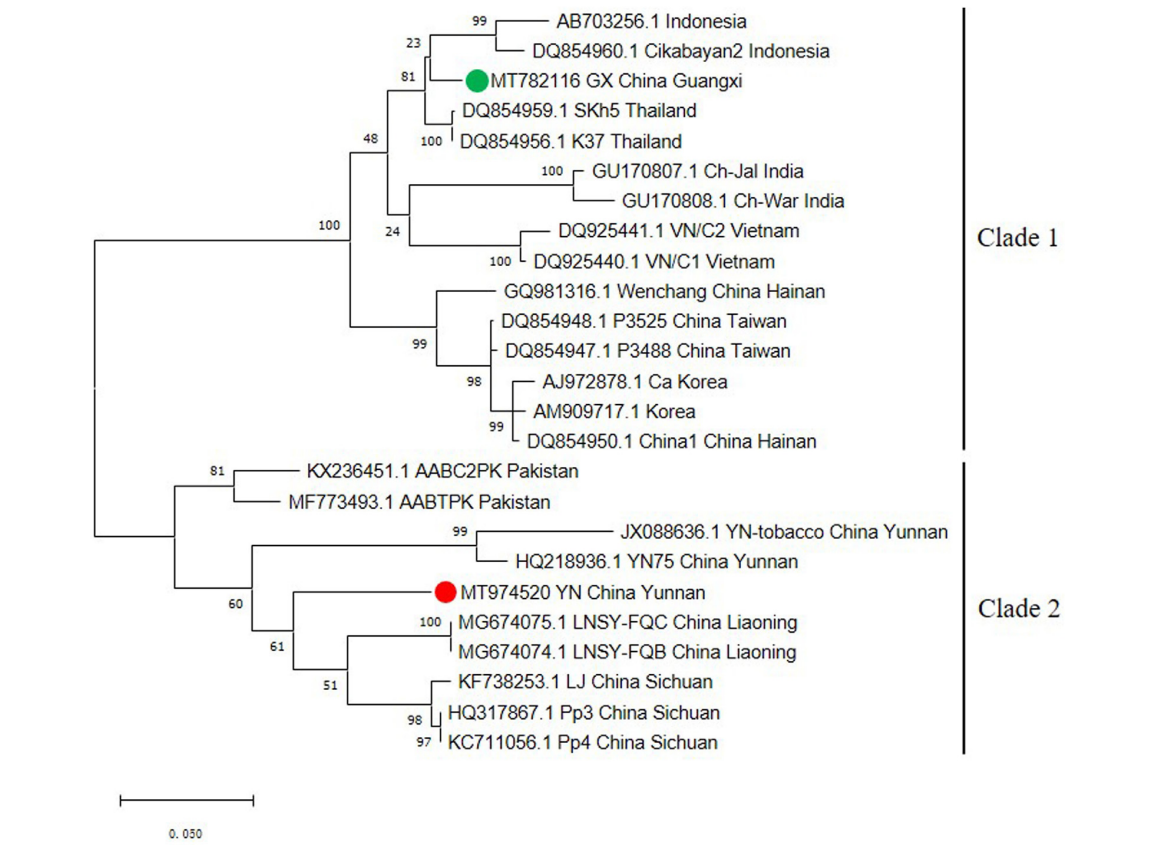


FIGURE 2 | Phylogenetic tree based on the nucleotide sequences of the CP gene of 25 ChiVMV isolates. The phylogenetic tree was produced using MEGA X by the maximum-likelihood (ML) method using the Tamura 3-parameter algorithm and 1,000 bootstrap replications. The number at each branch of the phylogenetic tree is the bootstrap percentage. The tree was divided into two clades and the new Guangxi and Yunnan isolates are marked by green and red circles, respectively.

TABLE 2 | Summary of possible recombination events among 16 full-length ChiVMV isolates identified by RDP4.

Event number	Begin	End	Recombinant sequence(s)	Minor parental sequence(s)	Major parental sequence(s)	*P*-value for the six detection methods in RDP4					
						RDP	GENECONV	Bootscan	Maxchi	Chimaera	SiSscan
1	2756	5284	MT782116 GX	GQ981316.1 Wenchang	Unknown (LN832362.1_hn)	3.55E-07	NS	4.92E-06	5.46E-08	1.34E-08	6.53E-07
2	31	1400	MT974520 YN	JX088636.1 YN tobacco	MN207122.1 PK	2.20E-44	4.85E-21	1.32E-38	3.83E-05	5.10E-17	1.53E-34
3	9316	9739	JX088636.1 YN tobacco	KC711055.1_Yp8	Unknown (MN207122.1 PK)	8.99E-06	NS	1.95E-06	NS	NS	2.23E-08
4	27	1696	JX088636.1 YN tobacco	MK405594.1 SichuanLuzhou	Unknown (GQ981316.1 Wenchang)	1.01E-25	4.01E-18	8.95E-22	8.12E-13	1.07E-09	8.80E-34
5	5604	5854	GU170808.1 Ch-War	Unknown (MN207122.1 PK)	GU170807.1 Ch-Jal	1.29E-19	7.58E-17	9.33E-20	2.27E-10	2.56E-09	1.56E-13
6	3035	4198	GU170807.1 Ch-Jal	KC711055.1_Yp8	AM909717.1 Korea	2.15E-10	1.72E-03	1.02E-10	2.69E-06	1.91E-07	4.29E-15
7	5658	8580	NC-005778.1 India	Unknown (GU170807.1 Ch-Jal)	AM909717.1 Korea	1.77E-09	NS	3.09E-09	2.00E-13	7.24E-07	9.45E-27
8	2401	2822	MN207122.1 PK	KC711056.1_Pp4	LN832362.1_hn	9.66E-13	1.24E-09	1.38E-12	2.05E-08	2.23E-05	1.22E-11
9	1894	4480	MN207122.1 PK	KR296797.1_HN	GU170808.1 Ch-War	4.93E-06	2.59E-05	8.94E-06	2.40E-10	3.35E-10	NS
10	27	1389	MN207122.1 PK	Unknown (JX088636.1 YN tobacco)	MK405594.1 SichuanLuzhou	1.21E-34	1.80E-20	1.37E-33	1.74E-14	9.05E-17	1.28E-34

*The whole genome sequences of the two newly identified isolates and 14 ChiVMV full-length sequences extracted from public nucleic acid databases were used for recombination analysis. The six methods in the RDP4 software, namely RDP, GENECONV, BOOTSCAN, MaxChi, Chimaera, and SISCAN were used to find possible parental isolates and recombination breakpoints with the default parameters. Recombination events identified by at least three methods and P value less than 1×10^{-6} are listed. For the geographical origin of the isolates, see **Table 1**. NS, Not Significant.*

173, 176, 231, 239, 261, and 268 were found to be under positive selection by MEME.

DISCUSSION

In this study, we detected ChiVMV in pepper disease samples collected from Guangxi and Yunnan provinces for the first time, and amplified the nearly full-length sequences of these isolates by overlapping PCR. Among the other complete ChiVMV sequences that have been made available, Guangxi isolate was least similar (79% nt identity) to the Yunnan tobacco isolate suggesting that the host imposes a selective effect on the evolution of the virus (**Table 1**). However, since the Yunnan tobacco isolate was predicted to be a minor parent of the Yunnan pepper isolate identified in our work (**Table 2**), the similarity between these two was slightly higher. The Yunnan pepper isolate has the lowest nt similarity with an isolate from India, with a value of 80.6% (**Table 1**). Published studies of the evolution and variation of ChiVMV isolates have usually been based on CP sequences (Moury et al., 2005; Tsai et al., 2008; Yang et al., 2013; Gao et al., 2016; Ahmad and Ashfaq, 2018). By constructing a phylogenetic tree from the full-length sequences, we found that the viral isolates from the same or similar regions tend to group together (**Figure 1**). The analysis suggests that the evolutionary adaptation of the virus is driven by geographic location, which is consistent with the conclusion of Gao et al. (2016). However, in addition to Chinese isolates, we found that there were isolates from South Korea and India that were in the same clade as the Guangxi isolate (**Figure 1**). This may be due to the frequent trade of vegetables and ornamentals between China and these countries.

Recombination is considered to be a significant source of plant virus genetic diversity (Simon-Loriere and Holmes, 2011). Recombination events have been identified in several potyviruses, including soybean mosaic virus, potato virus Y etc., (Revers et al., 1996; Gagarinova et al., 2008; Seo et al., 2009). Our analysis detected a total of 10 recombination events within the full-length ChiVMV sequences, and the newly identified Guangxi and Yunnan isolates were both recognized as recombinants (**Table 2**), indicating that recombination plays an important role in shaping the adaptability of the ChiVMV population. We also detected three recombination events among 25 selected CP sequences from different geographical locations, and the Guangxi isolate was predicted to be a minor parent of the Sichuan isolate (KF738253.1) (**Supplementary Table 5**).

The site-specific selection pressure analysis of the CP coding region using a variety of methods shows that codons at many sites are subject to negative selection, which is consistent with previous

findings (Ahmad and Ashfaq, 2018), suggesting a strong negative or purifying selection in the ChiVMV population.

In conclusion, our study has determined the nearly complete genome sequences of two ChiVMV isolates from Guangxi and Yunnan provinces, China. Comparisons with other sequences have shown that the genetic differences among ChiVMV isolates were likely correlated with geographical location. Recombination occurs actively in the ChiVMV population and may be a force driving the adaptive evolution of the virus. A comparative analysis of the genome sequences of additional ChiVMV isolates would be helpful to give a clearer picture of genetic variability and evolution in this important virus.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SR and FY: conceptualization. JC and FY: funding acquisition. SR, XC, and SQ: investigation. SR, JP, HZ, YL, and GW: methodology. WJ and YZ: resources. FY: supervision. SR: writing – original draft. FY: writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the Chinese Agriculture Research System (CARS-24-C-04) and K. C. Wong Magna Fund in Ningbo University.

ACKNOWLEDGMENTS

We thank M. J. Adams, Minehead, United Kingdom, for correcting the English of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.616171/full#supplementary-material

Supplementary Table 1 | Primers used in this study.

Supplementary Table 2 | Detailed information of the isolates with whole genome sequences used for analysis in this study.

Supplementary Table 3 | Detailed information of the isolates with CP sequences used for analysis in this study.

Supplementary Table 4 | Nucleotide sequence similarity and divergence for the 16 ChiVMV isolates based on their full genome sequences.

Supplementary Table 5 | Summary of possible recombination events among 25 ChiVMV CP coding region sequences identified by RDP4.

REFERENCES

Ahmad, A., and Ashfaq, M. (2018). Genetic diversity and recombination analysis based on capsid protein gene of *Chilli veinal mottle virus* isolates from Pakistan. *Eur. J. Plant Pathol.* 151, 891–900. doi: 10.1007/s10658-018-1423-x

Ball, L. A. (2005). "The universal taxonomy of viruses in theory and practice," in *Virus Taxonomy*, eds C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball (San Diego: Academic Press), 3–8. doi: 10.1016/b978-0-12-249951-7.50003-1

Ding, M., Yang, C., Zhang, L., Jiang, Z. L., Fang, Q., Qin, X. Y., et al. (2011). Occurrence of *Chilli veinal mottle virus* in *Nicotiana tabacum* in Yunnan, China. *Plant Dis.* 95, 357–357. doi: 10.1094/pdis-09-10-0686

Gagarinova, A. G., Babu, M., Stromvik, M. V., and Wang, A. (2008). Recombination analysis of *Soybean mosaic virus* sequences reveals evidence of RNA recombination between distinct pathotypes. *Virol. J.* 5:143. doi: 10.1186/1743-422x-5-143

Gao, F., Jin, J., Zou, W., Liao, F., and Shen, J. (2016). Geographically driven adaptation of *Chilli veinal mottle virus* revealed by genetic diversity analysis of the coat protein gene. *Arch. Virol.* 161, 1329–1333. doi: 10.1007/s00705-016-2761-7

Kaur, C., Kumar, S., and Krishna Raj, S. (2015). Association of a distinct strain of *Chilli veinal mottle virus* with mottling and distortion disease of *Datura inoxia* in India. *Arch. Phytopathol. Pflanzenschutz* 48, 545–554. doi: 10.1080/03235408.2015.1052260

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003.

Moury, B., Palloix, A., Caranta, C., Gognalons, P., Souche, S., Selassie, K. G., et al. (2005). Serological, molecular, and pathotype dversity of *Pepper veinal mottle virus* and *Chili veinal mottle virus*. *Phytopathology* 95, 227–232.

Ong, C. A., Varghese, G., and Ting, W. P. (1980). The effect of *Chilli veinal mottle virus* on yield of chilli (*Capsicum annuum* L). *MARDI Res. Bull.* 8, 74–78.

Pond, S. L. K., and Frost, S. D. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533. doi: 10.1093/bioinformatics/bti320

Revers, F., and García, J. A. (2015). Chapter three - molecular biology of potyviruses. *Adv. Virus Res.* 92, 101–199. doi: 10.1016/bs.aivir.2014.11.006

Revers, F., Le Gall, O., Candresse, T., Le Romancer, M., and Dunez, J. (1996). Frequent occurrence of recombinant potyvirus isolates. *J. Gen. Virol.* 77(Pt 8), 1953–1965. doi: 10.1099/0022-1317-77-8-1953

Seo, J. K., Ohshima, K., Lee, H. G., Son, M., Choi, H. S., Lee, S. H., et al. (2009). Molecular variability and genetic structure of the population of *Soybean mosaic virus* based on the analysis of complete genome sequences. *Virology* 393, 91–103. doi: 10.1016/j.virol.2009.07.007

Shah, H., Yasmin, T., Fahim, M., Hameed, S., and Haque, M. I. (2008). Transmission and host range studies of Pakistani isolate of *Chilli veinal mottle virus*. *Pak. J. Bot.* 40, 2669–2681.

Shah, H., Yasmin, T., Fahim, M., Hameed, S., and Haque, M. I. (2009). Prevalence, occurrence and distribution of *Chili veinal mottle virus* in Pakistan. *Pak. J. Bot.* 41, 955–965.

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

Tsai, W. S., Huang, Y. C., Zhang, D. Y., Reddy, K., Hidayat, S. H., Srithongchai, W., et al. (2008). Molecular characterization of the CP gene and 3' UTR of *Chilli*

veinal mottle virus from South and Southeast Asia. *Plant Pathol.* 57, 408–416. doi: 10.1111/j.1365-3059.2007.01780.x

Wang, J., Liu, Z., Niu, S., Peng, M., Wang, D., Weng, Z., et al. (2006). Natural occurrence of *Chilli veinal mottle virus* on *Capsicum chinense* in China. *Plant Dis.* 90:377. doi: 10.1094/pd-90-0377c

Yang, J., Dong, J., Zhang, T. J., Wang, R., Luo, Z. P., Luo, H. Y., et al. (2013). A new isolate of *Chilli veinal mottle virus* that infects tobacco in China. *J. Plant Pathol.* 95, 187–190.

Zhao, F. F., Xi, D. H., Liu, J., Deng, X. G., and Lin, H. H. (2014). First report of *Chilli veinal mottle virus* infecting tomato (*Solanum lycopersicum*) in China. *Plant Dis.* 98:1589. doi: 10.1094/pdis-11-13-1188-pdn

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 February 2021
doi: 10.3389/fmicb.2021.623875

A Multiyear Survey and Identification of Pepper- and Tomato-Infecting Viruses in Yunnan Province, China

Yueyue Li[1,2†], Guanlin Tan[3†], Long Xiao[1], Wenpeng Zhou[1], Pingxiu Lan[1], Xiaojiao Chen[1], Yong Liu[4], Ruhui Li[5]* and Fan Li[1]**

[1] State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, China, [2] College of Life Science, Luoyang Normal University, Luoyang, China, [3] Modern Education Technology Center, Yunnan Agricultural University, Kunming, China, [4] Hunan Plant Protection Institute, Hunan Academy of Agricultural Sciences, Changsha, China, [5] USDA-ARS, National Germplasm Resources Laboratory, Beltsville, MD, United States

OPEN ACCESS

Edited by:
Rajarshi Kumar Gaur,
Deen Dayal Upadhyay Gorakhpur University, India

Reviewed by:
Jose Trinidad Ascencio-Ibáñez, North Carolina State University, United States
Hanu R. Pappu, Washington State University, United States

***Correspondence:**
Yong Liu
haoasliu@163.com
Ruhui Li
Ruhui.Li@ars.usda.gov
Fan Li
fanlikm@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: 30 October 2020
Accepted: 25 January 2021
Published: 26 February 2021

Citation:
Li Y, Tan G, Xiao L, Zhou W, Lan P, Chen X, Liu Y, Li R and Li F (2021) A Multiyear Survey and Identification of Pepper- and Tomato-Infecting Viruses in Yunnan Province, China. Front. Microbiol. 12:623875. doi: 10.3389/fmicb.2021.623875

During pepper and tomato production seasons in 2013–2017, large-scale virus disease surveys were conducted in different regions of Yunnan Province, China. A total of 1,267 pepper and tomato samples with various virus-like symptoms were collected and analyzed for virus infections through dot enzyme-linked immunosorbent assay (dot-ELISA), polymerase chain reaction (PCR), and reverse-transcription (RT)-PCR. The detection results showed that 19 different viruses were present in about 50.9% of the assayed samples, and among these viruses, seven viruses were found in both pepper and tomato samples. Mixed infections with two to three of the 15 identified mixed infection types were found in the pepper samples and 10 identified mixed infection types were found in the tomato samples. Among the infected samples, *Tomato spotted wilt orthotospovirus* (TSWV) was the most common virus, with a detection rate of about 20.0% followed by *Pepper vein yellows virus* (PeVYV, 13.0%). This survey revealed for the first time that pepper is a natural host of *Tobacco vein distorting virus* (TVDV) worldwide and tomato is a natural host of *Potato leafroll virus* (PLRV) in China. PeVYV, *Tobacco mild green mosaic virus* (TMGMV) and *Wild tomato mosaic virus* (WTMV) were first time found in pepper and *Tomato mottle mosaic virus* (ToMMV) and *Chilli veinal mottle virus* (ChiVMV) were first time found in tomato in Yunnan Province. Finally, the virus incidences were higher in Kunming, Yuxi, Chuxiong, and Honghe region than other regions.

Keywords: pepper, tomato, virus disease, virus identification, mixed infection

INTRODUCTION

Pepper (*Capsicum annuum*) and tomato (*Solanum lycopersicum*) are two economically important solanaceous vegetable crops in China and many other countries. With the rapid increase of production areas and growth of new cultivars in the recent years, outbreaks of virus disease have become common in solanaceous crops, leading to serious

yield losses. Approximately 76 and 136 viruses have been reported for pepper and tomato plants throughout the world, respectively (Dong et al., 2008; Hanssen et al., 2010; Kenyon et al., 2014; Li et al., 2014; Feng et al., 2017; Mauricio-Castillo et al., 2017; Olaya et al., 2017; Wang et al., 2017a; Liu et al., 2019). To date, 28 RNA and three DNA viruses have been reported for pepper crops, and 22 RNA and 12 DNA viruses have been reported for tomato crops in China. Many studies have suggested that *Cucumber mosaic virus* (CMV) and *Tobacco mosaic virus* (TMV) were the most predominant viruses in both pepper and tomato fields in China (Dong et al., 2008; Tao and Zhou, 2008; Zhang et al., 2009, 2017a; Qing et al., 2010; Wang and Huang, 2010; Zhang, 2010; Ge et al., 2012; Ruan, 2012; Ding et al., 2013; Li et al., 2014; Wang et al., 2014, 2017b; Zhao et al., 2014; Guo et al., 2015; Padmanabhan et al., 2015; Yuan et al., 2015; Feng et al., 2017; Liu et al., 2019).

Yunnan Province is located at the southwest border of China, neighboring Laos, Myanmar, and Vietnam. The low-latitude and high-elevation geographic features create a multiclimate weather, suitable for many vegetable productions, including many solanaceous vegetable crops. The total vegetable production acreage in Yunnan Province in 2017 had reached about 1.1 million ha and produced approximately 20.78 million tons of vegetables products, valued about 7.15 billion United States dollars. Pepper and tomato are the two major solanaceous vegetable crops in Yunnan Province. Due to virus diseases, some pepper and tomato fields in this province had total yield losses in the recent years, resulting in a quick increase of pesticide applications, insect tolerance to pesticides, and environment pollutions (Chen and Liu, 2013; Bu et al., 2014).

Identification of viruses and their epidemiology in pepper and tomato fields are critical for the development of effective managements for the viruses. Although field surveys of virus diseases in pepper and tomato fields had been reported for other provinces of China (Wen et al., 2009; Wang et al., 2014, 2017b; Guo et al., 2015; Liu, 2015; Wang L. S. et al., 2015; Gao et al., 2016; Zhang et al., 2017b), similar survey has not been done in Yunnan Province. Through this study, we have found that virus infections in these two crops are more common in Yunnan Province than that in other Chinese provinces.

MATERIALS AND METHODS

Field Survey and Sample Collection

Leaves and fruits showing virus-like symptoms were first recorded using a Sony camera (DSC-RX100M5, Sony), harvested, and then stored separately in plastic sampling bags till use. This survey focused on pepper and tomato crops in 38 counties belonging to 12 different cities or autonomous regions from April to October in 2013, 2014, 2015, and 2016 and April in 2017. More tissue samples were collected from the surveyed fields with higher disease incidences and severer disease symptoms, while fewer samples were collected from the fields with lower disease incidences and milder disease symptoms (**Table 1**). The collected samples were analyzed for virus infections immediately or stored at −80°C or −20°C until use.

TABLE 1 | Numbers of pepper and tomato samples collected from different regions of Yunnan Province.

Sampling sites		Number of samples	
City/ Autonomous region	**County**	**Pepper**	**Tomato**
Chuxiong	Yuanmou, Shuangbai, Yaoan, Nanhua, Lufeng, Chuxiong	116	190
Yuxi	Hongta, Jiangchuan, Tonghai, Chengjiang, Huaning, Xinping	135	22
Honghe	Shiping, Jianshui, Kaiyuan, Mengzi	212	60
Kunming	Panlong, Dongchuan, Yiliang, Xundian, Songming, Jinning	140	97
Wenshan	Wenshan, Guangnan, Yanshan, Qiubei	89	25
Dehong	Mangshi	11	21
Dali	Dali	28	0
Baoshan	Shidian, Longling	51	6
Lijiang	Gucheng	10	22
Zhaotong	Zhaoyang	13	0
Qujing	Shizong, Luliang, Huize, Xuanwei	13	3
Puer	Ninger, Jiangcheng	3	0
Total number		821	446

Serological Tests

Serological tests were done using the dot enzyme-linked immunosorbent assay (dot-ELISA) as described (Xie et al., 2013) with slight modifications. Briefly, each sample (0.1 g) was placed in a precooled 2-ml tube with five 3 mm and one 6 mm stainless-steel beads. After addition of 1 ml 0.01 M phosphate-buffered saline (PBS, 0.13 M NaCl, 0.003 M KCl, 0.008 M Na_2HPO_4, 0.001 M KH_2PO_4, pH 7.4), the tube was placed at −80°C for 5 min, and then ground using a high-throughput tissue homogenizer (Scientz-48, Ningbo Scientz) set at 60 Hz for 2 × 90 s (twice). The tube was centrifuged at 8,000 × *g* and 4°C for 5 min, and the supernatant in each tube was blotted onto a nitrocellulose membrane (2 μl per dot, three dots per sample) followed by air-drying. The membrane was incubated for 30 min in 10 ml PBST buffer (0.01 M PBS containing 0.05% Tween-20, pH 7.4) supplemented with 5% skim milk powder and then probed with a diluted monoclonal antibody for 1 h at 37°C. Monoclonal antibodies specific for *Broad bean wilt virus 2* (BBWV2), *Cucumber green mottle mosaic virus* (CGMMV), CMV, TMV, *Tomato spotted wilt orthotospovirus* (TSWV), or *Turnip mosaic virus* (TuMV) were kindly provided by Prof. Xueping Zhou, Zhejiang University, Hangzhou, China. After four washes in the PBST solution, the membrane was incubated again in 10 ml 10,000 times diluted alkaline phosphatase (AP)-conjugated goat anti-mouse IgG (Sigma-Aldrich, St. Louis, MO, USA) in PBST for 1 h at 37°C. After five washes in PBST, the membrane was incubated in 5 ml AP substrate solution [0.1 M Tris, pH 9.5, 0.1 M NaCl and 0.025 M $MgCl_2$, supplemented with nitro-blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl phosphate (Sangon Biotech., Shanghai, China)] as instructed for color development at room temperature for

10–25 min. The reaction was terminated after the positive control dots developed a purple color, whereas the negative control dots remained a light green color. A sample was considered infected if all three dots developed the purple color. Positive control samples representing BBWV2, CMV, TMV, TSWV, or TuMV were maintained in our laboratory, and the CGMMV-infected control sample was kindly provided by Dr. Zhaobang Cheng, Jiangsu Academy of Agricultural Science, Nanjing, China.

Total Nucleic Acid Extraction, Polymerase Chain Reaction and Reverse Transcription-PCR

Total nucleic acid was extracted from tissue samples using a modified CTAB method (Li et al., 2008a). Degenerated and specific primers for polymerase chain reaction (PCR) or reverse transcription (RT)-PCR were as reported (Ha et al., 2006, 2008; Hirota et al., 2010; Zhou et al., 2011; Zhang et al., 2017b; Li et al., 2018a,b,c, 2020; Liu et al., 2019) or made in this study (**Supplementary Table 1**). First-strand cDNA was synthesized using the Reverse Transcriptase M-MLV (RNase H^-) Kit (TaKaRa Biotech, Dalian, China). Each RT reaction contained about 200 ng of total RNA (2 μl), 1 μl of Random Oligo-dT (N6) primer (TaKaRa Biotech., Dalian, China) or a specific reverse primer (10 μM), and 3 μl nuclease-free ddH_2O. The mixture was incubated at 70°C for 10 min and then chilled immediately on ice for 3 min. After addition of 2 μl of 5 × M-MLV buffer, 0.5 μl of dNTP mixture (10 mM each), 0.5 μl of RTase M-MLV (RNase H^-), and 1 μl of nuclease-free H_2O, the reaction mixture was incubated at 42°C for 1 h followed by 70°C for 15 min. Each PCR reaction was 10 μl (5 μl Green Taq Mix (Vazyme Biotech. Co., Ltd., China), 0.2 μl each primer (10 μM each), 1 μl cDNA (RT-PCR) or total nucleic acid (PCR), and 3.6 μl nuclease-free H_2O), and the reaction condition was initial denaturation at 94°C for 5 min, 35 cycles (for reactions using specific primers) or 40 cycles (for reactions using degenerated primers) of 94°C for 30 s, 40–55°C for 30 s, and 72°C for 1 min, followed by the final extension at 72°C for 7 min. The PCR or RT-PCR products were visualized in 1% agarose gels through electrophoresis in 0.5 × TBE buffer. The resulting agarose gels were stained with 5 μl Gold View as instructed (Solarbio Science & Technology Co., Ltd., Beijing, China) and examined under an UV illumination GenoSens1850 system.

Cloning and DNA Sequencing

PCR products were recovered using a Universal DNA Purification Kit (TIANGEN Biotech., Beijing, China) and then cloned into the pMD19-T Vector (TaKaRa Biotech., Dalian, China). At least three clones representing a single PCR product were sequenced by the Beijing Genomics Institute (BGI, Shenzhen, China). The resulting sequences were used to BLAST search the sequence database at the National Center for Biotechnology Information[1]. The results were then confirmed by both dot-ELISA and RT-PCR for RNA viruses or PCR for DNA viruses.

[1]https://blast.ncbi.nlm.nih.gov/Blast.cgi

RESULTS

Field Survey Result

Pepper and tomato fields in 38 counties, belonging to 12 different cities or autonomous regions (refer to as regions thereafter) were surveyed for virus diseases in 2013, 2014, 2015, 2016, and 2017 (**Table 1** and **Figure 1A**). A total number of 821 pepper samples and 446 tomato samples were collected during the surveys. The most common symptoms on the infected pepper and tomato plants were mottling or mosaic, necrosis, yellowing, and distortion in leaves. The survey results showed that most of the surveyed pepper and tomato fields in Chuxiong, Yuxi, Honghe, and Kunming regions had a near 100% virus disease incidence, resulting in complete production failures (**Figures 1B,C**). Most surveyed fields in Wenshan, Dehong, Dali, and Baoshan regions had 10–25% virus disease incidences, while diseased plants in most surveyed fields in the Lijiang, Zhaotong, Qujing, and Puer regions were sporadic.

Virus Diversity

A total number of 19 different viruses were detected in 645 of the total 1,267 samples (50.9%) (**Table 2**). Among these viruses, *Chilli veinal mottle virus* (ChiVMV), CMV, *Pepper mild mottle virus* (PMMoV), *Tobacco bushy top virus* (TBTV), *Tomato mosaic virus* (ToMV), *Tomato mottle mosaic virus* (ToMMV), and TSWV were detected in both pepper and tomato samples (**Figure 2A**). The virus detection result also showed that TSWV was the most common virus (253 out of 1,267 assayed samples) in Yunnan Province, followed by *Pepper vein yellows virus* (PeVYV) and CMV (**Figure 2B** and **Table 2**). Further assays showed that virus population structures in the pepper and tomato fields in these regions varied each year (**Figure 2C**). For instance, unlike other viruses, TSWV, PeVYV, CMV, and PMMoV were detected in each year in the surveyed pepper and tomato fields, and the incidence of TSWV in 2013, 2014, 2015, and 2016 was at 42.1, 30.3, 59.2, and 4.7%, respectively (**Table 2** and **Figure 2C**).

Geographical Distributions of Viruses

To determine the geographical distributions of these viruses in Yunnan Province, we analyzed the field survey and virus detection data. The result showed that the virus detection rates in Kunming, Yuxi, Honghe, and Chuxiong regions were higher, ranging from 49.3 to 69.6% (**Figure 3A**), and also had much higher virus incidences in fields. Dehong and Dali regions also had high virus detection rates (62.5 and 50.0%, respectively), but the virus incidences in fields were not high; their high virus detection rates were due possibly to the less number of samples. For virus population structures, a total of 13 viruses were detected in the pepper and tomato samples collected from Chuxiong region, followed by 10 viruses in the samples from Yuxi region, eight viruses in the samples from Honghe region, six viruses in the samples from Kunming region, four viruses in the samples from Dehong region, and two viruses in the samples from Dali region (**Figure 3B**). Only one virus was detected in the samples from Baoshan, Wenshan, Lijiang, and Zhaotong regions. No virus was found in the samples from Qujing and Puer regions

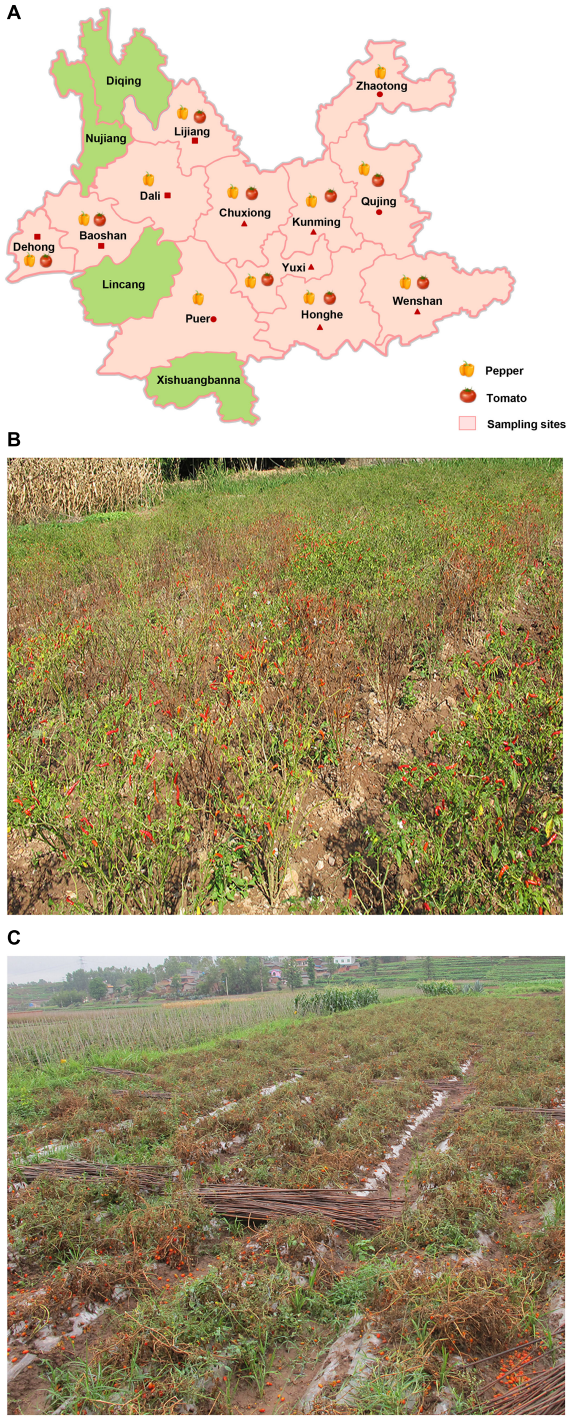


FIGURE 1 | Virus detection using field-collected pepper and tomato samples in Yunnan Province. **(A)** Distribution of sampling sites in this study is shown. The sampling sites are shown in pink, and pepper and tomato images indicate the samples collected in the indicated region. The sampling regions with a triangle symbol are the regions with more samples collected and with more fields having higher virus disease incidences. The sampling regions with a square symbol are the regions with less than 100 samples collected and with fewer fields having common virus infections. The sampling regions with a circular symbol are the regions with only a few samples collected and the *(Continued)*

FIGURE 1 | Continued
infected plants were only sporadic. Representative pepper **(B)** and tomato **(C)** fields with near 100% TSWV infection. These photos were taken in Honghe Autonomous region in 2013 **(B)** and Chuxiong Autonomous region in 2014 **(C)**.

TABLE 2 | Virus detection rates on pepper and tomato in different years in Yunnan.

Virus	Virus abbreviation	Detection rate in different years (%)				
		2013	2014	2015	2016	2017
Tomato spotted wilt orthotospovirus	TSWV	42.1	30.3	59.2	4.7	0.0
Pepper vein yellows virus	PeVYV	10.4	5.4	4.2	19.8	0.0
Cucumber mosaic virus	CMV	17.2	5.1	4.2	1.0	0.0
Tomato yellow leaf curl virus	TYLCV	1.8	0.0	0.0	2.3	47.5
Tomato mosaic virus	ToMV	0.0	3.1	0.0	3.4	0.0
Chilli veinal mottle virus	ChiVMV	0.0	5.1	0.0	1.9	0.0
Tomato chlorosis virus	ToCV	0.0	0.0	0.0	1.8	23.7
Tomato zonate spot orthotospovirus	TZSV	0.0	8.2	0.0	0.0	0.0
Tomato mottle mosaic virus	ToMMV	1.8	0.0	0.0	1.0	20.3
Tobacco bushy top virus	TBTV	5.4	1.0	0.0	0.0	0.0
Pepper mild mottle virus	PMMoV	0.5	1.7	2.8	1.0	0.0
Tobacco vein distorting virus	TVDV	2.7	0.7	7.0	0.0	0.0
Tomato yellow leaf curl Thailand virus	TYLCTHV	0.0	0.0	0.0	2.1	0.0
Papaya leaf curl China virus	PaLCuCNV	0.0	3.1	0.0	0.0	0.0
Tobacco mosaic virus	TMV	0.9	0.0	0.0	1.0	0.0
Tobacco mild green mosaic virus	TMGMV	0.0	2.7	0.0	0.0	0.0
Potato leaf roll virus	PLRV	0.0	0.0	0.0	0.5	0.0
Broad bean wilt virus 2	BBWV2	0.9	0.0	0.0	0.0	0.0
Wild tomato mosaic virus	WTMV	0.0	0.0	0.0	0.2	0.0

(**Figure 3A**), where lower virus incidence and small sample sizes may be responsible for the failure of virus detection.

Distributions of Pepper-Infecting Viruses

A total of 13 RNA viruses were detected in 429 of the 821 assayed pepper samples, namely: TSWV, PeVYV, CMV, ChiVMV, ToMV, *Tobacco vein distorting virus* (TVDV), PMMoV, ToMMV, TMV, *Tobacco mild green mosaic virus* (TMGMV), BBWV2, TBTV, and *Wild tomato mosaic virus* (WTMV) (**Figure 4A**). Among these

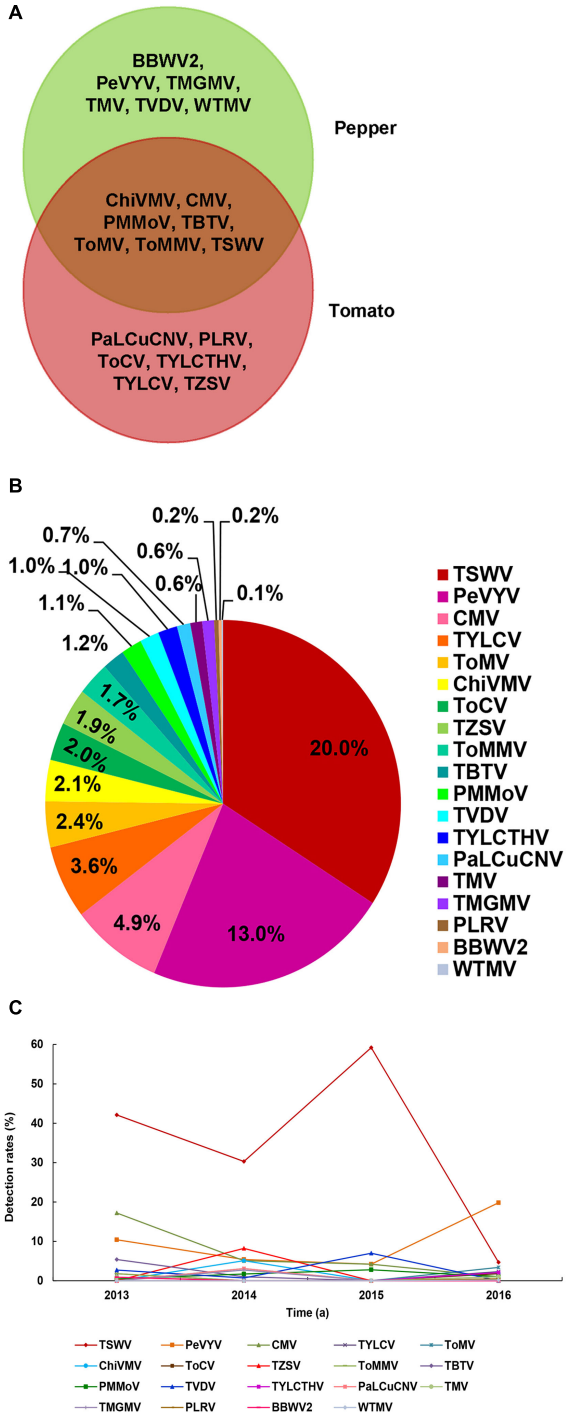


FIGURE 2 | Virus population structures in pepper and tomato fields in Yunnan Province. **(A)** Viruses identified in the pepper and tomato samples. The viruses identified in the pepper samples is shown in the green area, the viruses identified in tomato samples is shown in the red area, and the viruses identified in both pepper and tomato samples is shown in the overlapped area. **(B)** Detection rates of viruses in pepper and tomato samples. Abbreviation of each detected virus is represented with a defined color, and the detection rate of each virus is marked in the corresponding icon. **(C)** Occurrence of different viruses from 2013 to 2016. Abbreviation of each detected virus is represented by a defined color legend line.

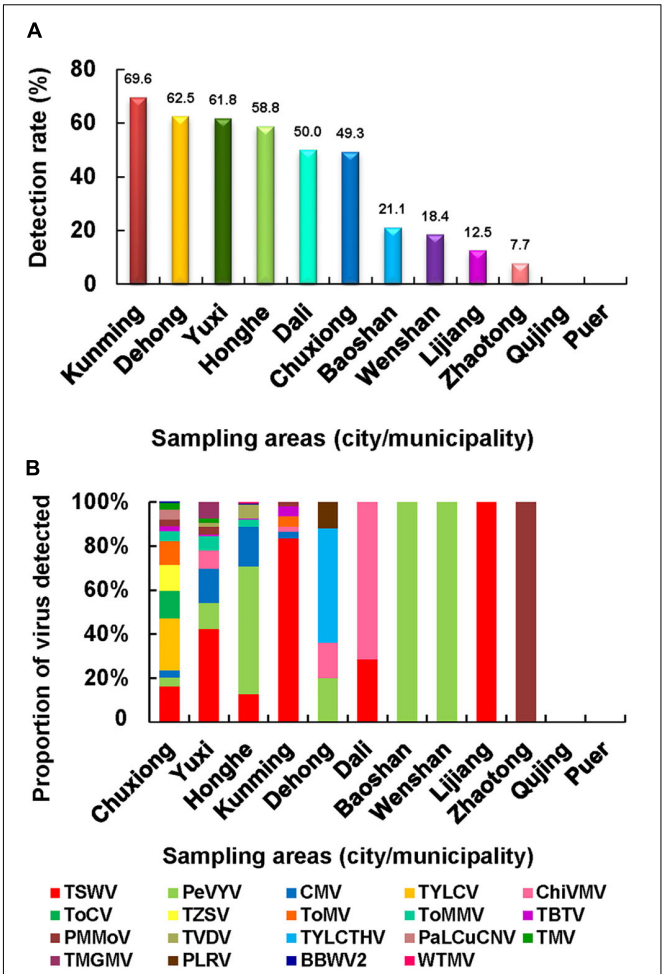


FIGURE 3 | Virus population structures in different surveyed regions. **(A)** Virus detection rates in the 12 surveyed regions. **(B)** Virus population structures in different regions surveyed in this study.

RNA viruses, TSWV and PeVYV were the most predominant viruses and were found in 20.5 and 20.1% of the assayed pepper samples, respectively (**Figure 4A**). In this study, pepper was first time found as a natural host of TVDV (GenBank accession no. MN889854). In addition, PeVYV (GenBank accession no. MN889851), TMGMV (GenBank accession no. MN889852), and WTMV (GenBank accession no. MN889853) were first time found in the pepper fields in Yunnan Province. Among the 49 pepper-infecting CMV isolates, 48 were CMV I subgroup isolates and one was CMV II subgroup isolate.

Of the 429 infected pepper samples, 51 samples were co-infected with two or three different viruses (i.e., CMV + TSWV, CMV + PeVYV, CMV + ToMMV, CMV + TVDV, CMV + WTMV, TSWV + PeVYV, TSWV + BBWV2, TMV + ToMV, ToMV + PMMoV, TMGMV + PMMoV, PeVYV + TVDV, CMV + ToMMV + PMMoV, CMV + TSWV + PeVYV, CMV + TVDV + PeVYV, and TMV + ToMV + PeVYV) (**Supplementary Table 2**). Among these mixed infections, CMV + TSWV and TSWV + PeVYV

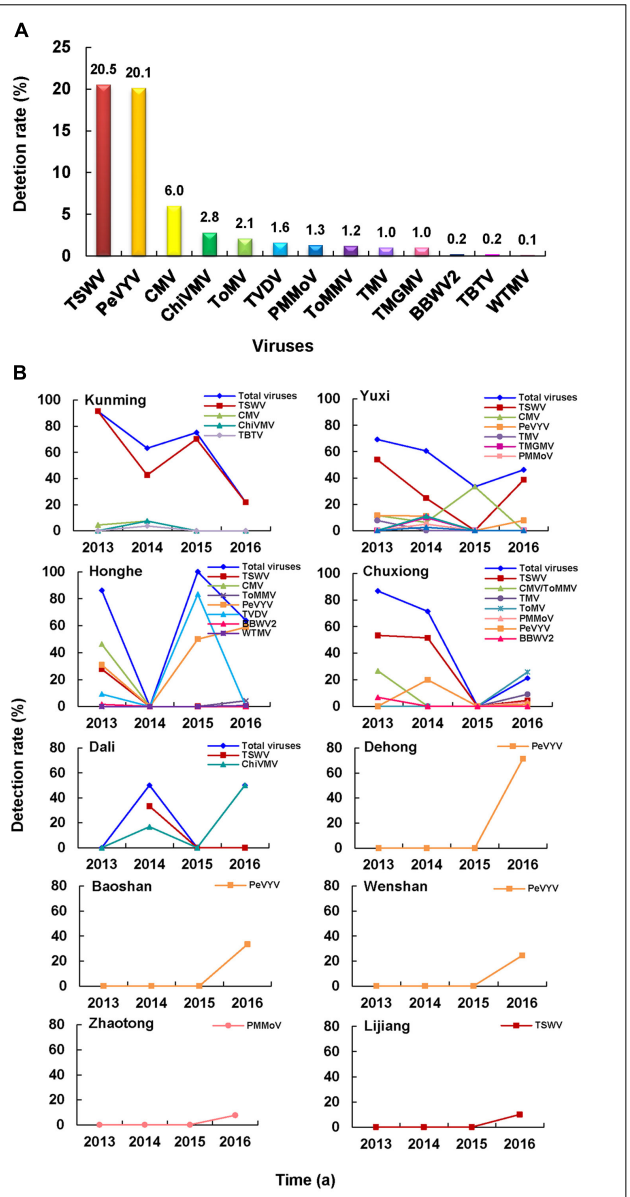


FIGURE 4 | Pepper virome in Yunnan Province. **(A)** Detection rates of viruses in the pepper samples. **(B)** Distribution and dynamics of pepper-infecting viruses from 2013 to 2016.

were most common combinations. The mixed infection of CMV + TSWV occurred mainly in Chuxiong, Honghe, Yuxi, and Kunming regions, while the mixed infection of TSWV + PeVYV occurred only in Honghe and Yuxi regions. Also, among these 15 co-infected types, eight had CMV and four contained TSWV (**Supplementary Table 2**).

TSWV was the most common virus in Kunming, Yuxi, and Chuxiong regions, while PeVYV was the most dominant virus in Honghe, Dehong, Baoshan, and Wenshan regions (**Figure 4B**). In Dali region, the most common virus was ChiVMV, while only PMMoV or TSWV was found in the pepper fields in Zhaotong and Lijiang regions (**Figure 4B**). The average detection rates of viruses in pepper samples collected in 2013, 2014, 2015, and 2016 were 77.7, 56.7, 75.5, and 38.8%, respectively. In general, the detection rate of TSWV in these years showed a downward trend, while the trend of PeVYV was in increase, especially in 2016 (**Figure 4B**).

Distributions of Tomato-Infection Viruses

A total of 10 RNA viruses and three DNA viruses were detected in 216 of the 446 assayed tomato samples (**Figure 5A**). The identified RNA viruses were TSWV, *Tomato chlorosis virus* (ToCV), *Tomato zonate spot orthotospovirus* (TZSV), CMV, ToMV, TBTV, ToMMV, ChiVMV, PMMoV, and *Potato leafroll virus* (PLRV), and the identified DNA viruses were *Tomato yellow leaf curl virus* (TYLCV), *Tomato yellow leaf curl Thailand virus* (TYLCTHV), and *Papaya leaf curl China virus* (PaLCuCNV). Among these viruses, TSWV and TYLCV were most common viruses found in the tomato samples. In this survey, PLRV (GenBank accession no. MN894826) was first time found in the infected tomato samples in China, while ToMMV (GenBank accession no. MN894824) and ChiVMV (GenBank accession no. MN894825) were first time found in the tomato samples in Yunnan Province. It is noteworthy that all the CMV isolates detected in the tomato samples belonged to the CMV II subgroup.

Among the 216 infected tomato samples, 38 were co-infected with two or three viruses (i.e., TSWV + ToMV, TSWV + TBTV, ToMV + PMMoV, TYLCV + ToCV, TYLCV + ToMMV, TYLCV + TBTV, TYLCTHV + ChiVMV, TYLCTHV + PLRV, TYLCV + ToCV + ToMMV, and TYLCTHV + ChiVMV + PLRV) (**Supplementary Table 3**). Among these mix-infected samples, the most samples contained begomoviruses. The TYLCV + ToCV mixed infection was mainly found in the tomato fields collected from Chuxiong region, followed by TSWV + TBTV and then TYLCV + ToMMV mixed infection (**Supplementary Table 3**).

The dominant tomato-infecting viruses in different regions varied. For instance, TSWV was the most common tomato-infecting virus in Kunming region from 2013 to 2016, whereas CMV and TYLCV were the most common viruses in Yuxi and Chuxiong regions, respectively (**Figure 5B**). In 2013 and 2014, virus diseases were more common in Kunming, Yuxi, and Chuxiong regions, Honghe region in 2013, and Dehong region in 2016 (**Figure 5B**).

Correlation Between Disease Symptoms and the Causal Viruses

In this study, the most common disease symptoms observed on the infected pepper and tomato plants were leaf mottling or mosaic, distortion, necrosis and yellowing, and plant stunting. To investigate the potential correlations between disease symptoms and the 19 identified viruses, we compared disease symptom and the virus identification data obtained in this study. The result indicated that the correlations were complex. Same virus often caused different disease symptoms in these two host plants of different varieties at different growth stages and in different infection seasons. In contrast, some different viruses

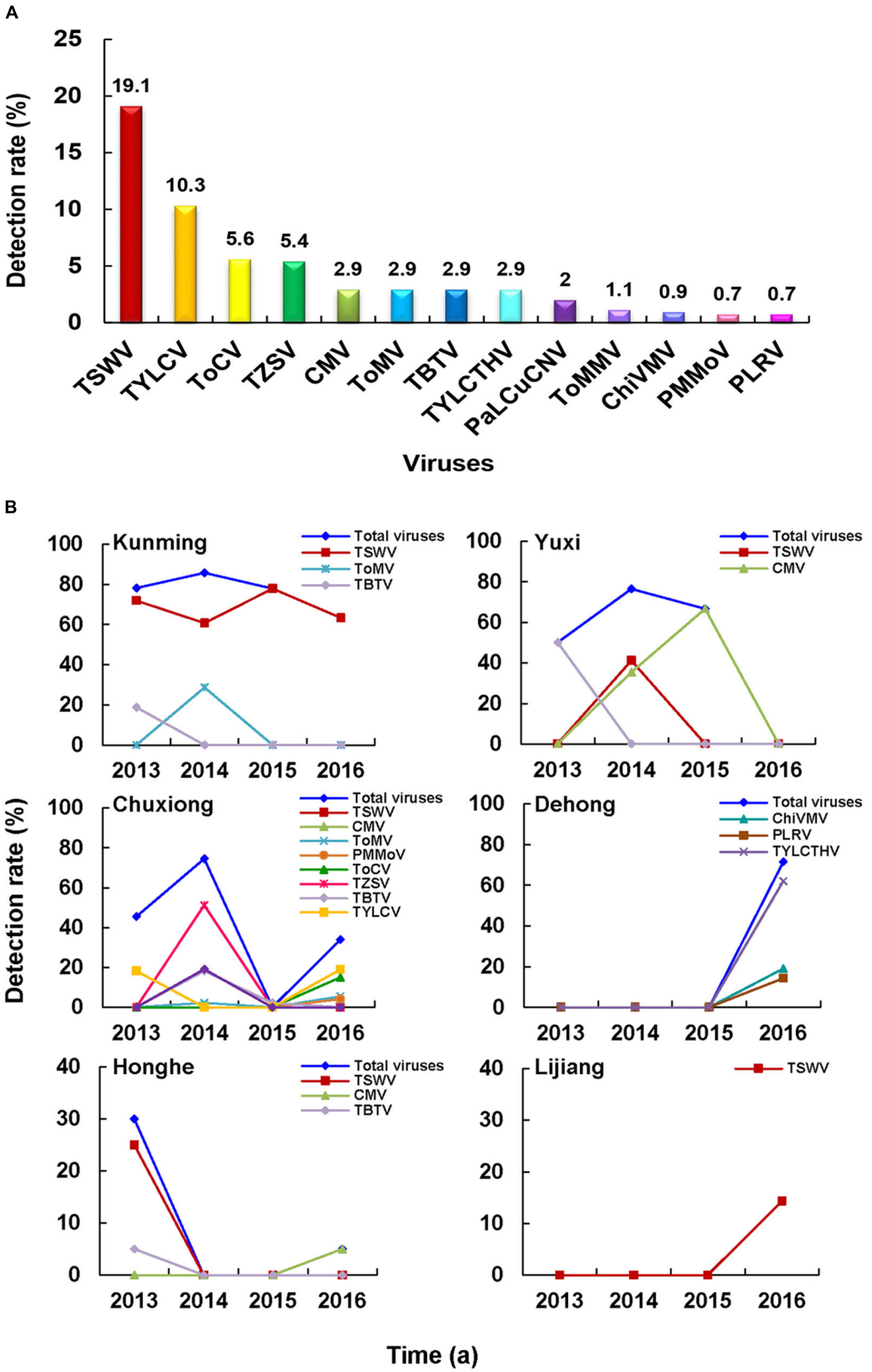


FIGURE 5 | Tomato virome in Yunnan Province. **(A)** Detection rates of viruses in the tomato samples. **(B)** Distribution and dynamics of tomato-infecting viruses from 2013 to 2016.

caused similar disease symptoms in different infected pepper or tomato plants of the same variety. Furthermore, disease symptoms caused by different combinations of viruses in the pepper and tomato plants also varied and were influenced by the environmental conditions. Therefore, disease symptoms shown on infected plants cannot be directly linked to the causal virus(es). However, disease symptoms caused by a single certain virus in a particular host plant were similar, including blistering and ring spots in the TSWV-infected leaves or leaf distortion (i.e., shoestring) in CMV-infected leaves. Disease symptoms caused by viruses belonging to the same genus (i.e., TMV, ToMV, ToMMV, PMMoV, and TMGMV) were similar (i.e., mottling and mosaic in leaves).

DISCUSSION

In this study, we surveyed pepper and tomato fields in different regions in Yunnan Province, China (**Figure 1A**). The result showed that the disease severities in different regions varied significantly. A total of 19 viruses were identified in assayed pepper and tomato samples collected in this study (**Figures 2A,B**). This number was higher than that previously reported for other regions in China (Wen et al., 2009; Wang et al., 2014, 2017b; Guo et al., 2015; Liu, 2015; Wang L. S. et al., 2015; Gao et al., 2016; Zhang et al., 2017b) and indicated that the virus population structures in the pepper and tomato fields in this region were more complex than that in the other regions of China. Our result showed that TSWV, PeVYV, CMV, and PMMoV occurred every year in the pepper and tomato fields in this province from 2013 to 2016 (**Figure 2C**). Among these four viruses, the occurrence of TSWV and CMV showed a decline trend, while the occurrence of PeVYV was gradually increasing. The occurrence of PMMoV maintained relatively stable from 2013 to 2016. Other viruses occurred sporadically in this period. During this survey, 13 viruses were detected in 2016 (**Figure 2C**), which can be useful for the design and development of pepper and tomato virus disease management strategy in this region.

The occurrence of TSWV in the pepper and tomato fields in 2013–2016 showed a downward trend (**Figures 2C**, **4B**, **5B**), probably due to the new control strategy designed for this virus in this area since 2014 (Zheng et al., 2015), which indicated that timely prevention and control was effective for virus disease management. The virus distributions in the pepper and tomato fields observed in the study (**Figures 4B**, **5B**) revealed that virus diseases were more common in Chuxiong, Yuxi, Kunming, and Honghe regions than other regions, and the virus population structures in different surveyed regions varied (**Figures 4B**, **5B**). For example, TSWV was the most common virus in Kunming and Yuxi regions, while PeVYV was a more common virus in Honghe, Baoshan, and Wenshan regions. Because different viruses are transmitted *via* different ways, management strategies for different regions should also vary accordingly. For example, applications of insect pesticides and resistant pepper and tomato varieties should be used in Kunming, Dehong, Honghe, Dali, and Baoshan regions because most dominant viruses in these regions are transmitted by insect vectors (**Figure 3B**). In contrast, a more integrated management measure should be used in Chuxiong region because this region has multiple viruses but none of them is prominent (**Figure 3B**).

Several recent studies have shown that CMV, tobamoviruses (i.e., TMV, ToMV, and/or PMMoV), and TYLCV were the major viruses in pepper and tomato fields in different provinces of China (Wen et al., 2009; Wu et al., 2013; Wang et al., 2014, 2017b; Guo et al., 2015; Liu, 2015; Wang L. S. et al., 2015; Gao et al., 2016; Liu et al., 2019). The results presented in this study showed that the virus population structures in the pepper and tomato fields in Yunnan Province changed in these years compared with an earlier report (Zhang et al., 1998). In that report, Zhang et al. (1998) had indicated that the main viruses in the pepper and tomato fields in this province were CMV and TMV. Due to the limitation of detection technology in that time, we cannot rule out the possibility of existence of other viruses in that study. TSWV is currently the most important virus in both pepper and tomato fields in Yunnan Province (**Figures 4A**, **5A**) and has severe pathogenicity (**Figures 1B,C**) and very wide host range (Li et al., 2015; Xiao et al., 2015). The infection rates of CMV, TMV, ToMV, PMMoV, and TYLCV in Yunnan Province have declined in recent years. This reduction may be caused by the fact that the corresponding control measures such as virus-resistant varieties were made in the management of these viruses based on the previous reports in the province (Zhang et al., 1998; Zhang, 2010). This report indicated that regular survey of the virus diseases was very important in management of viral diseases. Currently, CMV II subgroup isolates are more common in Yunnan Province than in other areas of China. Although both CMV I and II subgroup isolates are present in Yunnan Province, CMV I subgroup isolates are more common in the pepper fields, while CMV II subgroup isolates are more common in the tomato fields. The dominance of CMV I subgroup isolates in the pepper fields is consistent with the report published previously (Li et al., 2008b). The dominance of CMV II subgroup isolates in tomato fields has not been reported previously in other regions of China. It was reported that CMV I subgroup isolates were common in other vegetable crops in China (Chen et al., 2016) and CMV II subgroup isolates were common in the tobacco, lily, and tomato in Yunnan Province (Li et al., 2000; Zhao et al., 2007).

TBTV and TVDV were known to infect tobacco to cause bushy top disease (Wang D. Y. et al., 2015), it may be a challenge for the prevention and control of tobacco bush top disease, since these two viruses are newly identified in pepper and/or tomato fields in Yunnan Province (Li et al., 2018a). PeVYV is a new virus in the pepper fields in this province and is gradually becoming a dominant virus (**Figure 4A**). Its infection rate in pepper fields in 2016 has reached 19.8% (**Figure 2C**). The first report of ToMMV in the pepper fields in Yunnan Province was in 2014 (Li et al., 2014) and is now in the pepper, tomato, and eggplant in other parts of China (Chai et al., 2018; Li et al., 2020). Because ToMMV can cause severe diseases in many crops and its infection rate in vegetable crops is gradually increasing in China (Sui et al., 2017; Nagai et al., 2019; Li et al., 2020), it may become a devastating virus in many vegetable crops in China. We consider that more investigations are needed to identify new viruses so that specific virus resistance can be integrated in the pepper and tomato breeding programs.

In this survey, no virus was detected in some samples with virus-like symptoms. It is possible that these virus-like symptoms were induced by other factors, including fungi, insects, abiotic stresses, environmental conditions, and/or pesticides. It is also possible that the causal virus(es) were not covered by the detection methods. As the sample sizes from some locations were small, the disease incidence in these locations needs further validation. Nevertheless, the information presented here should allow a better management of virus diseases in Yunnan Province.

CONCLUSION

We have conducted a comprehensive survey of virus diseases in the pepper and tomato fields in Yunnan Province, one of the main vegetable production provinces in China. Our result indicates that the virus population structure in the pepper fields is more complex than that in the tomato fields. This information is useful for a better management of virus diseases in these two vegetable fields in this region.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

YLi, GT, and FL conceived and designed the experiments. YLi, LX, and WZ performed the experiments. YLi and GT collected the samples and analyzed the data. PL, XC, and YLiu contributed reagents, materials, and analysis tools. YLi, RL, and FL wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This project was financially supported by the National Natural Science Foundation of China (31660039 and 31701767), Special Fund for Agro-scientific Research in the Public Interest (201303028), Program for Innovative Research Team (in Science and Technology) in University of Yunnan province (Yunjiaoke 2014-22), and National Program Cultivation Fund of Luoyang Normal University (2019-PYJJ-009).

ACKNOWLEDGMENTS

The authors thank Dr. Xinshun Ding (retired, New York, USA) for his help on the language of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.623875/full#supplementary-material

Supplementary Table 1 | Primers used in this study to detect viruses in pepper and tomato samples.

Supplementary Table 2 | Numbers and detection rates of different viruses in the mix-infected pepper samples collected from different regions in this study.

Supplementary Table 3 | Numbers and detection rates of different viruses in the mix-infected tomato samples collected from different regions in this study.

REFERENCES

Bu, Y. Q., Kong, Y., Zhi, Y., Wang, J. Y., and Shan, Z. J. (2014). Pollution of chemical pesticideson environment and suggestion for prevention and control counter measures. *J. Agric. Sci. Technol.* 16, 19–25. doi: 10.13304/j.nykjdb.2014.129

Chai, A. L., Chen, L. D., Li, B. J., Xie, X. W., and Shi, Y. X. (2018). First report of a mixed infection of Tomato mottle mosaic virus and Tobacco mild green mosaic virus on eggplants in China. *Plant Dis.* 102:2668. doi: 10.1094/PDIS-04-18-0686-PDN

Chen, L., and Liu, L. J. (2013). Chemical control of plant virus diseases. *Agrochemicals* 52, 787–789. doi: 10.16820/j.cnki.1006-0413.2013.11.002

Chen, Y. Z., Tan, X. Q., Zhu, C. H., Sun, S., Liu, Y., and Zhang, D. Y. (2016). Sequence diversity analysis of cucumber mosaic virus isolates from common crops in China. *J. Plant Prot.* 43, 427–433.

Ding, Y. N., Li, S. F., Zhang, S., Cui, Y. Y., Yang, H., He, W., et al. (2013). Molecular detection and analysis of the complete nucleotide sequences of TYLCV isolates from Xinjiang. *Plant Prot.* 39, 51–55.

Dong, J. H., Cheng, X. F., Yin, Y. Y., Fang, Q., Ding, M., Li, T. T., et al. (2008). Characterization of *Tomato zonate* spot virus, a new tospovirus in China. *Arch. Virol.* 153, 855–864. doi: 10.1007/s00705-008-0054-5

Feng, G., Xin, M., Cao, M. J., Wang, L. S., Li, L., and Wang, X. F. (2017). Identification of multiple viruses infecting hot pepper in Guiyang by deep sequencing. *Acta Phytopathol. Sin.* 47, 591–597. doi: 10.13926/j.cnki.apps.000113

Gao, W., Wang, Y., Zhang, C. X., Zhang, A. S., and Zhu, X. P. (2016). Investigation and pathogen preliminary identification of pepper virus disease in Tianjin. *Shandong Agric. Sci.* 48, 91–94.

Ge, B. B., He, Z., Jiang, D. M., Zhang, Z. X., Liu, G. J., and Wang, H. Q. (2012). Characterization and complete nucleotide sequence of potato virus M isolated from tomato in China. *Acta Virol.* 56, 261–263.

Guo, S. Y., Tong, Y., Huang, Y., Luo, X. F., and Qing, L. (2015). Preliminary identification and analyses of viruses causing pepper virus disease in Chongqing. *China. Acta Hortic. Sin.* 42, 263–270. doi: 10.16420/j.issn.0513-353x.2014-0774

Ha, C., Coombs, S., Revill, P., Harding, R., Vu, M., and Dale, M. (2006). Corchorus yellow vein virus, a new world geminivirus from the old world. *J. Gen. Virol.* 87, 997–1003. doi: 10.1099/vir.0.81631-0

Ha, C., Coombs, S., Revill, P. A., Harding, R. M., Vu, M., and Dale, J. L. (2008). Design and application of two novel degenerate primer pairs for detection and complete genomic characterization of potyviruses. *Arch. Virol.* 153, 25–36. doi: 10.1007/s00705-007-1053-7

Hanssen, I. M., Lapidot, M., and Thomma, B. P. H. J. (2010). Emerging viral diseases of tomato crops. *Mol. Plant Microbe* 23, 539–548. doi: 10.1094/MPMI-23-5-0539

Hirota, T., Natsuaki, T., Murai, T., Nishigawa, H., Niibori, K., Goto, K., et al. (2010). Yellowing disease of tomato caused by *Tomato chlorosis* virus newly recognized in Japan. *J. Gen. Plant Pathol.* 76, 168–171. doi: 10.1007/s10327-010-0219-4

Kenyon, L., Kumar, S., Tsai, W. S., Hughes, J., and d'A. (2014). Virus diseases of peppers (*capsicum spp.*) and their control. *Adv. Virus Res.* 90, 297–354. doi: 10.1016/B978-0-12-801246-8.00006-8

Li, F., Zhou, X. P., Qi, Y. J., Chen, H. R., and Li, D. B. (2000). Detection of Cucumber mosaic virus subgroup II isolate from tobacco plants in Yunnan province. *Acta Microbiol. Sin.* 40, 346–351.

Li, R., Mock, R., Huang, Q., Abad, J., Hartung, J., and Kinard, G. (2008a). A reliable and inexpensive method of nucleic acid extraction for the PCR-based detection of diverse plant pathogens. *J. Virol. Methods* 154, 48–55. doi: 10.1016/j.jviromet.2008.09.008

Li, W., Kong, B. H., Chen, H. R., and Bai, W. (2008b). The identification of Cucumber mosaic virus subgroups of Capsicum in the main producing areas of Yunnan. *J. Yunnan Agric. Univ.* 23, 167–172. doi: 10.16211/j.issn.1004-390x(n).2008.02.019

Li, Y. Y., Liu, Q. L., Xiang, D., Li, Z. N., Ma, Y., Tan, S. T., et al. (2018a). First report of natural infection of *Tobacco bushy* top virus on tomato and pepper in China. *Plant Dis.* 102:1466. doi: 10.1094/PDIS-11-17-1820-PDN

Li, Y. Y., Ma, Y., Meng, Y., Huang, M. Z., Ren, G. M., Zhao, J. F., et al. (2018b). First report of *Chilli veinal* mottle virus infecting *Solanum aethiopicum* in China. *Plant Dis.* 102:1181. doi: 10.1094/PDIS-09-17-1351-PDN

Li, Y. Y., Tan, G. L., Lan, P. X., Zhang, A. S., Liu, Y., Li, R. H., et al. (2018c). Detection of tobamoviruses by RT-PCR using a novel pair of degenerate primers. *J. Virol. Methods.* 259, 122–128. doi: 10.1016/j.jviromet.2018.06.012

Li, Y. Y., Wang, C. L., Xiang, D., Li, R. H., Liu, Y., and Li, F. (2014). First report of *Tomato mottle* mosaic virus infection of pepper in China. *Plant Dis.* 98:1447. doi: 10.1094/PDIS-03-14-0317-PDN

Li, Y. Y., Xiao, L., Tan, G. L., Fu, X. P., Li, R. H., and Li, F. (2015). First report of Tomato spotted wilt virus on celery in china. *Plant Dis.* 99:734. doi: 10.1094/PDIS-11-14-1105-PDN

Li, Y. Y., Zhou, W. P., Lu, S. Q., Chen, D. R., Dai, J. H., Ge, Q. Y., et al. (2020). Occurrence and biological characteristics of Tomato mottle mosaic virus on solanaceae crops in China. *Sci. Agric. Sin.* 53, 539–550. doi: 10.3864/j.issn.0578-1752.2020.03.007

Liu, X. J. (2015). *Identification and Molecular Variation of Virus Infecting Vegetables in Zhejiang and Jiangxi provinces*. Ph. D. thesis, Chinese Academy of Agricultural Sciences: Hangzhou.PhDthesis

Liu, Y., Li, F., Li, Y. Y., Zhang, S. B., Gao, X. W., Xie, Y., et al. (2019). Identification, distribution, and occurrence of viruses in the main vegetables of China. *Sci. Agric. Sin.* 52, 239–261.

Mauricio-Castillo, J. A., Reveles-Torres, L. R., Mena-Covarrubias, J., Argüello-Astorga, G. R., Creamer, R., Franco-Bañuelos, A., et al. (2017). First report of Beet curly top virus-PeYD associated with a new disease in chile pepper plants in Zacatecas. *Mexico. Plant Dis.* 101:513. 10.1094/PDIS-09-16-1277-PDN

Nagai, A., Duarte, L. M. L., Chaves, A. L. R., Peres, L. E. P., and dos Santos, D. Y. A. C. (2019). Tomato mottle mosaic virus, in brazil and its relationship with Tm-22 gene. *Eur. J. Plant Pathol.* 155, 353–359. doi: 10.1007/s10658-019-01762-7

Olaya, C., Velásquez, N., Betancourt, M., Cuellar, W. J., and Pappu, H. R. (2017). First report of natural infection of *Alstroemeria necrotic* streak virus on tomato (*Solanum lycopersicum*) and bell pepper (*Capsicum annuum*) in Colombia. *Plant Dis.* 101:1065. doi: 10.1094/PDIS-10-16-1396-PDN

Padmanabhan, C., Zheng, Y., Li, R., Sun, S. E., Zhang, D. Y., Liu, Y., et al. (2015). Complete genome sequence of Southern tomato virus identified in China using next-generation sequencing. *Genome Announc* 3:e01226-e15. doi: 10.1128/genomeA.01226-15

Qing, L., Xiong, Y., Sun, X. C., Yang, S. Y., and Zhou, C. Y. (2010). First report of *Tobacco curly* shoot virus infecting pepper in China. *Plant Dis.* 94:637. doi: 10.1094/PDIS-94-5-0637A

Ruan, T. (2012). *Molecular Identification and Sequence Analysis of Geminiviruses on Three Host Plants in China*. Master's thesis, University of Southwest: Chongqing.Mastersthesis

Sui, X. L., Zheng, Y., Li, R. G., Padmanabhan, C., Tian, T. Y., Groth-Helms, D., et al. (2017). Molecular and biological characterization of Tomato mottle mosaic virus and development of RT-PCR detection. *Plant Dis.* 101, 704–711. doi: 10.1094/PDIS-10-16-1504-RE

Tao, X. R., and Zhou, X. P. (2008). Pathogenicity of a naturally occuring recombinant DNA satellite associated with Tomato yellow leaf curl China virus. *J. Gen. Virol.* 89, 306–311. doi: 10.1099/vir.0.83388-0

Wang, A. W., and Huang, P. Z. Q. (2010). Symptom, pathogeny and chemical control of tomato virus disease. *J. Changjiang Vegetables* 3, 38–40. doi: 10.3865/j.issn.1001-3547.2010.03.022

Wang, D. Y., Yu, C. M., Wang, G. L., Shi, K. R., Li, F., and Yuan, X. F. (2015). Phylogenetic and recombination analysis of Tobacco bushy top virus in china. *Virol. J.* 12:111. doi: 10.1186/s12985-015-0340-2

Wang, L. S., Chen, W., Tan, Q. Q., Chen, X. J., Wu, S. P., He, H. Y., et al. (2015). Detection of viruses on pepper in Guizhou. *Guizhou Agric. Sci.* 43, 99–101. doi: 10.3969/j.issn.1001-3601.2015.08.024

Wang, S. L., Sun, X. H., Tan, W. P., Wang, C. L., Yang, Y. Y., Wang, X. Y., et al. (2017a). First report of Melon aphid-borne yellows virus (MABYV) in pepper plants in China. *Plant Dis.* 101:262. doi: 10.1094/PDIS-03-16-0349-PDN

Wang, S. L., Tan, W. P., Yang, Y. Y., Dai, H. J., Sun, X. H., Qiao, N., et al. (2017b). Molecular detection and identification of main viruses on pepper in Shandong province. *Sci. Agric. Sin.* 50, 2728–2738. doi: 10.3864/j.issn.0578-1752.2017.14.009

Wang, Y., Gao, W., and Zhang, C. X. (2014). Occurrence and ELISA detection of tomato virus in Tianjin area. *Tianjin Agric. Sci.* 20, 1–5. doi: 10.3969/j.issn.1006-6500.2014.12.001

Wen, C. H., Liu, Y. L., Liu, Q., Wang, J. P., and Song, R. (2009). Identification of viruses infecting tomato in Hexi region of Gansu province. *Acta Agric. Boreali-occidentalis Sin.* 18, 291–294.

Wu, H. F., Cao, A. C., Zheng, J. Q., Guo, M. X., Wang, Q. X., Li, Y., et al. (2013). Research progress of tomato yellow leaf curl disease. *Crops* 1, 18–26.

Xiao, L., Li, Y. Y., Lan, P. X., Tan, G. L., Ding, M., Li, R. H., et al. (2015). First report of Tomato spotted wilt virus infecting cowpea in china. *Plant Dis.* 100:233. doi: 10.1094/PDIS-04-15-0495-PDN

Xie, Y., Jiao, X. Y., Zhou, X. P., Liu, H., Ni, Y. Q., and Wu, J. X. (2013). Highly sensitive serological methods for detecting tomato yellow leaf curl virus in tomato plants and whiteflies. *Virol. J.* 10:142. doi: 10.1186/1743-422X-10-142

Yuan, W., Yong, R. J., Zuo, L. Y., Du, K. T., and Zhou, T. (2015). First report of Beet western yellow virus on pepper in China. *J. Plant Pathol.* 97, 391–403. doi: 10.4454/JPP.V97I2.009

Zhang, H., Gong, H. R., and Zhou, X. P. (2009). Molecular characterization and pathogenicity of Tomato yellow leaf curl virus in China. *Virus Genes* 39, 249–255. doi: 10.1007/s11262-009-0384-8

Zhang, S., Zhang, D., Liu, Y., Luo, X. W., Liu, M. Y., Du, J., et al. (2017a). First report of *Lettuce chlorosis* virus infecting tomato in China. *Plant Dis.* 101:846. doi: 10.1094/PDIS-09-16-1315-PDN

Zhang, Y. C., Hou, M. S., and Cai, L. (2017b). Detection of main viral species of vegetable virus diseases in Hubei province. *J. Huazhong Agric. Univ.* 6, 31–38. doi: 10.13300/j.cnki.hnlkxb.2017.06.005

Zhang, Z. K. (2010). *Geographic Distribution of Geminiviruses in Yunnan and its Effect on the Disease Occurrence and Damage*. Ph. D. thesis, Chinese Academy of Agricultural Sciences: Beijing.

Zhang, Z. K., Fang, Q., Wu, Z. Q., He, Y. K., Li, Y. H., and Huang, X. Q. (1998). The occurrence and distribution of main *Solanacae crops* viruses in Yunnan. *J. Yunnan Univ.* 20, 128–131.

Zhao, D., Kong, B. H., Li, F., Cai, H., Wang, L. C., and Chen, H. R. (2007). Cloning and sequencing on CP gene of Cucumber mosaic virus subgroup II isolated from Lily (Lilium cv. Oriental hybrids). *Acta Phytopathol. Sin.* 37, 456–460.

Zhao, F. F., Xi, D. H., Liu, J., Deng, X. G., and Lin, H. H. (2014). First report of Chilli veinal mottle virus infecting tomato (Solanum lycopersicum) in China. *Plant Dis.* 98:1589. doi: 10.1094/PDIS-11-13-1188-PDN

Zheng, X., Zhang, J., Chen, Y. D., Wu, K., and Dong, J. H. (2015). Occurrence dynamics and control status of Tospoviruses in vegetables of Yunnan in 2014. *Shandong Agri c. Sci.* 47, 83–87. doi: 10.3969/j:issn.2095-1191.2015.3.428

Zhou, C. J., Xiang, H. Y., Zhuo, T., Li, D. W., Yu, J. L., and Han, C. G. (2011). A novel strain of Beet western yellows virus, infecting sugar beet with two distinct genotypes differing in the 5'-terminal half of genome. *Virus Genes* 42, 141–149. doi: 10.1007/s11262-010-0553-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 April 2021
doi: 10.3389/fmicb.2021.650773

Identification of a New Genetic Clade of Cowpea Mild Mottle Virus and Characterization of Its Interaction With Soybean Mosaic Virus in Co-infected Soybean

Zhongyan Wei, Chenyang Mao, Chong Jiang, Hehong Zhang, Jianping Chen and Zongtao Sun**

State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University, Ningbo, China

OPEN ACCESS

Edited by:
Xifeng Wang, Institute of Plant Protection (CAAS), China

Reviewed by:
Nicolas Bejerman, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina
Pedro Luis Ramos-González, Biological Institute of São Paulo, Brazil

***Correspondence:**
Jianping Chen
chenjianping@nbu.edu.cn; jpchen2001@126.com
Zongtao Sun
sunzongtao@nbu.edu.cn

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *08 January 2021*
Accepted: *15 March 2021*
Published: *08 April 2021*

Citation:
Wei Z, Mao C, Jiang C, Zhang H, Chen J and Sun Z (2021) Identification of a New Genetic Clade of Cowpea Mild Mottle Virus and Characterization of Its Interaction With Soybean Mosaic Virus in Co-infected Soybean. Front. Microbiol. 12:650773. doi: 10.3389/fmicb.2021.650773

Cowpea mild mottle virus (CPMMV; genus *Carlavirus*) can be a destructive pathogen of soybean but there is little information about its distribution on soybean in China. Here, we collected soybean plants with virus-like symptoms from 11 fields widely scattered within China, and used high-throughput sequencing to determine their virome. Most samples (8/11) were co-infected by the well-studied potyvirus soybean mosaic virus (SMV) and CPMMV, and the remaining three samples were singly infected with CPMMV. The near-complete genome sequences of the 11 CPMMV isolates were determined and phylogenetic analysis showed that they constituted a new genetic clade. One recombination event was detected among the CPMMV sequences, and the isolate CPMMV_JL_CC was identified as recombinant. In mechanical inoculation assays, co-infection by CPMMV and SMV resulted in an enhancement of disease symptoms, but decreased the expression level of the genomic RNAs and CP of CPMMV, without significantly affecting SMV accumulation. The interaction between these viruses needs further investigation.

Keywords: high-throughput sequencing, phylogenetic analysis, genetic variation, recombination, co-infection

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.], one of the most important oil and cash crops worldwide, provides protein-rich food and feed for humans and animals (Wilson, 2008). Many different pests and pathogens attack soybean crops causing significant losses in yield and quality (Widyasari et al., 2020) and some of the most serious pathogens are viruses (Wrather and Koenning, 2006). Soybean virus diseases typically cause 10–30% yield losses but losses of 50–100% have been reported from severe outbreaks (Hartman et al., 2011; Song et al., 2016). Soybean mosaic virus (SMV) is the most important virus infecting soybean worldwide but a wide variety of other viruses have been reported, including cucumber mosaic virus, cowpea mild mottle virus (CPMMV) and alfalfa mosaic virus (Hill and Whitham, 2014).

SMV, a member of the genus *Potyvirus* in the family *Potyviridae*, has a single-strand RNA genome that encodes a single large open reading frame (ORF) (Hajimorad et al., 2018; Gao et al., 2019). The translated polyprotein yields a series of multifunctional proteins through proteolysis, which are commonly named P1, HC-Pro, P3, PIPO (a product of slippage in the P3 coding sequence), 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP (Hajimorad et al., 2018; Gao et al., 2019). SMV is naturally transmitted by aphids in a nonpersistent manner but can also be transmitted through infected seeds (Widyasari et al., 2020). Plants infected with SMV usually have mottling on their leaves, stem necrosis and are stunted. Strains of SMV have been recognized based on their genomic similarity and the response of different soybean cultivars. In this way, isolates were classified into seven strains (G1–G7) in the United States (Cho and Goodman, 1979), and more recently into 22 strains (SC1–SC22) in China (Li et al., 2010; Wang et al., 2013).

CPMMV is another economically important virus that causes serious damage to soybean production particularly in Brazil (Zanardo et al., 2014a,b). It was first reported from cowpea in Ghana and subsequently found in other leguminous crops across the world (Brunt and Kenten, 1973; Zanardo and Carvalho, 2017). CPMMV is a positive-strand RNA virus that belongs to the genus *Carlavirus* in the family *Betaflexiviridae*. The genome of CPMMV (7.8–8.9 kb) encodes six ORFs (ORF 1-6) that are translated into the corresponding proteins: Replicase including the RNA-dependent RNA polymerase (RdRp; ORF1), triple gene block proteins (TGBs; ORFs2-4), coat protein (CP; ORF5) and a nucleic-acid-binding protein (NABP; ORF6) (Zanardo and Carvalho, 2017). CPMMV is transmitted by the whitefly *Bemisia tabaci* in a nonpersistent manner, and the predominant symptoms in soybean are mosaic, stem necrosis and dwarfing (Muniyappa, 1983; Zanardo and Carvalho, 2017). The genomes of CPMMV isolates infecting soybean have been characterized from Brazil, Ghana and India (Brunt and Kenten, 1973; Yadav et al., 2014; Zanardo et al., 2014a,b). Recently, we found CPMMV infecting soybean in Anhui province in China (Wei et al., 2020) but the distribution, importance and biological characteristics of CPMMV in China are largely unknown.

In this study, we collected soybean plants with virus-like symptoms from different regions in China and used high-throughput sequencing (HTS) to investigate the presence of viruses. All the samples contained CPMMV, often in co-infections with SMV, and these CPMMV isolates formed a single clade, distinct from previously reported isolates from elsewhere in the world. Plants co-infected with CPMMV and SMV had enhanced disease symptoms, but had reduced expression levels of the genomic RNA and CP of CPMMV compared to those infected by CPMMV alone.

MATERIALS AND METHODS

Virus Isolates

Leaves of soybean plants with typical virus-like symptoms were collected from 11 fields among six soybean-growing provinces of China (Jilin, Shandong, Henan, Anhui, Jiangsu and Hubei) during September 2019. Pools of leaves from three to five symptomatic plants in each field were wrapped in plastic bags and placed in dry ice.

RNA Sequencing

Total RNA was extracted from each sample by TRIzol Reagent (Invitrogen, United States). Approximately 10 μg of total RNA was used for transcriptome sequencing. The RNA sequencing was performed by Zhejiang Tiangen (Company, Hangzhou, China). The final cDNA library was constructed using the TruSeq Stranded mRNA Library Prep Kit (Illumina, United States) and sequenced on an Illumina HiSeq 4000 (LC Sciences, United States).

Analysis of Sequencing Reads

The reads were generated through the Illumina paired-end RNA-seq approach, the average insert size for the paired-end library was 300 ± 50 bp. Prior to assembly, the low-quality reads were removed. Clean data were assembled using Trinity software (version 2.8.5). The assembled contigs were first compared with Barcode of Life Data (BOLD) Systems[1], and searched against NCBI virus RefSeqs[2] using a BlastX algorithm with a cutoff E-value of 10^{-5}.

RT-PCR Detection

About 1.5 μg of total RNA used in the RNA sequencing assay was reversely transcribed to cDNA using a reverse transcription kit (Tiangen Company, Beijing, China). The cDNA (0.5 μL) was amplified by PCR using KOD-FX (Toyobo, Osaka, Japan), with the following conditions: 94°C for 3 min, 34 cycles of 98°C for 10 s, 58°C for 30 s, and 72°C for 2 min. Primers used for amplifying sequences of specific viruses (CPMMV and SMV) are listed in **Supplementary Table 2**. PCR products were separated on a 1.2% agarose gel.

Phylogenetic Analysis

The nearly complete genome sequences of isolates obtained in this study and reported isolates retrieved from NCBI[3] were aligned by MUSCLE method in MEGAX. Phylogenetic trees were constructed using Maximum-Likelihood (ML) methods using the best-fitting model: GTR + G + I (General Time Reversible + Gama Distributed With Invariant Sites). To check the reliability of the constructed trees, the bootstrap test with 1,000 bootstrap replications was used. For the CPMMV phylogenetic tree construct, Indian citrus ringspot virus (*Mandarivirus*, *Alphaflexiviridae*) was used as the outgroup. Turnip mosaic virus (*Potyvirus, Potyviridae*) was used as the outgroup for the SMV phylogenetic tree construct.

Recombination Analysis

Putative recombination events amongst CPMMV isolates and SMV isolates were identified using the recombination detection program RDP5 (Martin et al., 2020), and evaluated using different

[1]https://www.boldsystems.org/

[2]ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral

[3]https://www.ncbi.nlm.nih.gov/

methods: RDP, GENECONV, BOOTSCAN, MaxiChi, Chimera, SiScan, and 3Seq. Alignments of nucleotide sequences produced in MEGAX were run in RDP5 (P-value cut-off of 0.05). Only recombination events that were detected by three or more methods were considered.

Mechanical Transmission Assay

For inoculation of soybean plants, sap from each of the 11 symptomatic plant pools was used to inoculate a susceptible soybean variety (Jiunong 9). Each inoculum was prepared from 1 g of symptomatic field plant leaves homogenized with 10 mL of 0.01 M phosphate-buffered saline, pH 7.0. After mixing with carborundum, inoculation was performed manually before the trifoliate leaves emerged and the inoculum finally rinsed with tap water. Plants inoculated with phosphate-buffered saline were used as controls. Inoculated plants were grown at 25–28°C (16 h light/8 h dark) in an incubator for 7–10 days, and classified into four phenotype classes based on symptoms of 12 plants per inoculation: symptomless, mosaic, crinkling, stem necrosis. Three independent experiments were conducted to provide data for statistical analysis, Values are means ± Standard Deviation (*SD*).

RNA Extraction and RT-qPCR

Total RNA was extracted from young trifoliate leaves using TRIzol reagent (Invitrogen, United States). About 1.5 μg of total RNA was reversely transcribed to cDNA using a reverse transcription kit (Tiangen Company, Beijing, China). Real-time PCR was conducted using ChamQ^{TM} SYBR qPCR Master Mix (Low ROX Premixed) by an ABI7900HT Sequence Detection System (Applied Biosystems, Carlsbad, CA, United States). The RT-qPCR conditions were as follows: 95°C for 4 min; 40 cycles of 95°C for 10 s and 60°C for 30 s. The relative expression levels of genes were determined using the $2^{-\Delta\Delta C(t)}$ method (Livak and Schmittgen, 2001). The soybean Actin11 gene was used as an internal control. Three biological and two technical replicates were conducted to determine gene expression. The RT-qPCR primer sequences used in this study are listed in **Supplementary Table 2**.

Western Blot

Approximately 200 mg soybean leaves were homogenized in 0.3 mL sodium dodecylsulfate (SDS) lysis buffer (100 mm Tris-HCl, pH 6.8, 10% SDS and 2.0% β -mercaptoethanol). The crude extracts were centrifuged at 12,000 × g for 10 min at room temperature, and the resulting supernatant (8 μL per sample plus 2 μL 5 × SDS loading buffer) was electrophoresed in 10–12% SDS-PAGE gels. Western blot analysis was done as previously reported (Zhang et al., 2019). Proteins were transferred to polyvinylidene difluoride (PVDF, Millipore, United States) membranes using the Trans-Blot Turbo transfer system (Bio-Rad, United States). Polyclonal rabbit anti-CPMMV-CP and anti-SMV-CP (at 1:3,000 dilution, synthesized by Huabio, China) were used to detect the respective viruses.

RESULTS

Occurrence of SMV and CPMMV in China

Soybean samples with mosaic and crinkling symptoms on their leaves were collected from 11 locations in six provinces (Jilin, Shandong, Henan, Anhui, Jiangsu and Hubei) in the main soybean producing region of China (**Figures 1A–C** and **Table 1**). After extraction of total RNA, high-throughput sequencing (HTS) yielded more than 25 M paired-end reads from each sample (ranging from 25.38 to 54.50 M). BLASTX analysis of the assembled contigs searched against the NCBI virus RefSeq database showed that SMV and CPMMV were widely found in these samples. Most samples (8/11) were co-infected by SMV

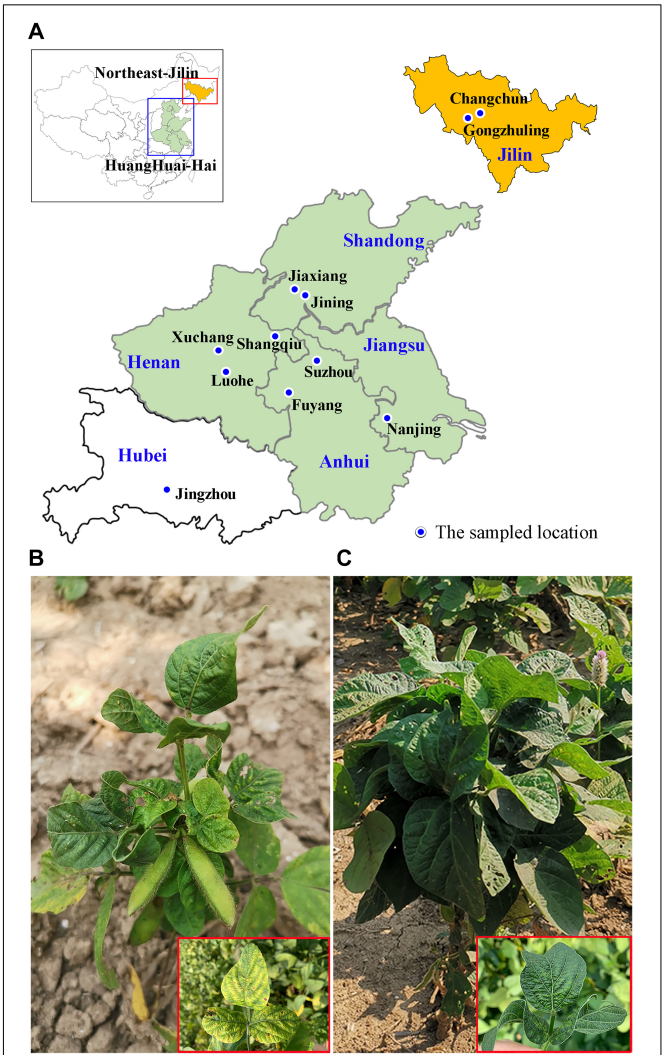


FIGURE 1 | The geographical locations of the symptomatic soybean samples collected in China and the symptoms on leaves of the sampled plants. **(A)** Map of China (top left corner) showing the study area (red and blue rectangular box). The 11 sampled fields are marked on the enlarged abridged map. **(B,C)** Leaves of soybean plants displaying typical yellow mosaic **(B)** and crinkling **(C)** virus-like symptoms.

TABLE 1 | Summary of Illumina sequencing and assembling statistics.

Location	Sample ID	Total reads	SMV			CPMMV		
			Acc. no.	Length (nt)	Alignment rate*	Acc. No.	Length (nt)	Alignment rate*
Anhui-Suzhou	AH_SZ	29.96 M	MW354946	9,990 bp	0.03%	MN908944	8,202 bp	23.07%
Anhui-Fuyang	AH_FY	32.56 M	–	–	–	MW354945	8,212 bp	18.23%
Shandong-Jiaxiang	SD_JX	54.50 M	MW354949	7,326 bp	1.09%	MW354937	8,202 bp	0.19%
Shandong-Jining	SD_JN	30.85 M	MW354953	7,137 bp	0.08%	MW354941	8,200 bp	0.11%
Henan-Luohe	HN_LH	41.61 M	MW354950	7,882 bp	0.12%	MW354936	8,200 bp	3.02%
Henan-Xuchang	HN_XC	28.70 M	–	–	–	MW354938	8,175 bp	0.24%
Henan-Shangqiu	HN_SQ	25.74 M	MW354952	7,167 bp	0.03%	MW354940	8,204 bp	4.22%
Jingsu-Nanjing	JS_NJ	30.23 M	MW354948	10,008 bp	43.11%	MW354944	8,338 bp	12.31%
Jilin-Changchun	JL_CC	45.80 M	MW354951	9,947 bp	18.61%	MW354943	8,212 bp	0.03%
Jilin-Gongzhuling	JL_GZL	26.38 M	MW354947	9,604 bp	21.09%	MW354942	8,172 bp	0.01%
Hubei-Jingzhou	HB_JZ	25.38 M	–	–	–	MW354939	8,178 bp	0.12%

**Virus-specific reads/reads.*

and CPMMV, and the remaining three samples (AH_FY, HN_XC and HB_JZ) were infected by CPMMV alone (**Table 1**). These results were confirmed by reverse transcription PCR (RT-PCR) using primers specific for SMV or CPMMV (**Supplementary Figure 1**). These results indicate that CPMMV is distributed in several soybean producing areas in China and is commonly present in co-infections with SMV.

Phylogenetic Analysis of SMV and CPMMV Isolates

From the transcriptome sequencing data, the near-complete genome sequences of four SMV isolates and >7kb of four others (**Supplementary Figure 2**) were assembled and deposited in GenBank (**Table 1**). To determine the phylogenetic relationships between these eight SMV sequences and other known SMV genomes (G1-G7, United States; SC3, SC7 and SC15, the most important strains in China), we performed phylogenetic analysis based on the nucleotide sequence of the region from P3 to 6K2 using Maximum-Likelihood (ML) methods. The results showed four major clades and that the SMV isolates from China clustered separately from those of USA (**Figure 2**). Clade IV included all the new SMV isolates from this study (except SMV_JL_GZL from Jilin province) and the SC7 strain, while the isolate SMV_JL_GZL from northeast of China clustered with the SC3 strain. Clades II and III included the USA isolates G1–G7.

We assembled the complete CPMMV sequences from each of our 11 samples, deposited them in GenBank (**Table 1**), and examined the phylogenetic relationships between them and some representative complete CPMMV sequences from Brazil, United States, India, Mexico and Ghana. As shown in **Figure 3**, there were three lineages. The CPMMV isolates from our samples were closely related to each other within clade III and with a Chinese cowpea isolate (KY420906.1_Hainan). The CPMMV isolates from Brazil (soybean), Mexico (common bean) and Florida (whitefly) were placed in Clade II, while the isolates from Ghana and India were in clade I. Thus, the CPMMV isolates from China form a very distinct cluster. In phylogenetic analyses of each ORF separately, the same pattern of clustering was consistently obtained (**Supplementary Table 1** and **Supplementary Figure 3**).

Sequence Comparisons of CPMMV Isolates

Because the genetic diversity among CPMMV isolates has been little studied, sequence comparisons were made separately for each ORF of the CPMMV isolates. Among the isolates from China, the nucleotide identity of each ORF was in the range from 95 to 100% but the values were much lower when the Chinese isolates were compared with those from other countries (63.5–82.5%) (**Supplementary Figure 4**). ORF1 of the Chinese isolates all encoded 1,860 amino acids (**Supplementary Table 1**) and were very similar to one another (95.3–99.6% nt identity), but had 79.3–81.0% identity with Brazilian isolates and 75 and 64% identity with those from India and Ghana, respectively. In ORF2, the isolates collected in this study had 98.1–99.7% nucleotide identity to one another and the ORF was 11 codons shorter than that of the previously reported CPMMV isolate on cowpea from China. ORF3 and ORF4 were the same size among the isolates from China, Brazil, Florida and India, but not Ghana (**Supplementary Table 1**). ORF5 encodes the coat protein of 289 aa (**Supplementary Table 1**), and is the most conserved region among all isolates (**Supplementary Figure 4**). Among the isolates from China, there was 97.6–100% nt identity and 100% amino acid identity in this ORF, which had 78.3–78.9% nt identity with that of the isolate from Ghana. The ORF6 of Chinese isolates were all the same size and shared 94.9–100% nt identity, but only 59.4–88.5% identity isolates from India and Ghana (**Supplementary Figure 4**). Overall, our results confirm that the isolates from China are different from those reported in other countries, and represent a unique CPMMV strain.

Recombination Analyses of CPMMV and SMV Isolates

To detect possible recombination events amongst the CPMMV isolates, the complete coding sequences of all known CPMMV isolates in NCBI were analyzed using RDP 5. Four putative

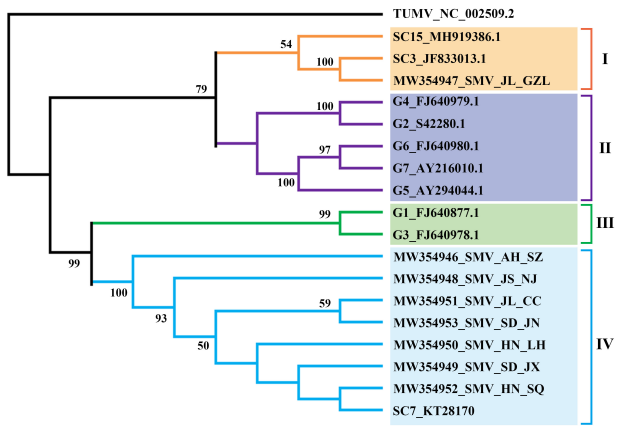


FIGURE 2 | ML-phylogenetic relationships among soybean mosaic virus (SMV) isolates from different countries. The phylogenic tree was constructed using MEGAX. Near-complete genome sequences of SMV isolates were obtained from NCBI (https://www.ncbi.nlm.nih.gov/). The colored bars (blue, orange, green, purple) indicate the different clades observed in the phylogenies.

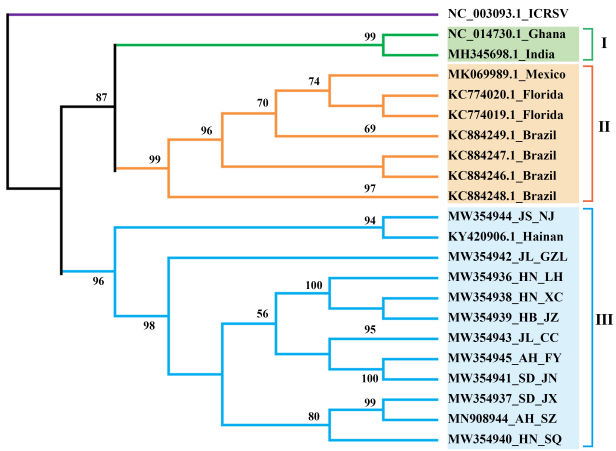


FIGURE 3 | ML-phylogenetic analysis of the complete genomes of cowpea mild mottle virus (CPMMV) isolates from different countries. The phylogenic tree was constructed using MEGAX. Near-complete genome sequences of CPMMV isolates were obtained from NCBI (https://www.ncbi.nlm.nih.gov/). The colored bars (blue, orange, green) indicate the different clades observed in the phylogenies. Blue, China; Orange, Brazil and Florida; Green, India, and Ghana.

recombination events were detected among the isolates but only one among the Chinese isolates (**Figure 4A**). The isolates CPMMV_HN-SQ and CPMMV_JS-NJ were identified as the respective major and minor parents of CPMMV_JL_CC in a single recombination event identified in the region from nt 2,164 to 3,308 in ORF1 (**Figures 4A–C**). The beginning and ending breakpoints were identified at nt 2,891 and 3,146. This result was supported by five methods (RDP, $p = 2.022 \times 10^{-05}$; Geneconv, $p = 5.858 \times 10^{-04}$; Bootscan, $p = 1.931 \times 10^{-05}$; MAXchi, $p = 2.586 \times 10^{-02}$; SiScan, $p = 2.781 \times 10^{-04}$) with high confidence and phylogenetic analysis of the putative recombinant and non-recombinant portions also support the conclusion (**Supplementary Figure 5**).

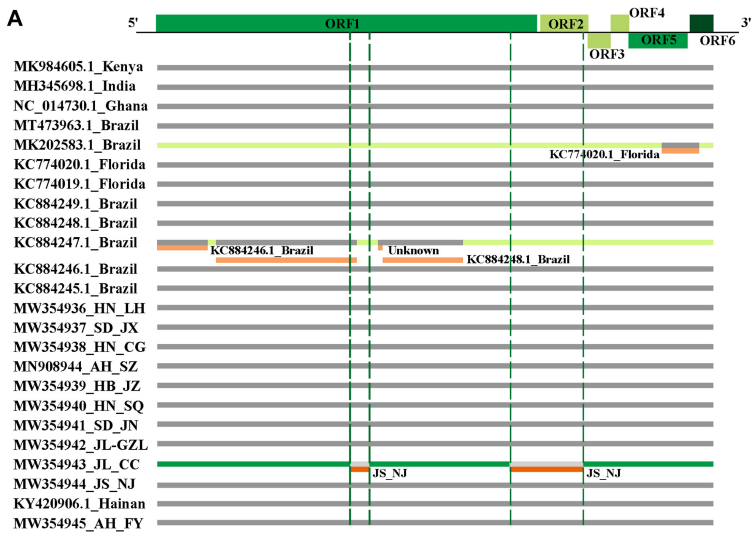
We also detected six recombination events in the SMV population using the nucleotide sequence covering the region of the genome from P3 to 6K2. Two recombination events occurred in the SMV isolates we identified in this study. Recombination event 1 showed SC7 as the putative major parent and SMV_JL_GZL as the minor parent, which led to recombinant isolate SMV_AH_SZ. SMV_JL_GZL was identified as anther recombinant isolate, with G4 and isolate SMV_AH_SZ as possible parents (**Supplementary Figure 6**). This result was supported by four to six methods with high confidence.

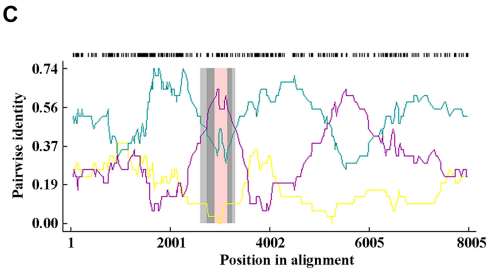
Symptoms on Mechanically Inoculated Plants

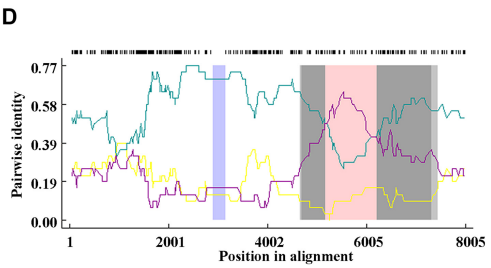
To further characterize the symptoms of CPMMV/SMV caused by the different isolates studied, the samples were used for mechanical transmission assays on susceptible soybean variety "Jiunong 9." Successful infection was established following inoculation by seven of the mixed sap samples (three to five symptomatic plants) from each field, with systemic symptoms including stem necrosis, mosaic, and crinkled leaves appearing from 7 to 14 days post-inoculation (dpi) (**Figure 5A**). Most of the plants inoculated with JL_GZL (8.67 ± 1.15/12) developed leaf crinkling while the other plants had leaf mosaic (3.33 ± 1.15/12). About 60% of plants inoculated with sample JL_CC became infected, mainly showing mosaic symptoms (7.00 ± 0.00/12) (**Figures 5A,B**). RT-qPCR and western-blotting detected RNA and coat protein (CP) of SMV but not CPMMV in these plants (**Figures 5C,D**) even though the field samples had been co-infected with both viruses. By contrast, plants inoculated with sample SD_JX, developed leaf mosaic (3.33 ± 0.58/12) or stem necrosis (8.67 ± 0.58/12) (**Figures 5A,B**) and the RT-qPCR and immunoblot analyses showed that CPMMV, but not SMV, was present in the inoculated plants. Inoculation with either JS_NJ or AH_SZ, resulted in co-infection by CPMMV and SMV (**Figures 5C,D**) with mosaic symptoms on the upper leaves at the early infection stages, followed by systemic necrosis throughout the plants (**Figure 5B**). HTS results had shown that field sample AH_FY contained only CPMMV (**Table 1**). Most of the plants mechanically inoculated with AH_FY developed mild stem necrosis symptoms (7.67 ± 1.15/12), while the remainder had leaf mosaic (4.33 ± 1.15/12).

Co-infection of SMV and CPMMV Decreased the Accumulation of CPMMV

To understand the interaction between CPMMV and SMV in co-infection, the sap from mechanically inoculated leaves of isolates AH_FY (CPMMV) and JL_GZL (SMV) were inoculated onto soybean seedlings separately or together. Control (mock) plants were inoculated with phosphate buffer. The plants singly infected with CPMMV or SMV developed mild stem necrosis or mosaic, respectively, at 7 dpi (**Figure 6A**), while co-infection caused more severe symptoms:







B

Event number	Recombination	Major parent	Minor parent	Detection methods R	G	B	M	C	S	T
1	MW354943_JL_CC	MW354940_HN_SQ	MW354944_JS_NJ	+	+	+	+	-	+	-
2	MK202583.1_Brazil	MK884248.1_Brazil	KC774020.1_Florida	+	+	+	+	+	+	+
3	MK884247.1_Brazil	MK882426.1_Brazil	Unknown	+	+	-	+	+	-	+
4	MK884247.1_Brazil	Unknown	MK884248.1_Brazil	+	+	+	+	+	+	+

FIGURE 4 | Recombination events among the CPMMV isolates. **(A)** Schematic representation of the recombination events identified among CPMMV isolates by the RDP 5 program. Each box represents a viral isolate with the recombination events identified by different color (green and orange). The CPMMV genome is shown at the top of the figure. **(B)** Details of recombination events in the complete coding sequence of CPMMV isolates. Recombination detection methods are represented by letters: R, RDP; G, Genecov; B, Bootscan; M, Maximum v2; C, Chimera; S, Sister scan; 3, 3 Seq. **(C,D)** Evidence of recombination event in the isolate JL_SMV1.

systemically infected leaves became crinkled at 5 dpi and developed systemic necrosis at 7 dpi (**Figure 6A**). RT-qPCR (**Figures 6B,C**) and western-blotting (**Figures 6D,E**) at 7 and 14 dpi confirmed the presence of both viruses. Compared to plants inoculated with a single virus, the accumulation of CPMMV genomic RNA and of CP were decreased in co-infected leaves but there were no significant effects on the levels of SMV RNA or CP.

DISCUSSION

CPMMV was first discovered on cowpea in Ghana in 1973 (Brunt and Kenten, 1973), and subsequently found infecting soybeans in Thailand and the Ivory Coast in the 1980s (Iwaki, 1982; Thouvenel et al., 1982). CPMMV was considered of minor importance until the occurrence of outbreaks in soybean across Brazil in the 2000s (Zanardo et al., 2014a). CPMMV is transmitted by the widespread whitefly *Bemisia tabaci*. It induces stem necrosis symptoms, and can cause economic losses in soybean production (Jossey, 2013; Zanardo and Carvalho, 2017). There are few studies on CPMMV: of the 16 complete genomic sequences of CPMMV available in GenBank, most (8/16) are from Brazil. The complete sequences of two CPMMV isolates from China have previously been reported, one from cowpea in Hainan (Yang et al., 2019), and our recent study providing the first report of CPMMV infecting soybean in China showing that it was the cause of leaf mosaic and crinkling symptoms (Wei et al., 2020). We have now demonstrated that CPMMV is widespread in the soybean producing areas in China and was present in all 11 samples collected. Their genomic sequences were obtained through HTS and phylogenetic analysis showed that all the Chinese isolates form a distinct cluster with 95–100% nucleotide identity to one another but rather distantly related to CPMMV isolates from other countries. In the family *Betaflexiviridae*, viruses are classified in the same species if their CP or replicase proteins have more than 72% nt or 80% aa identity (Adams et al., 2012). The ORF1 (RdRp) and CP of isolates from China have 79.3–81.0% and 83.0–90.0% nt identity to Brazilian isolates and clearly belong to the same species.

In Brazil, CPMMV causes a significant threat to soybean production and has received particular attention (Zanardo and Carvalho, 2017). The molecular variability of Brazilian CPMMV isolates has been investigated and it was concluded that the topology of the phylogenetic tree was not related to the geographical origin of isolates within Brazil (Zanardo et al., 2014a,b). Although the Chinese isolates form a separate branch of the CPMMV tree, there was similarly no evidence of further sub-clades based on their origin within China. This may suggest that CPMMV is a relatively recent introduction in both Brazil and China, or perhaps that the vector is able to rapidly distribute virus variants within the country. In addition, our phylogenetic analysis of SMV isolates also did not show clear relationships with

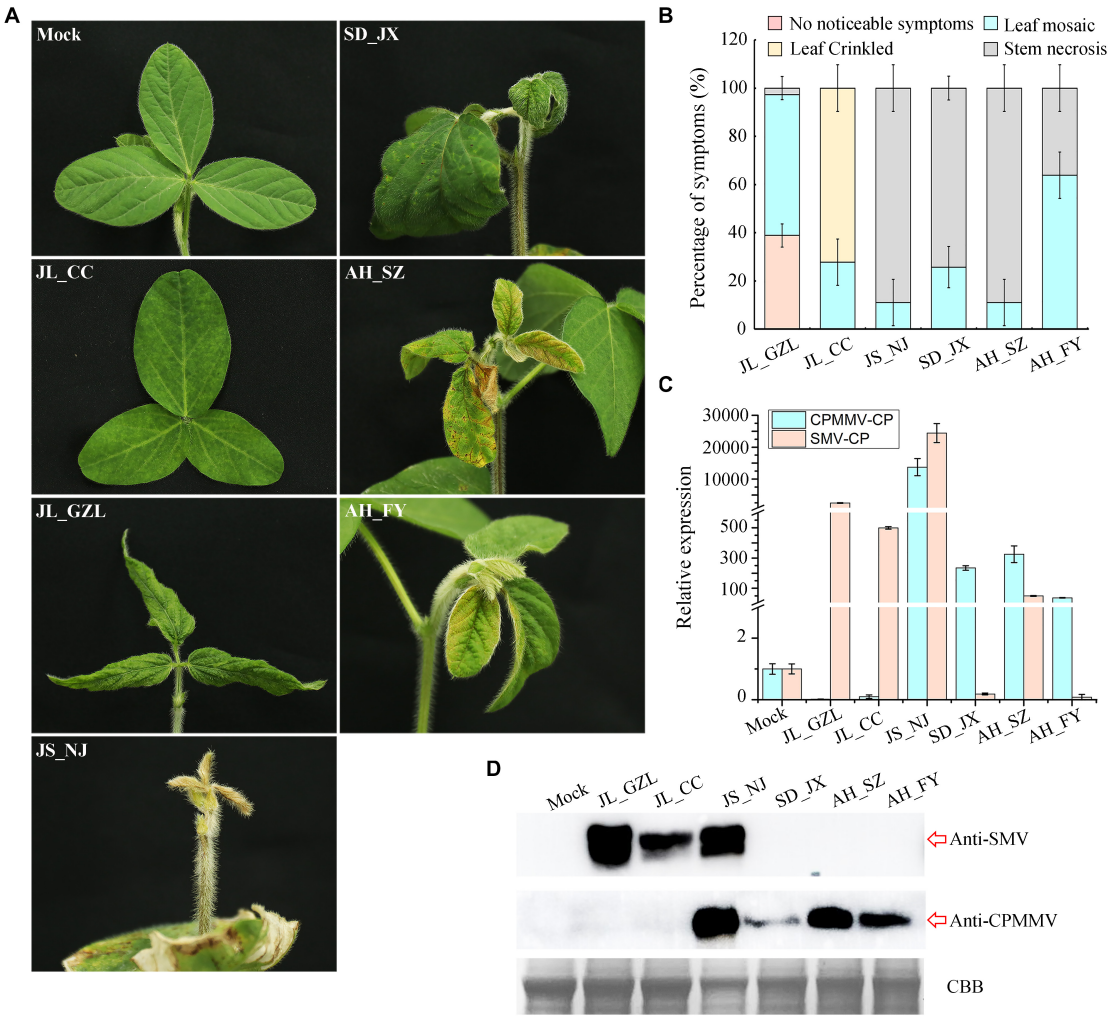


FIGURE 5 | Symptoms of SMV/CPMMV infection on soybean. **(A)** Photographs of soybean (Jiunong 9) pants 10 days after inoculation with sap from different field locations. **(B)** The percentages of SMV/CPMMV-infected plants with different grades of disease symptoms, $n = 12$. **(C)** SMV and CPMMV RNAs in infected plants as determined by RT-qPCR. Data are shown as means ± *SD* of three biological replicates. **(D)** SMV and CPMMV coat protein (CP) protein levels in mock-inoculated and SMV/CPMMV-infected plants as determined by immunoblot analysis. Coomassie brilliant blue (CBB) staining of the same extracts is shown to demonstrate equal loading.

their geographical origin. All SMV isolates (except SMV_JL_GZL from Jilin province) of this study were clustered in one clade with SC7 strain, and distinct from the US strains (G1–G7) (**Figure 2**). Recombination may explain the position of the SMV_JL_GZL isolate (**Supplementary Figure 6**). Recombination events have been identified in several viruses (Gagarinova et al., 2008; Zanardo et al., 2014b; Kalyandurg et al., 2017; Chikh-Ali et al., 2019). Our analysis detected one recombination event within the CPMMV isolates and two recombination events within the SMV isolates from China (**Figure 4** and **Supplementary Figure 6**), indicating that recombination is one of the most important factors that contribute to the variability and evolution of SMV and CPMMV.

Previous studies have suggested that virus epidemics in soybean production in China are mainly caused by SMV and co-infection of soybean by SMV and CPMMV has not previously been reported. Our attempts to transmit virus from the co-infected field samples by mechanical inoculation sometimes led to the loss of one of the viruses. The mechanisms of co-infection are diverse in different host-pathogen systems (Bance, 1991; Gonzalez-Jara et al., 2004). Several lines of biochemical and genetic evidence have shown that viral interaction patterns in co-infection may depend on the host cultivar. For example, Wheat streak mosaic virus and Triticum mosaic virus induced cultivar-specific disease synergism in three wheat cultivars (Tatineni et al., 2010). We suppose that the different outcomes of mechanical inoculation in our study may be a consequence of differences in the virus content of field samples and host cultivar-pathogen specificity.

Co-infections by two or more viruses are common in the field. Although co-infections can be either synergistic or antagonistic (Syller, 2012; Bian et al., 2020), they usually cause more severe

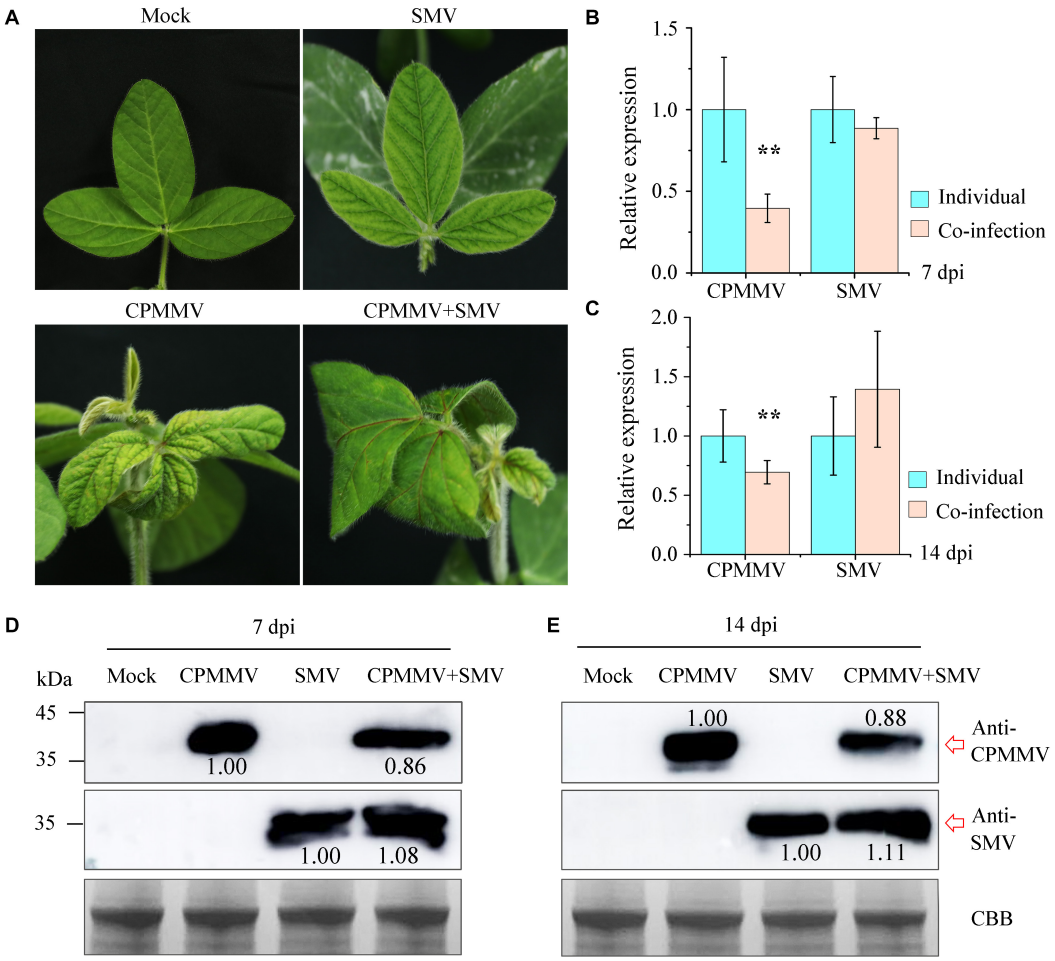


FIGURE 6 | Co-infection of CPMMV and SMV increased the accumulation of SMV. **(A)** Symptoms on the first newly-emerged leaves 7 days after inoculation with SMV, CPMMV or SMV+CPMMV. **(B,C)** The relative expression levels of SMV and CPMMV RNAs as determined by RT-qPCR at 7 dpi **(B)** and 14 dpi **(C)**. Data are shown as means ± *SD* of three biological replicates, $^{**}P < 0.01$, Student's *t*-tests. **(D,E)** SMV and CPMMV coat protein (CP) protein levels in mock-inoculated and SMV/CPMMV-infected plants as determined by immunoblot analysis at 7 dpi **(D)** and 14 dpi **(E)**. Coomassie brilliant blue (CBB) staining of the same extracts is shown to demonstrate equal loading.

symptoms and significant damage to crops (Malapi-Nelson et al., 2009; Mahuku et al., 2015; Redinbaugh and Stewart, 2018). In our case, field observation and seedling inoculations suggest that co-infection causes more severe symptoms. However, at the molecular level, there would appear to be some antagonism between the viruses since RT-qPCR and western-blotting results showed that the accumulation level of CPMMV genomic RNAs and CP were decreased in co-infected leaves. This suggests that in co-infection with SMV, CPMMV interacts antagonistically. Similar observations have been reported with SMV and AMV infection in soybean, which showed severe symptoms in doubly infected plants, but the level of SMV accumulation was reduced. Conversely, the level of AMV accumulation was increased indicating that the interaction of AMV with SMV is synergistic (Malapi-Nelson et al., 2009). In the current study, there was no significant effect on the accumulation level of SMV in co-infected leaves. There appears to be a complex mechanism of interaction between SMV and CPMMV that requires further investigation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The sequences of CPMMV and SMV isolates were deposited in The National Center for Biotechnology Information (NCBI) GenBank, https://www.ncbi.nlm.nih.gov/genbank/ and the accession numbers can be found in the article/**Supplementary Material**. The other raw data supporting the conclusions of this article are available upon request from the corresponding author, without undue reservation.

AUTHOR CONTRIBUTIONS

ZW and ZS designed the research and performed the experiments. ZW drafted the manuscript. ZS revised the manuscript. CM and CJ supervised the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Science Fund for Excellent Young Scholars (32022072), Zhejiang Provincial Natural Science Foundation of China (LQ21C140004), the China National Novel Transgenic Organisms Breeding Project (2019ZX08004-004), and K. C. Wong Magna Fund in Ningbo University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.650773/full#supplementary-material

Supplementary Figure 1 | Detection of CPMMV and SMV in the total RNA of symptomatic field soybean samples as shown by RT-PCR.

Supplementary Figure 2 | The genome organization of SMV isolates.

Supplementary Figure 3 | NJ-phylogenetic relationships among the ORFs of CPMMV isolates from different countries.

Supplementary Figure 4 | Pairwise identity matrix of CPMMV isolates for each open reading frame (ORF).

Supplementary Figure 5 | Phylogenetic trees analysis based on different nucleotide fragments of CPMMV isolates from China.

Supplementary Figure 6 | Recombination events among the SMV isolates.

Supplementary Table 1 | Genomic organization of the 11 CPMMV isolates described in this study.

Supplementary Table 2 | List of primer sequences used in this study.

REFERENCES

Adams, M. J., Candresse, T., Hammond, J., Kreuze, J. F., and Yoshikawa, N., et al. (2012). "Family *Betaflexiviridae*," in *Virus Taxonomy: The Classification and Nomenclature of Viruses.* The 9th Report of the ICTV, eds A. M. Q. King, M. J. Adams, E. B. Carstens, and E. J. Lefkowitz (London: Academic Press), 920–941.

Bance, V. B. (1991). Replication of potato virus X RNA is altered in coinfections with potato virus Y. *Virology* 182, 486–494. doi: 10.1016/0042-6822(91)90589-4

Bian, R. L., Andika, I. B., Pang, T. X., Lian, Z. Q., Wei, S., Niu, E., et al. (2020). Facilitative and synergistic interactions between fungal and plant viruses. *Proc. Natl. Acad. Sci. U.S.A.* 117, 3779–3788. doi: 10.1073/pnas.1915996117

Brunt, A. A., and Kenten, R. H. (1973). Cowpea mild mottle, a newly recognized virus infecting cowpeas (*Vigna unguiculata*) in Ghana. *Ann. Appl. Biol.* 74, 67–74. doi: 10.1111/j.1744-7348.1973.tb07723.x

Chikh-Ali, M., Rodriguez-Rodriguez, M., Green, K. J., Kim, D. J., Chung, S. M., Kuhl, J. C., et al. (2019). Identification and molecular characterization of recombinant Potato Virus Y (PVY) in potato from South Korea, PVY(NTN) strain. *Plant Dis.* 103, 137–142. doi: 10.1094/PDIS-05-18-0715-RE

Cho, E. K., and Goodman, R. M. (1979). Strains of soybean mosaic virus: classification based on virulence in resistant soybean cultivars. *Phytopathology* 69, 467–470. doi: 10.1094/Phyto-69-467

Gagarinova, A. G., Babu, M., Stromvik, M. V., and Wang, A. (2008). Recombination analysis of Soybean mosaic virus sequences reveals evidence of RNA recombination between distinct pathotypes. *Virol. J.* 5:143. doi: 10.1186/1743-422X-5-143

Gao, L., Luo, J. Y., Ding, X. N., Wang, T., Hu, T., Song, P. W., et al. (2019). Soybean RNA interference lines silenced for eIF4E show broad potyvirus resistance. *Mol. Plant Pathol.* 21, 303–317. doi: 10.1111/mpp.12897

Gonzalez-Jara, P., Tenllado, F., Martinez-Garcia, B., Atencio, F. A., Barajas, D., Vargas, M., et al. (2004). Host-dependent differences during synergistic infection by Potyviruses with potato virus X. *Mol. Plant Pathol.* 5, 29–35. doi: 10.1111/j.1364-3703.2004.00202.x

Hajimorad, M. R., Domier, L. L., Tolin, S. A., Whitham, S. A., and Saghai Maroof, M. A. (2018). Soybean mosaic virus: a successful potyvirus with a wide distribution but restricted natural host range. *Mol. Plant Pathol.* 19, 1563–1579. doi: 10.1111/mpp.12644

Hartman, G. L., West, E. D., and Herman, T. K. (2011). Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Secur.* 3, 5–17. doi: 10.1007/s12571-010-0108-x

Hill, J. H., and Whitham, S. A. (2014). Control of virus diseases in soybeans. *Adv. Virus Res.* 90, 355–390. doi: 10.1016/B978-0-12-801246-8.00007-X

Iwaki, M. (1982). Whitefly transmission and some properties of cowpea mild mottle virus on soybean in Thailand. *Plant Dis.* 66:365. doi: 10.1094/PD-66-365

Jossey, S. (2013). Role of soybean mosaic virus-encoded proteins in seed and aphid transmission in soybean. *Phytopathology* 103, 941–948. doi: 10.1094/PHYTO-09-12-0248-R

Kalyandurg, P., Gil, J. F., Lukhovitskaya, N. I., Flores, B., Muller, G., Chuquillanqui, C., et al. (2017). Molecular and pathobiological characterization of 61 Potato mop-top virus full-length cDNAs reveals great variability of the virus in the centre of potato domestication, novel genotypes and evidence for recombination. *Mol. Plant Pathol.* 18, 864–877. doi: 10.1111/mpp.12552

Li, K., Yang, Q. H., Zhi, H. J., and Gai, J. Y. (2010). Identification and distribution of soybean mosaic virus strains in southern China. *Plant Dis.* 94, 351–357. doi: 10.1094/PDIS-94-3-0351

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Mahuku, G., Lockhart, B. E., Wanjala, B., Jones, M. W., Kimunye, J. N., Stewart, L. R., et al. (2015). Maize lethal necrosis (MLN), an emerging threat to maize-based food security in sub-saharan Africa. *Phytopathology* 105, 956–965. doi: 10.1094/PHYTO-12-14-0367-FI

Malapi-Nelson, M., Wen, R. H., Ownley, B. H., and Hajimorad, M. R. (2009). Co-infection of soybean with soybean mosaic virus and alfalfa mosaic virus results in disease synergism and alteration in accumulation level of both viruses. *Plant Dis.* 93, 1259–1264. doi: 10.1094/PDIS-93-12-1259

Martin, D. P., Varsani, A., Roumagnac, P., Botha, G., Maslamoney, S., Schwab, T., et al. (2020). RDP5: a computer program for analysing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* veaa087. doi: 10.1093/ve/veaa087

Muniyappa, V. (1983). Transmission of cowpea mild mottle virus by *Bemisia tabaci* in a nonpersistent manner. *Plant Dis.* 67, 391–393. doi: 10.1094/PD-67-391

Redinbaugh, M. G., and Stewart, L. R. (2018). Maize lethal necrosis: an emerging, synergistic viral disease. *Ann. Rev. Virol.* 5, 301–322. doi: 10.1146/annurev-virology-092917-04341

Song, Y. P., Cui, L., Lin, Z., Karthikeyan, A., Na, L., Kai, L., et al. (2016). Disease spread of a popular soybean mosaic virus strain (SC7) in southern China and effects on two susceptible soybean cultivars. *Philipp. Agric. Sci.* 99, 355–364.

Syller, J. (2012). Facilitative and antagonistic interactions between plant viruses in mixed infections. *Mol. Plant Pathol.* 13, 204–216. doi: 10.1111/j.1364-3703.2011.00734.x

Tatineni, S., Graybosch, R. A., Hein, G. L., Wegulo, S. N., and French, R. (2010). Wheat cultivar-specific disease synergism and alteration of virus accumulation during co-infection with Wheat streak mosaic virus and *Triticum* mosaic virus. *Phytopathology* 100, 230–238. doi: 10.1094/PHYTO-100-3-0230

Thouvenel, J. C., Monsarrat, A., and Fauquet, C. (1982). Isolation of cowpea mild mottle virus from diseased soybeans in the Ivory Coast. *Plant Dis.* 66, 336–337. doi: 10.1094/PD-66-336

Wang, D. G., Tian, Z., Li, K., Li, H. W., Huang, P. P., Hu, G. Y., et al. (2013). Identification and variation analysis of soybean mosaic virus strains in Shan-Dong, Henan and Anhui Provinces of China. *Soybean Sci.* 32, 806–809.

Wei, Z. Y., Wu, G. W., Ye, Z. X., Jiang, C., Mao, C. Y., Zhang, H. H., et al. (2020). First report of cowpea mild mottle virus infecting soybean in China. *Plant Dis.* 104, 2534–2534. doi: 10.1094/PDIS-01-20-0063-PDN

Widyasari, K., Alazem, M., and Kim, K. H. (2020). Soybean resistance to soybean mosaic virus. *Plants (Basel)* 9:219. doi: 10.3390/plants9020219

Wilson, R. F. (2008). *Soybean: Market Driven Research Needs*. New York, NY: Springer, 3–14.

Wrather, J. A., and Koenning, S. R. (2006). Estimates of disease effects on soybean yields in the United States 2003 to 2005. *J. Nematol.* 38, 173–180.

Yadav, M. K., Biswas, K. K., Lal, S. K., Baranwal, V. K., and Jain, R. K. (2014). A distinct strain of cowpea mild mottle virus infecting soybean in India. *J. Phytopathol.* 161, 739–744. doi: 10.1111/jph.12119

Yang, X., Zhang, D. Y., Zhang, Y., Chen, S., Li, S. S., Gong, Y. N., et al. (2019). Genomic sequences measure and molecular characteristics of cowpea mild mottle virus of hainan isolate. *China Vegetables* 364, 41–44.

Zanardo, L. G., and Carvalho, C. M. (2017). Cowpea mild mottle virus (Carlavirus, Betaflexiviridae): a review. *Trop. Plant Pathol.* 42, 417–430. doi: 10.1007/s40858-017-0168-y

Zanardo, L. G., Silva, F. N., Bicalho, A. A. C., Urquiza, G. P. C., Lima, A. T. M., Almeida, A. M. R., et al. (2014a). Molecular and biological characterization of Cowpea mild mottle virus isolates infecting soybean in Brazil and evidence of recombination. *Plant Pathol.* 63, 456–465. doi: 10.1111/ppa.12092

Zanardo, L. G., Silva, F. N., Lima, A. T., Milanesi, D. F., Castilho-Urquiza, G. P., Almeida, A. M., et al. (2014b). Molecular variability of cowpea mild mottle virus infecting soybean in Brazil. *Arch. Virol.* 159, 727–737. doi: 10.1007/s00705-013-1879-0

Zhang, H. H., Tan, X. X., He, Y. Q., Xie, K. L., Li, L. L., Wang, R., et al. (2019). Rice black-streaked dwarf virus P10 acts as either a synergistic or antagonistic determinant during superinfection with related or unrelated virus. *Mol. Plant Pathol.* 20, 641–655. doi: 10.1111/mpp.12782

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 April 2021
doi: 10.3389/fmicb.2021.662352

Biological and Genetic Characterization of Pod Pepper Vein Yellows Virus-Associated RNA From *Capsicum frutescens* in Wenshan, China

Jiejun Peng[1†], Shan Bu[2,3†], Yueyan Yin[4,5], Mengying Hua[1], Kuangjie Zhao[1], Yuwen Lu[1], Hongying Zheng[1], Qionglian Wan[4], Songbai Zhang[2], Hairu Chen[4], Yong Liu[2], Jianping Chen[1], Xiaohan Mo[6] and Fei Yan[1]**

[1] State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agroproducts, Institute of Plant Virology, Ningbo University, Ningbo, China, [2] Key Laboratory of Pest Management of Horticultural Crop of Hunan Province, Hunan Plant Protection Institute, Hunan Academy of Agricultural Sciences, Changsha, China, [3] Longping Branch of Graduate College, Hunan University, Changsha, China, [4] College of Plant Protection, Yunnan Agricultural University, Kunming, China, [5] Institute of Alpine Economic Plants, Yunnan Academy of Agricultural Sciences, Lijiang, China, [6] Yunnan Academy of Tobacco Agricultural Sciences, Kunming, China

OPEN ACCESS

Edited by:
Xiaofei Cheng,
Northeast Agricultural University, China

Reviewed by:
Xiangdong Li,
Shandong Agricultural University, China
Ying Wang,
China Agricultural University, China

***Correspondence:**
Xiaohan Mo
xhmo@163.net
Fei Yan
yanfei@nbu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 01 February 2021
Accepted: 19 March 2021
Published: 15 April 2021

Citation:
Peng J, Bu S, Yin Y, Hua M, Zhao K, Lu Y, Zheng H, Wan Q, Zhang S, Chen H, Liu Y, Chen J, Mo X and Yan F (2021) Biological and Genetic Characterization of Pod Pepper Vein Yellows Virus-Associated RNA From Capsicum frutescens in Wenshan, China. Front. Microbiol. 12:662352. doi: 10.3389/fmicb.2021.662352

Tombusvirus-like associated RNAs (tlaRNAs) are positive-sense single-stranded RNAs found in plants co-infected with some viruses of the genus *Polerovirus*. Pod pepper vein yellows virus (PoPeVYV) was recently reported as a new recombinant polerovirus causing interveinal yellowing, stunting, and leaf rolling in *Capsicum frutescens* plants at Wenshan city, Yunnan province, China. The complete genome sequence of its associated RNA has now been determined by next-generation sequencing and reverse transcription (RT) polymerase chain reaction (PCR). PoPeVYV-associated RNA (PoPeVYVaRNA) (GenBank Accession No. MW323470) has 2970 nucleotides and is closely related to other group II tlaRNAs, particularly tobacco bushy top disease-associated RNA (TBTDaRNA, GenBank Accession No. EF529625). In infection experiments on *Nicotiana benthamiana* and *C. frutescens* plants, synergism between PoPeVYVaRNA and PoPeVYV was demonstrated, leading to severe interveinal yellowing of leaves and stunting of plants. The results provide further information on the genetic and biological properties of the various agents associated with pepper vein yellows disease (PeVYD).

Keywords: pod pepper vein yellows virus, tombusvirus-like associated RNA, *Polerovirus*, *Umbravirus*, recombination, biological characterization

INTRODUCTION

Tombusvirus-like associated RNAs (tlaRNAs) are often found in plants infected by some poleroviruses, including those that cause carrot motley dwarf disease, tobacco bushy top disease, and beet western yellows (Sanger et al., 1994; Mo et al., 2011; Campbell et al., 2020; Yoshida, 2020). TlaRNAs are single-stranded positive-sense RNAs of about 3 kb that encode two open reading frames (ORFs): ORF1a ends at an amber stop codon (UAG) and translational readthrough of this

codon results in a large protein ORF1b that contains amino acid motifs characteristic of viral polymerases. TlaRNAs lack a coat protein (CP) gene and depend on helper viruses of the genus *Polerovirus* for their encapsidation and transmission. The association of these RNAs with their poleroviruses facilitates movement and increases the accumulation of virus progeny within co-infected cells (Sanger et al., 1994; Syller, 2002; Mo et al., 2015; Yoshida, 2020). Phylogenetic analysis of the full-length genomes of tlaRNAs confirms their relationship to viruses in the genus *Tombusvirus* and that they can be classified into at least two distinct groups (Campbell et al., 2020). The tlaRNAs have a GGL amino acid triplet encoded by the nucleotides immediately following the amber stop codon and eight characteristic motifs of + ssRNA virus RdRps within the deduced amino acid sequences of ORF1b (Koonin, 1991). Notably, the ORF1b of all tlaRNAs has the GDD amino acid triplet characteristic of viral polymerases (Kamer and Argos, 1984).

Pepper vein yellows disease (PeVYD) is a major threat to pepper production in many different countries (Murakami et al., 2011; Dombrovsky et al., 2013; Knierim et al., 2013; Liu et al., 2016; Maina et al., 2016; Lotos et al., 2017). Pepper vein yellows viruses (PeVYVs) induce interveinal yellowing, stunting, and leaf rolling (Kamran et al., 2018). They are phloem-restricted viruses and are currently classified into six species within the genus *Polerovirus* (International Committee on Taxonomy of Viruses [ICTV] 2019 release)[1], named *Pepper vein yellows virus 1–6* (Murakami et al., 2011; Dombrovsky et al., 2013; Knierim et al., 2013; Liu et al., 2016; Maina et al., 2016; Lotos et al., 2017). A PeVYD outbreak on pod pepper (*Capsicum frutescens*) in Wenshan city, Yunnan province in 2019 was associated with a new recombinant polerovirus named pod PeVYV (PoPeVYV) (GenBank Accession No. MT188667). PoPeVYV is predicted to result from a single recombination event with PeVYV-3 as the major parent and the region 4126–5192 nt derived from TVDV as the minor parent. However, a full-length clone of PoPeVYV caused only symptomless infection in *Nicotiana benthamiana* and *C. frutescens* (Zhao et al., 2021).

In this study, we have identified a tlaRNA [PoPeVYV-associated RNA (PoPeVYVaRNA)] associated with PoPeVYV and belonging to Group II of tlaRNAs. This tlaRNA increases the titer of PoPeVYV and has destructive effects on plants. The genomic properties of PoPeVYVaRNA provide insights into the etiological roles of these agents in pod PeVYD (PoPeVYD).

MATERIALS AND METHODS

Sample Collection and RNA Extraction

In July 2019, 89 pepper (*C. frutescens*) samples were collected from three regions of Wenshan city. All the samples had typical viral symptoms of interveinal leaf yellowing and fruit discoloration (**Supplementary Figure 1**). Total RNA was extracted from fresh leaves/fruits using TRIzol™ Reagent (Invitrogen) in compliance with the manufacturer's instructions.

[1]https://talk.ictvonline.org/files/master-species-lists/m/msl/9601

Sequence and *de novo* Assembly

A total amount of 1 μg RNA was used as input material for the RNA sample preparations. The mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. RNA integrity was checked by Agilent 2100 Bioanalyzer (Agilent Technologies). The TruSeq RNA Sample Preparation Kit (Illumina) was used to construct cDNA libraries according to the manufacturer's instructions.

An Illumina NovaSeq 6000 platform with PE150 bp and CLC Genomics Workbench 20 (QIAGEN) was used for sequencing and data analysis. A total of 6,452,174 paired-end reads were obtained; 78,793 contigs (average length 579 bp) were generated *de novo* and compared with nucleotide and amino acid sequences in GenBank using BLASTn or BLASTx, respectively.

RT-PCR

Reverse transcription (RT) polymerase chain reaction (PCR) was performed using the ReverTra Ace™ qPCR RT Master Mix (Toyobo) and KOD-plus-Neo (Toyobo) following manufacturer's protocol. RT was performed with random primers at 42°C for 60 min. The cycling conditions for the subsequent PCR reaction were: 98°C 3 min, and then 30 cycles of 98°C for 30 s, 55°C for 90 s, 68°C for 1 kb/min; and 68°C for 10 min. RT-PCR products were purified, ligated into *pEASY®*-Blunt Zero Cloning vector (TRANS, China), and transformed into *Escherichia coli XL10* competent cells, and purified plasmids were sequenced.

5′ RACE and 3′ RACE

In order to obtain the full-length sequence of tlaRNA in pod pepper, part of the sequence was amplified using the primer pair tlaRNA-WS (**Supplementary Table 1**), and specific primers were designed for 5′ RACE R and 3′ RACE F. Then, the 5′ and 3′ RACE reactions were performed to obtain the complete 5′and 3′ terminal sequences. The 5′ and 3′ RACE reactions were performed as previously described (Zhao et al., 2021).

Phylogenetic and Sequence Analysis

Complete genome sequences of tlaRNAs were obtained from GenBank (**Supplementary Table 2**). Sequences were aligned using MUSCLE; the evolutionary history was inferred using the maximum likelihood (ML) method (Le and Gascuel, 2008). The best-fit nucleotide substitution model was determined to be ML (GTR + G) by MEGA X (Kumar et al., 2018). Evolutionary analyses were conducted in MEGA X with 1000 bootstrap replicates.

Plasmid Construction and Agroinfiltration

Reverse transcription PCR was performed using KOD-plus-Neo (Toyobo) and following the manufacturer's protocol. PCR products were purified with the Gel Extraction Kit (Omega). To generate infectious clones, the ClonExpress II One Step Cloning Kit (Vazyme) was used for homologous recombination. The full length tlaRNA was amplified with primer pair (Inf-tlaRNA)

and recombined with the linearized binary vector pCB301-MD (Zhao et al., 2021).

To confirm infectivity, the infectious clone (pCB-PoPeVYVaRNA) was transformed into *Agrobacterium tumefaciens* (GV3101) which was mixed inoculation with pCB-PoPeVYV and delivered to *N. benthamiana* (Zhao et al., 2021). The tissue was harvested 15–28 days post inoculation (dpi). RT-PCR detection was done as described before (Murakami et al., 2011).

RT-qPCR

Quantitative RT-PCR was used to determine whether the presence of the tlaRNA affects the accumulation levels of PoPeVYV. Fold changes in accumulation of each component were determined using the relative quantification method and normalized to the mean values of those at 28 dpi. For relative quantification of each RNA, the UBC gene of *N. benthamiana* was selected as an internal control.

Virion Purification

Virions were purified from plants using procedures developed previously (Mo et al., 2010) with some modifications. Virus-infected leaf tissues (250 g) were harvested 28 days post infiltration and homogenized in 500 mL of extraction buffer [0.1 M sodium phosphate buffer; 0.5% (w/v) cellulase; 0.5% (w/v) pectinase; 0.1% (v/v) sodium azide; 0.5% (v/v) β-mercaptoethanol, pH 6.0]. The homogenate was stirred at 25°C for 5 h and emulsified in a mixture of equal volumes of chloroform and 1-butanol. The emulsion was broken by centrifugation at 10,000 *g* for 15 min. Then, Triton X-100 was added to the upper aqueous phase to a final concentration of 1% (v/v) and stirred gently for 30 min. After addition of 8% PEG6000 (w/v) and 0.4 M NaCl, the mixture was stirred gently for 1 h at room temperature, then kept at 4°C for 2 h, and centrifuged at 8000 *g* for 15 min. The resultant pellet was suspended in 50 mL of storage buffer (0.1 M sodium phosphate, pH 7.0) and clarified by centrifugation at 5000 *g* for 15 min. The suspension was concentrated and purified by centrifugation at 70,000 *g* for 4 h through a 30% sucrose cushion. After centrifugation, the pellet was suspended in 1 mL of storage buffer. To exclude any free viral RNA, the virion preparation was digested by RNase at 37°C for 10 min.

Transmission by Aphids

Virus-free aphids (*Myzus persicae*) were reared from newly born ones. Approximately 100 apterous aphids (3–4 days old nymphs) were transferred using a paintbrush from the virus-free stock plants to a 50 mL centrifuge tube for a starvation period of 1 h. The aphids were then transferred to a cylindrical Perspex cage and allowed to feed on aqueous 20% sucrose solution containing PoPeVYV virions, PoPeVYV + tlaRNA virions or sucrose solution (control). After 24 h, RT-PCR was used to detect virion in aphids (10 per treatment). Then, aphids (30 per treatment) were released onto disease-free pod peppers (six plants per treatment) and kept under controlled environmental conditions (∼25°C, 60% relative humidity, and a 14-h photoperiod). RNA was extracted from the new leaves of these plants after 45 days to test for the presence of viral RNA.

RESULTS

Sequence Comparison and Phylogenetic Analysis of PoPeVYV-Associated RNA From Wenshan City

Symptoms of interveinal leaf yellowing resembling those caused by viruses were observed in pod pepper fields throughout Wenshan, China and, as we previously reported, a new recombinant polerovirus, PoPeVYV, was identified in 16 of 58 symptomatic samples (Zhao et al., 2021). To assess whether an associated RNA was also present, these leaves were mixed into a pooled sample and sent for next-generation RNA-Seq sequencing (NGS). A large contig of 2915 nt was detected that had the highest nucleotide identity (83.6%) to tobacco bushy top disease-associated RNA (TBTDaRNA; GenBank: EF529625). To confirm our sequencing data, we used primers AR3F and AR5R to amplify a fragment of the tlaRNA by RT-PCR (Campbell et al., 2020). Fragments with a predicted size of approximately 650 bp were obtained in eight of the 16 PoPeVYV-infected samples. Following 5′ and 3′ RACE analysis, we determined the complete sequence of the tlaRNAs, which were identical in sequence and 2970 nt long (GenBank accession number: MW323470). We tentatively designated the isolated RNA as PoPeVYVaRNA (**Figure 1A**).

The full-length genome sequence of PoPeVYVaRNA had nucleotide identities of 85.5, 76.0, and 65.2% with TBTDaRNA, PeVYVaRNA-PRO54353, and BWYVaRNA-ST9, respectively (**Table 1**). It had the predicted two ORFs (ORF1a and readthrough protein ORF1b) characteristic of other tlaRNAs, a short 5′ non-coding region of 8 nt preceding the start of ORF1a and a long 3′ non-coding region of 595 nt (**Figure 1A**).

To investigate the relationships between PoPeVYVaRNA and other tlaRNAs, the full genome sequences of 13 tlaRNAs were retrieved from GenBank and an ML phylogenetic tree was inferred (**Figure 1B**). PoPeVYVaRNA clustered into Group II with BWYVaRNA-st9, TuYVaRNA-JKI29345, PeVYVaRNA-PRO54353, and TBTDaRNA (**Figure 1B**). Alignment analysis also showed that PoPeVYVaRNA was most closely related to TBTDaRNA (**Table 1**). The predicted amino acid sequence of PoPeVYVaRNA ORF1b contained the eight characteristic motifs of + ssRNA virus RdRps (**Supplementary Figure 2**) (Koonin, 1991; Campbell et al., 2020).

PoPeVYV Induces Typical Viral Symptoms in *N. benthamiana* by Co-infection With PoPeVYV-Associated RNA

To examine the effects of PoPeVYVaRNA on the symptoms caused in mixed infections with PoPeVYV, *N. benthamiana* plants were inoculated by infiltrating their leaves with *A. tumefaciens* harboring different virus–RNA combinations: SI, singly infected with the virus PoPeVYV; SIa, singly infected with PoPeVYVaRNA; MI, mixed infection of

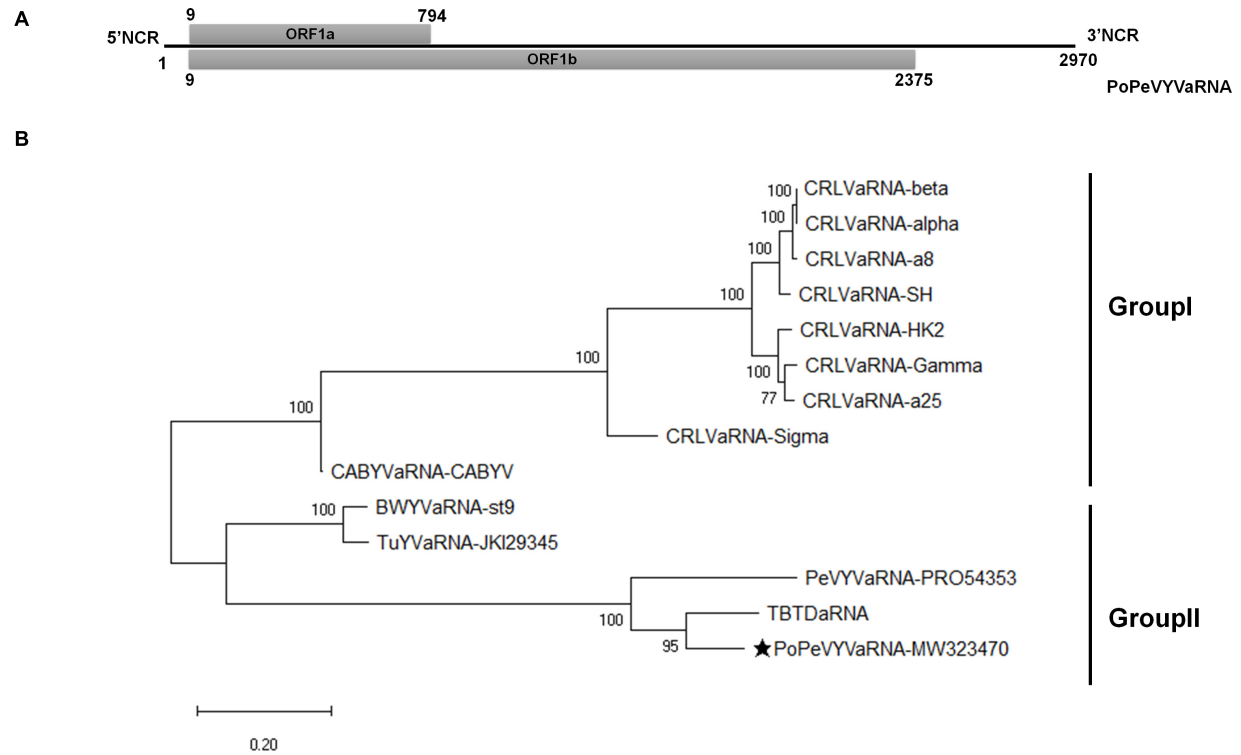


FIGURE 1 | Genome organization and phylogenetic analysis of PoPeVYVaRNA. **(A)** Genome organization of PoPeVYVaRNA. **(B)** Maximum likelihood phylogenetic tree constructed using MEGA X, showing the relationship of PoPeVYVaRNA to other tlaRNAs using their nt sequences. Numbers on branches are bootstrap support values (1000 replicates). Multiple nucleotide sequences were aligned using MUSCLE, the best model (GTR + G), and subsequent analysis was determined by MEGAX. The tlaRNA abbreviations and accession numbers are described in **Supplementary Table 2**.

TABLE 1 | Comparisons (nucleotide/amino acid identity, %) between the genome of PoPeVYVaRNA and closely related RNAs.

		5′NCR	ORF1[a]		ORF1[b]		RTD		3′NCR	Full genome
		nt	aa	nt	aa	nt	aa	nt	nt	nt
Group II	PoPeVYVaRNA-MW323470	100	100	100	100	100	100	100	100	100
	BWYVaRNA-st9	85.7	49.0	64.0	64.0	67.0	69.5	68.2	57.0	65.2
	TuYVaRNA-JKI29345	85.7	50.5	63.7	63.9	66.6	69.0	67.7	57.3	65.0
	PeVYVaRNA-PRO54353	62.5	63.6	69.7	77.5	76.9	84.2	80.5	73.1	76.0
	TBTDaRNA	**100**	**81.5**	**81.0**	**88.6**	**85.3**	**92.9**	**87.4**	**86.1**	**85.5**
Group I	CRLVaRNA-Sigma	87.5	35.0	58.3	48.9	60.1	54.2	60.9	67.3	59.8

[a] *NCR = non-coding region, aa = amino acid, and nt = nucleotide. Highest percentages are underlined and in bold.*
[b] *PoPeVYVaRNA (GenBank, MW323470), BWYVaRNA-st9 (GenBank, L04281), PeVYVaRNA-PRO54353 (GenBank, MT321510), TBTDaRNA (GenBank, EF529625), CRLVaRNA-Sigma (GenBank, KM486093), and TuYVaRNA-JKI29345 (GenBank, MK450521).*

PoPeVYV + PoPeVYVaRNA; CK, control with no virus or viral RNA. There were no symptoms in any SI, SIa, or CK plants 14 dpi, whereas at 28 dpi, MI-inoculated plants had typical viral symptoms of interveinal leaf yellowing and plants were stunted (**Figure 2A**).

Reverse transcription PCR using primer pair (qPoPeVYV-P3) (**Supplementary Table 1**) to detect the PoPeVYV in systemic leaves showed that viral RNA was present and had spread systemically in SI- and MI-inoculated plants, but not in the controls or SIa-inoculated plants. RT-PCR also showed that PoPeVYVaRNA had spread systemically in MI-inoculated plants and that it was encapsidated in virions purified from MI plants (**Figure 2B** and **Table 2**). Quantitative RT-PCR indicated that the level of PoPeVYV RNA in plants inoculated with MI was more than 15-fold that in plants inoculated with SI (**Figures 2C,D**).

To investigate the mechanism of the synergism, mutant infectious clones of PoPeVYVaRNA were constructed that abolished expression of one or both ORFs (*PoPeVYVaRNA-orf1a*, *-orf1b*, and *-orf1ab*). These were used in co-infection experiments with PoPeVYV (six plants per treatment, pCB301-MD as control) (**Figure 3A**). Quantitative RT-PCR indicated that the level of PoPeVYV RNA in local leaves

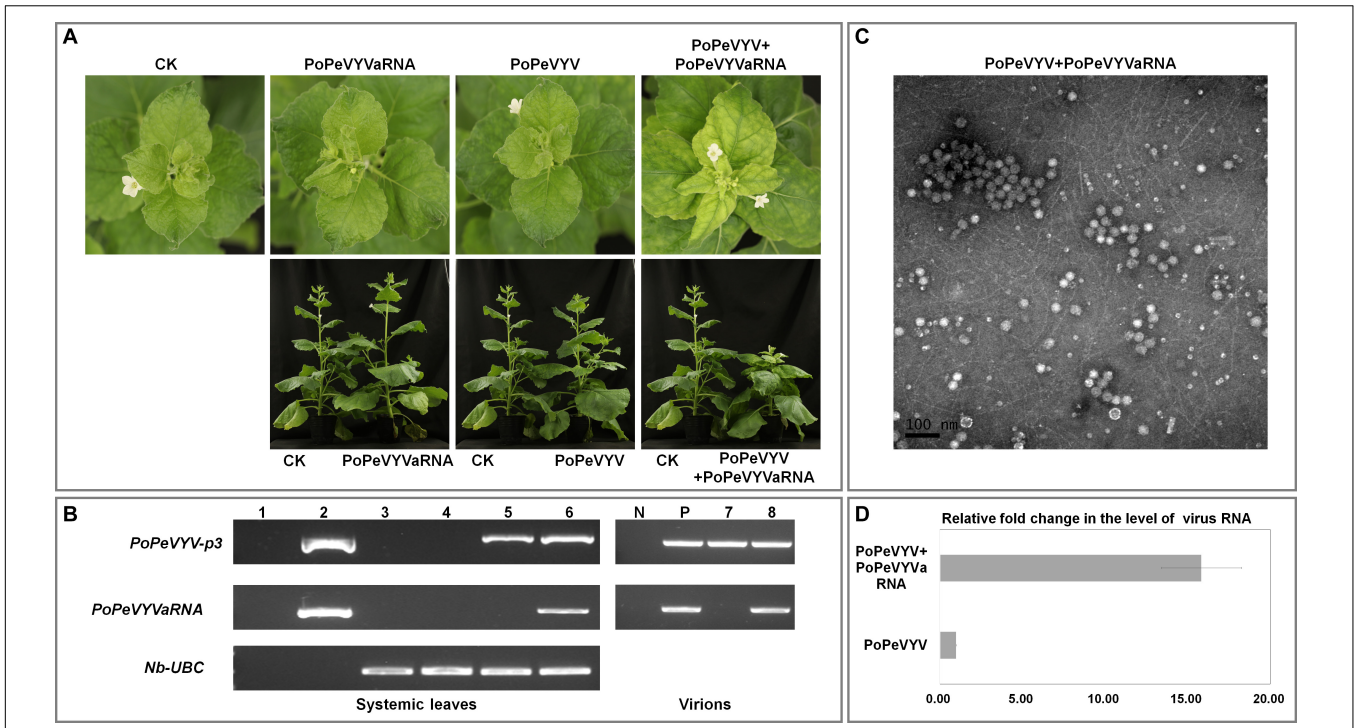


FIGURE 2 | Symptoms caused by PoPeVYV in *N. benthamiana* co-infected with PoPeVYVaRNA. **(A)** Phenotype of *N. benthamiana* plants agroinfiltrated with viral infectious clone combinations or empty agrobacterium (CK) at 28 days post infiltration. **(B)** RT-PCR confirming the presence of viral RNAs in systemic leaves of inoculated plants and virions. 1, negative control; 2, positive control; 3, CK; 4, PoPeVYVaRNA; 5, PoPeVYV; 6, PoPeVYV + PoPeVYVaRNA; N, negative control; P, positive control; 7, virion from PoPeVYV-infected plant; 8, virion from PoPeVYV + PoPeVYVaRNA infected plant. **(C)** Virions purified from leaves infected by the PoPeVYV + PoPeVYVaRNA infectious clones and observed by TEM. Bars represent 100 nm. **(D)** Relative fold changes of PoPeVYV in systemically infected leaves of *N. benthamiana* inoculated with PoPeVYV or PoPeVYV + PoPeVYVaRNA, as shown by quantitative real-time reverse transcription PCR. The means (±*SE*) were calculated from the RNA levels of 12 individual plants at 28 days post inoculation.

TABLE 2 | Systemic infection of two host plant species following agroinoculation or aphid transmission with different virus–RNA combinations.

Virus–RNA combinations	*Nicotiana benthamiana*[1]	*Myzus persicae*[2]	*Capsicum frutescens*[3]
PoPeVYV	12/12	10/10	2/6
	12/12	10/10	1/6
	12/12	10/10	3/6
PoPeVYVaRNA	0/12	–	–
	0/12	–	–
	0/12	–	–
PoPeVYV + PoPeVYVaRNA	12/12	10/10	5/6
	12/12	10/10	4/6
	12/12	10/10	6/6

[1]*Plants were tested by RT-PCR at 28 days post inoculation.*
[2]*Aphids were tested by RT-PCR at 12 h post virus acquisition.*
[3]*Plants were tested by RT-PCR at 45 days post inoculation.*

co-inoculated with PoPeVYVaRNA or *PoPeVYVaRNA-orf1a* was more than 6- or 2.9-fold that in leaves co-inoculated with pCB301-MD. Co-inoculation with *PoPeVYVaRNA-orf1b or PoPeVYVaRNA-orf1ab* did not significantly affect the level of PoPeVYV RNA (**Figure 3B**). RT-PCR showed that viral RNA was present and had spread systemically in all the inoculated plants, but that mutants of PoPeVYVaRNA did not spread systemically in co-infections with PoPeVYV (**Figure 3C**). There were leaf rolling symptoms only in plants co-infected with PoPeVYV and PoPeVYVaRNA at 14 dpi (**Figure 3C**).

Interveinal Yellowing Symptoms Are Caused by Co-infection With PoPeVYV and PoPeVYV-Associated RNA in *C. frutescens*

Aphid transmission was used to examine the biological significance of PoPeVYVaRNA in *C. frutescens*. The newly-emerged leaves of plants inoculated with aphids fed only on PoPeVYV virions (SI) had mild interveinal symptoms after 45 days, but when the aphids were fed on a mixture of PoPeVYV and the tlaRNA (MI), the symptoms were much more severe (**Figure 4A**). RT-PCR showed that the tlaRNA had spread systemically in the MI-treated plants (**Figure 4B** and **Table 2**). Quantitative RT-PCR indicated that the level of PoPeVYV RNA in plants transmitted with MI was more than 7.9-fold that in plants transmitted with SI (**Figure 4C**). RT-PCR indicated that all the aphids fed on SI or MI acquired virus (10/10), but the virus

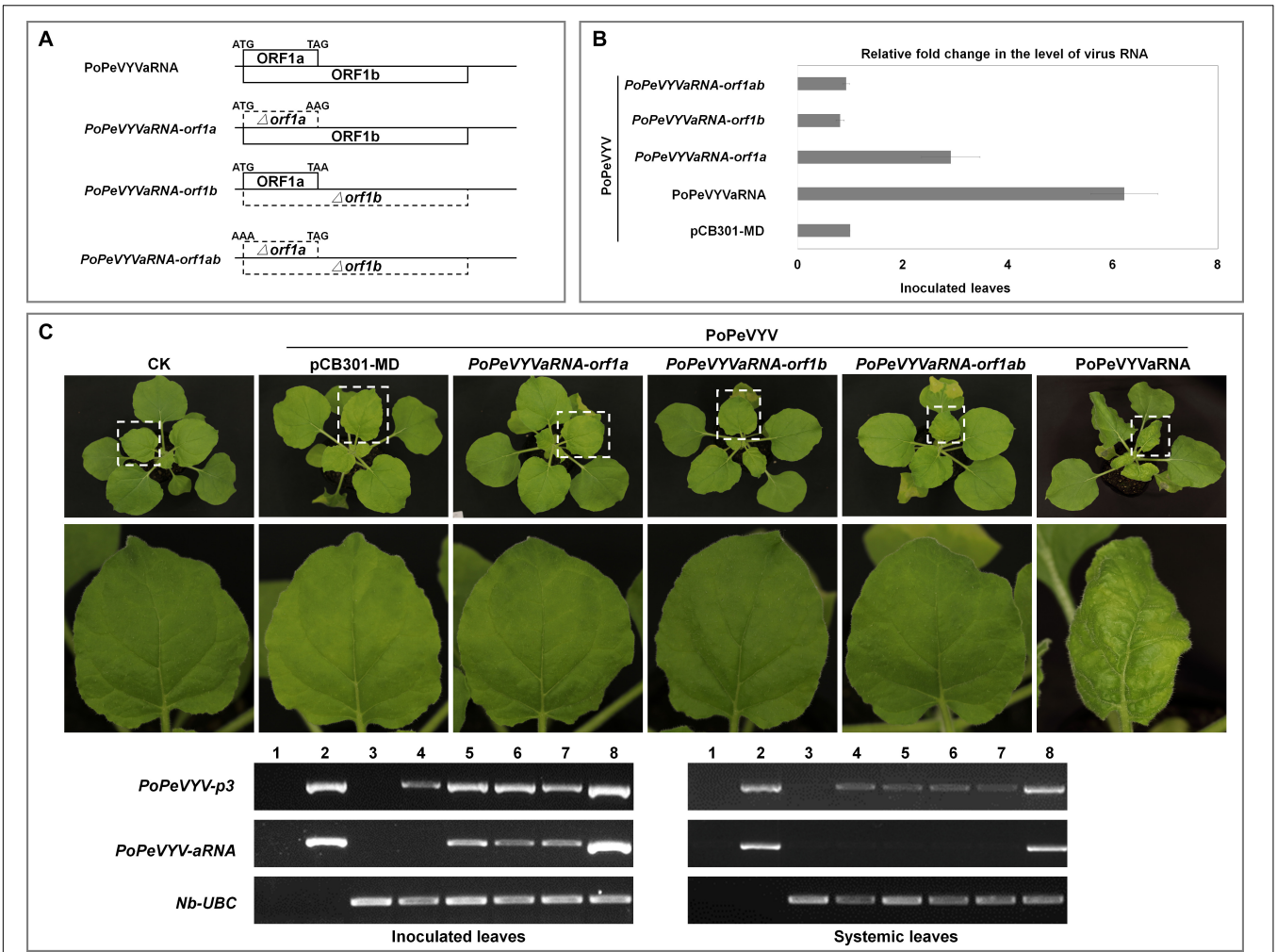


FIGURE 3 | PoPeVYVaRNA is responsible for the synergism with PoPeVYV in *N. benthamiana*. **(A)** The PoPeVYVaRNA mutant infectious clones prepared. **(B)** Relative fold changes of PoPeVYV in local leaves of *N. benthamiana* co-inoculated with PoPeVYVaRNA or mutants of PoPeVYVaRNA (pCB301-MD as control), as shown by quantitative real-time reverse transcription PCR. The means (±*SE*) were calculated from the RNA levels of six individual plants at 3 days post inoculation. **(C)** Phenotype of *N. benthamiana* plants agroinfiltrated with viral infectious clone combinations or empty agrobacterium (CK) at 14 days post infiltration. RT-PCR confirming the presence of viral RNAs in local/systemic leaves of inoculated plants. 1, negative control; 2, positive control; 3, CK; 4, PoPeVYV + pCB301-MD; 5, PoPeVYV + *PoPeVYVaRNA-orf1a*; 6, PoPeVYV + *PoPeVYVaRNA-orf1b*; 7, PoPeVYV + *PoPeVYVaRNA-orf1ab*; 8, PoPeVYV + PoPeVYVaRNA.

transmission rate by the aphids was very different at, respectively, 17–50 and 67–100% (**Table 2**).

DISCUSSION

In this study, we have identified a tombusvirus-like RNA associated with PoPeVYV. tlaRNAs are found in plants co-infected with several viruses in the genus *Polerovirus*. All the tlaRNAs have a very short non-coding region preceding ORF1a at the 5′ end, which encodes a putative product of 25.1–29.3 kDa. Readthrough of the ORF1a amber termination codon allows expression of an 84.6–89.0 kDa protein (ORF1b). The genetic properties of tlaRNAs are similar to viruses in the genus *Umbravirus*, but the conserved polymerase is interrupted by readthrough of the ORF1a amber termination codon instead of slightly overlapping the end (−1 frameshift). Umbraviruses also have a movement protein (MP) that enables them to spread very efficiently within infected plants, but tlaRNAs do not (Ryabov et al., 1998; Campbell et al., 2020). Despite these similarities with umbraviruses and the presence of distinct phylogenetic clades, tlaRNAs have never been formally classified to genera (Lefkowitz et al., 2018; Campbell et al., 2020).

Viral synergism is caused by co-infection of two unrelated viruses, leading to more severe symptoms. Synergistic infection of phloem-restricted poleroviruses and umbraviruses has destructive effects on crop plants and has been well studied (Yoo et al., 2017; Zhou et al., 2017; Yoshida, 2020). Only a few RNAs associated with poleroviruses have been reported, and synergism between them has often been overlooked in the past. It has been shown that several tlaRNAs stimulate the titers of the poleroviruses and enhance the disease symptoms

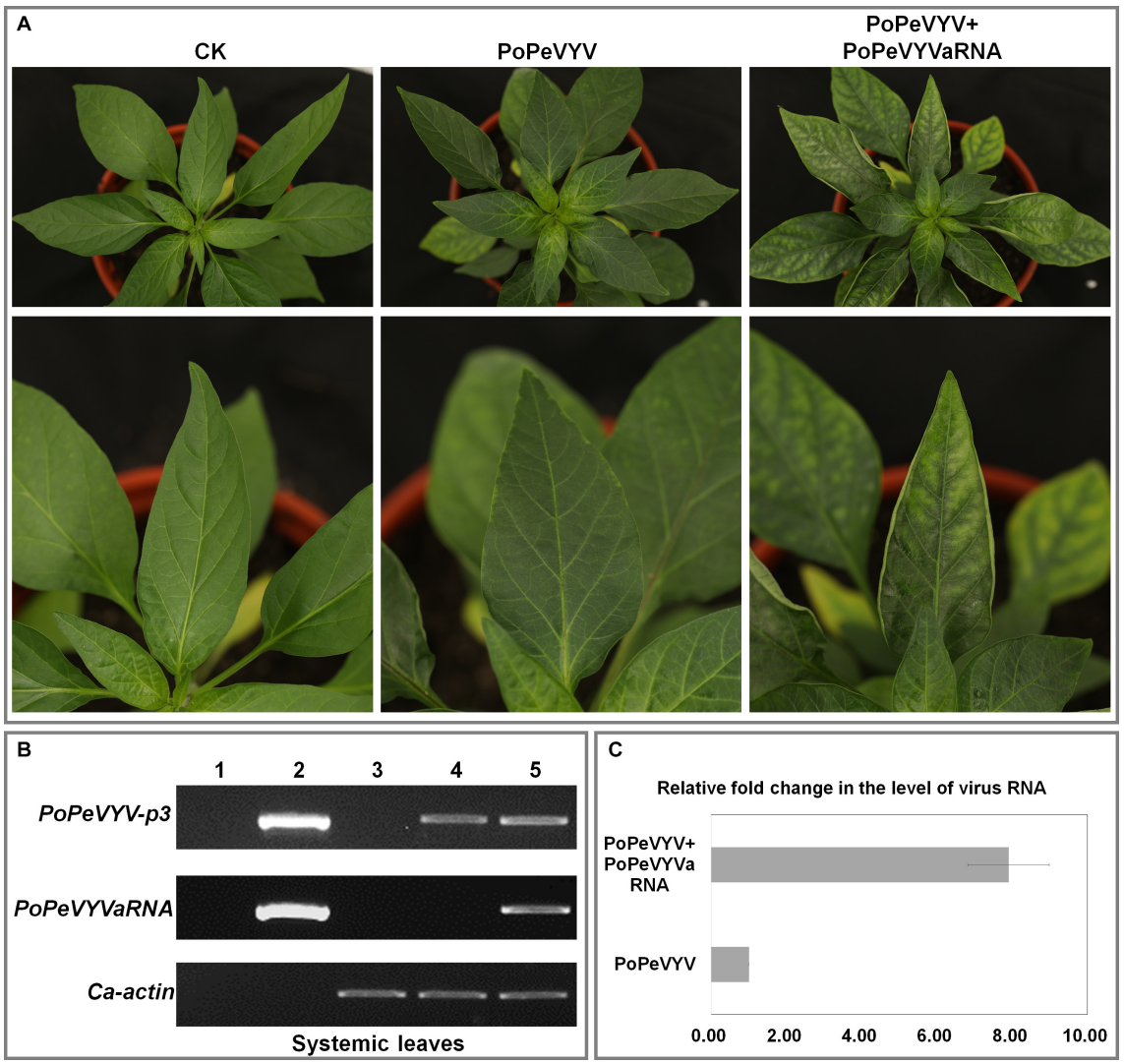


FIGURE 4 | Symptoms caused by PoPeVYV in *Capsicum frutescens* co-infected with PoPeVYVaRNA. **(A)** Phenotype of *C. frutescens* plants infected by aphids with different combinations or empty agrobacterium (CK) 45 days post infiltration. **(B)** RT-PCR confirming the presence of viral RNAs in systemic leaves of infected plants. 1, Negative control; 2, positive control; 3, CK; 4, PoPeVYV; 5, PoPeVYV + PoPeVYVaRNA. **(C)** Relative fold changes of PoPeVYV in systemically infected leaves of *C. frutescens* inoculated with PoPeVYV or PoPeVYV + PoPeVYVaRNA, as shown by quantitative real-time reverse transcription PCR. The means (±*SE*) were calculated from the RNA levels of three individual plants at 45 days post inoculation.

in plants co-infected with their respective poleroviruses (Sanger et al., 1994; Mo et al., 2015; Yoshida, 2020). In this study, we have now also shown synergism between PoPeVYV and its associated RNA (PoPeVYVaRNA) with increased viral titers and symptom severity consistent with field observations (**Figures 2**, **3** and **Supplementary Figure 1**). Earlier studies showed that TBTDaRNA could be detected by RT-PCR in 16 of 17 TBTD-affected samples collected from different locations in Yunnan province, showing that TBTDaRNA is a normal component of the tobacco bushy top complex in China (Mo et al., 2011). However, PoPeVYVaRNA was detected by RT-PCR in only eight of 16 PoPeVYV-infected samples, and tlaRNAs do not appear to be essential components of the infections by other PeVYV complexes in fields in Wenshan city (data not shown). This apparent difference between tlaRNAs in their biological effects needs to be examined further.

The plants from the fields described here were infected with various viruses (PeVYVs, ChiVMV, ChiRSV, CMV, etc.), and the severe viral symptoms of PeVYD in the field may therefore be a complicated synergistic effect of mixed infection (Cheng et al., 2011; Laprom et al., 2019; Zhao et al., 2021). The results of this study indicate that one factor affecting PeVYD symptoms is likely to be the co-infection of PoPeVYV and PoPeVYVaRNA.

In conclusion, PoPeVYVaRNA is an associated RNA that depends upon co-infection and encapsidation with PoPeVYV for its systemic movement.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

XM, JP, and FY conceived and designed the experiments. HZ, QW, SZ, and YLiu collected the samples. SB, KZ, and MH performed the experiments. YLu and JC analyzed the data. JP, SB, XM, and FY wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was financially supported by the Chinese Agriculture Research System (CARS-24-C-04). This work also was supported by grant-in-aid from State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agroproducts (KF20190107) and K. C. Wong education foundation.

ACKNOWLEDGMENTS

We thank Dr. Mike Adams for manuscript correction.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.662352/full#supplementary-material

Supplementary Figure 1 | Symptoms of virus-infected pod peppers from the field.

Supplementary Figure 2 | The deduced amino acid sequences of ORF1b with eight characteristic motifs of + ssRNA virus RdRps in Group II.

Supplementary Table 1 | Primers used in this study.

Supplementary Table 2 | Accession numbers of tombusvirus-like associated RNAs used for phylogenetic analysis.

REFERENCES

Campbell, A. J., Erickson, A., Pellerin, E., Salem, N., Mo, X., Falk, B. W., et al. (2020). Phylogenetic classification of a group of self-replicating RNAs that are common in co-infections with poleroviruses. *Virus Res.* 276:197831. doi: 10.1016/j.virusres.2019.197831

Cheng, Y. H., Deng, T. C., Chen, C. C., Liao, J. Y., Chang, C. A., and Chiang, C. H. (2011). First report of pepper mottle virus in bell pepper in Taiwan. *Plant Dis.* 95:617. doi: 10.1094/pdis-10-10-0721

Dombrovsky, A., Glanz, E., Lachman, O., Sela, N., Doron-Faigenboim, A., and Antignus, Y. (2013). The complete genomic sequence of pepper yellow leaf curl virus (PYLCV) and its implications for our understanding of evolution dynamics in the genus polerovirus. *PLoS One* 8:e70722. doi: 10.1371/journal.pone.0070722

Kamer, G., and Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* 12, 7269–7282. doi: 10.1093/nar/12.18.7269

Kamran, A., Lotos, L., Amer, M. A., Al-Saleh, M. A., Alshahwan, I. M., Shakeel, M. T., et al. (2018). Characterization of pepper leafroll chlorosis virus, a new polerovirus causing yellowing disease of bell pepper in Saudi Arabia. *Plant Dis.* 102, 318–326. doi: 10.1094/pdis-03-17-0418-re

Knierim, D., Tsai, W. S., and Kenyon, L. (2013). Analysis of sequences from field samples reveals the presence of the recently described pepper vein yellows virus (genus Polerovirus) in six additional countries. *Arch. Virol.* 158, 1337–1341. doi: 10.1007/s00705-012-1598-y

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* 72, 2197–2206. doi: 10.1099/0022-1317-72-9-2197

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Laprom, A., Nilthong, S., and Chukeatirote, E. (2019). Incidence of viruses infecting pepper in Thailand. *Biomol. Concepts* 10, 184–193. doi: 10.1515/bmc-2019-0021

Le, S. Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320. doi: 10.1093/molbev/msn067

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* 46, D708–D717. doi: 10.1093/nar/gkx932

Liu, M., Liu, X., Li, X., Zhang, D., Dai, L., and Tang, Q. (2016). Complete genome sequence of a Chinese isolate of pepper vein yellows virus and evolutionary analysis based on the CP, MP and RdRp coding regions. *Arch. Virol.* 161, 677–683. doi: 10.1007/s00705-015-2691-9

Lotos, L., Olmos, A., Orfanidou, C., Efthimiou, K., Avgelis, A., Katis, N. I., et al. (2017). Insights into the etiology of polerovirus-induced pepper yellows disease. *Phytopathology* 107, 1567–1576. doi: 10.1094/PHYTO-07-16-0254-R

Maina, S., Edwards, O. R., and Jones, R. A. (2016). First complete genome sequence of pepper vein yellows virus from Australia. *Genome Announc.* 4, 450–416. doi: 10.1128/genomeA.00450-16

Mo, X. H., Chen, Z. B., and Chen, J. P. (2010). Complete nucleotide sequence and genome organization of a Chinese isolate of tobacco vein distorting virus. *Virus Genes* 41, 425–431. doi: 10.1007/s11262-010-0524-1

Mo, X. H., Chen, Z. B., and Chen, J. P. (2011). Molecular identification and phylogenetic analysis of a viral RNA associated with the Chinese tobacco bushy top disease complex. *Ann. Appl. Biol.* 158, 188–193. doi: 10.1111/j.1744-7348.2010.00452.x

Mo, X. H., Xu, P., Zhao, X. N., Zhang, L. F., Qin, X.Y., and Xia, Z. Y. (2015). "The interactions between tobacco vein distorting virus and tobacco bushy top disease-associated RNA," in *Proceedings of the 2015 CORESTA Meeting, Agronomy/Phytopathology* (Izmir: CORESTA). Available online at: https://www.coresta.org/abstracts/interactions-between-tobacco-vein-distorting-virus-and-tobacco-bushy-top-disease

Murakami, R., Nakashima, N., Hinomoto, N., Kawano, S., and Toyosato, T. (2011). The genome sequence of pepper vein yellows virus (family Luteoviridae, genus Polerovirus). *Arch. Virol.* 156, 921–923. doi: 10.1007/s00705-011-0956-5

Ryabov, E. V., Oparka, K. J., Santa Cruz, S., Robinson, D. J., and Taliansky, M. E. (1998). Intracellular location of two groundnut rosette umbravirus proteins delivered by PVX and TMV vectors. *Virology* 242, 303–313. doi: 10.1006/viro.1997.9025

Sanger, M., Passmore, B., Falk, B. W., Bruening, G., Ding, B., and Lucas, W. J. (1994). Symptom severity of beet western yellows virus-strain ST9 is comferred by the ST9-associated RNA and is not associated with virus release from the phloem. *Virology* 200, 48–55. doi: 10.1006/viro.1994.1161

Syller, J. (2002). Umbraviruses: The unique plant viruses that do not encode a capsid protein. *Acta Microbiol. Polon.* 51, 99–113.

Yoo, R. H., Lee, S. W., Lim, S., Zhao, F., Igori, D., Baek, D., et al. (2017). Complete genome analysis of a novel umbravirus-polerovirus combination isolated from Ixeridium dentatum. *Arch. Virol.* 162, 3893–3897. doi: 10.1007/s00705-017-3512-0

Yoshida, N. (2020). Biological and genetic characterization of carrot red leaf virus and its associated virus/RNA isolated from carrots in Hokkaido, Japan. *Plant Pathol.* 69, 1379–1389. doi: 10.1111/ppa.13202

Zhao, K., Yin, Y., Hua, M., Wang, S., Mo, X., Yuan, E., et al. (2021). Pod pepper vein yellows virus, a new recombinant polerovirus infecting Capsicum frutescens in Yunnan province, China. *Virol. J.* 18:42. doi: 10.1186/s12985-021-01511-5

Zhou, C. J., Zhang, X. Y., Liu, S. Y., Wang, Y., Li, D. W., Yu, J. L., et al. (2017). Synergistic infection of BrYV and PEMV 2 increases the accumulations of both BrYV and BrYV-derived siRNAs in Nicotiana benthamiana. *Sci. Rep.* 7:45132. doi: 10.1038/srep45132

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 May 2021
doi: 10.3389/fmicb.2021.673218

A Survey Using High-Throughput Sequencing Suggests That the Diversity of Cereal and Barley Yellow Dwarf Viruses Is Underestimated

Merike Sõmera[1], Sébastien Massart[2], Lucie Tamisier[2], Pille Sooväli[3], Kanitha Sathees[4] and Anders Kvarnheden[4,5]*

[1] Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia, [2] Laboratory of Integrated and Urban Phytopathology, Gembloux Agro-Bio Tech – University of Liège, Gembloux, Belgium, [3] Department of Plant Protection, Estonian Crop Research Institute, Jõgeva, Estonia, [4] Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden, [5] Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Tartu, Estonia

OPEN ACCESS

Edited by:
Xifeng Wang, State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection (CAAS), China

Reviewed by:
W. Allen Miller, Iowa State University, United States
Peipei Zhang, Langfang Normal University, China

***Correspondence:**
Merike Sõmera
merike.somera@taltech.ee

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: *01 March 2021*
Accepted: *06 April 2021*
Published: *11 May 2021*

Citation:
Sõmera M, Massart S, Tamisier L, Sooväli P, Sathees K and Kvarnheden A (2021) A Survey Using High-Throughput Sequencing Suggests That the Diversity of Cereal and Barley Yellow Dwarf Viruses Is Underestimated. Front. Microbiol. 12:673218. doi: 10.3389/fmicb.2021.673218

Worldwide, barley/cereal yellow dwarf viruses (YDVs) are the most widespread and damaging group of cereal viruses. In this study, we applied high-throughput sequencing technologies (HTS) to perform a virus survey on symptomatic plants from 47 cereal fields in Estonia. HTS allowed the assembly of complete genome sequences for 22 isolates of cereal yellow dwarf virus RPS, barley yellow dwarf virus GAV, barley yellow dwarf virus PAS (BYDV-PAS), barley yellow dwarf virus PAV (BYDV-PAV), and barley yellow dwarf virus OYV (BYDV-OYV). We also assembled a near-complete genome of the putative novel species BYDV-OYV from Swedish samples of meadow fescue. Previously, partial sequencing of the central part of the coat protein gene indicated that BYDV-OYV represented a putative new species closely related to BYDV-PAV-CN, which currently is recognized as a subtype of BYDV-PAV. The present study found that whereas the 3′gene block of BYDV-OYV shares the closest relationship with BYDV-PAV-CN, the 5′gene block of BYDV-OYV shows the closest relationships to that of BYDV-PAS. Recombination detection analysis revealed that BYDV-OYV is a parental virus for both. Analysis of complete genome sequence data indicates that both BYDV-OYV and BYDV-PAV-CN meet the species criteria of genus *Luteovirus*. The study discusses BYDV phylogeny, and through a systematic *in silico* analysis of published primers for YDV detection, the existing gaps in current diagnostic practices for detection of YDVs, proposing primer pairs based on the most recent genomic information for the detection of different BYDV species. Thanks to the rising number of sequences available in databases, continuous updating of diagnostic primers can improve test specificity, e.g., inclusivity and exclusivity at species levels. This is needed to properly survey the geographical and host distribution of the different species of the YDV complex and their prevalence in cereal/barley yellow dwarf disease epidemics.

Keywords: HTS, *Luteovirus*, BYDV, CYDV, diagnostics, wheat, epidemiology, OYV

INTRODUCTION

Yellow dwarf viruses (YDVs) of the family *Luteoviridae* constitute a complex of ssRNA viruses that are the most widespread group of cereal viruses worldwide (a list of the countries with recorded entries can be found at https://www.cabi.org/isc/datasheet/10539). YDVs are transmitted by more than 25 aphid species in a persistent non-propagative manner (Halbert and Voegtlin, 1995). Perennial grasses play an important role in the epidemiology of YDVs as more than 150 potential reservoir host species have been recorded in the family Poaceae (Gramineae; D'Arcy, 1995). Yield losses of 13–45 kg ha^{-1} for each 1% increase in YDV incidence are reported, ranging up to 80% of a total yield. The actual losses are dependent on symptom severity, which depends on the particular YDV species infecting the crop plants and on the varying transmission efficiency of YDVs by different vector species (Van den Eynde et al., 2020).

The aphid transmission properties of polero- and luteoviruses are determined by their capsids, which share a common evolutionary origin (Martin et al., 1990). Currently, in the genus *Polerovirus*, viruses of five recognized species have been found infecting cereals or grasses: *Cereal yellow dwarf virus RPV*, *Cereal yellow dwarf virus RPS* (CYDV-RPS), *Maize yellow dwarf virus RMV* (MYDV-RMV), *Maize yellow mosaic virus* (MaYMV), and *Sugarcane yellow leaf virus*. In addition, recent reports have indicated two novel tentative members of the genus *Polerovirus*: Barley virus G (BVG; Zhao et al., 2016) and Wheat leaf yellowing-associated virus (Zhang et al., 2017). In the genus *Luteovirus*, there are five species with members infecting cereals or grasses: *Barley yellow dwarf virus PAV* (BYDV-PAV), *Barley yellow dwarf virus PAS* (BYDV-PAS), *Barley yellow dwarf virus MAV*, *Barley yellow dwarf virus kerII,* and *Barley yellow dwarf virus kerIII*. Currently, Barley yellow dwarf virus GAV (BYDV-GAV) is considered as a subspecies of *Barley yellow dwarf virus MAV*, and Barley yellow dwarf virus PAV-CN is considered as a subspecies of *Barley yellow dwarf virus PAV*. Partial sequencing of the coat protein (CP) gene suggested a putative novel cereal-infecting luteovirus Barley yellow dwarf virus OYV (BYDV-OYV), whose taxonomic status remained unclear because of the lack of a complete genome sequence (Bisnieks et al., 2004). Two established species, Barley yellow dwarf virus GPV (BYDV-GPV) and Barley yellow dwarf virus SGV (BYDV-SGV), have not been assigned to either genus yet.

In the current study, we identified and sequenced the genome of BYDV-OYV for isolates from spring wheat and oat samples collected in Estonia and from meadow fescue samples collected in Sweden. In addition to BYDV-OYV, we also detected BYDV-PAS, BYDV-PAV, BYDV-GAV, and CYDV-RPS in a field survey in Estonia using high-throughput sequencing (HTS). Interestingly, half of these species have been reported seldomly. Combining the genome sequences generated during this study and the available sequences in the database, we propose also new diagnostic primers for specific detection of the BYDV different species.

MATERIALS AND METHODS

Plant Material

In 2012–2015, the leaf samples of cereal plants (wheat, barley, oat, rye, and triticale) showing chlorotic mottle or stripes, yellowing, reddening, stunting, or other symptoms characteristic of possible viral infection were collected from 47 fields, mainly located in south-eastern Estonia (Sõmera et al., 2020). Two samples of meadow fescue (*Festuca pratensis*) that were analyzed in this study originated from Sweden (Lit, County of Jämtland) and were collected in 2010.

Characterization of Swedish BYDV-OYV Isolate Using IC-RT-PCR and Sanger Sequencing

Extracts of two meadow fescue (*F. pratensis*) plants were used as the source of viral RNA in immunocapture (IC) RT-PCR (Bisnieks et al., 2004) with polyclonal antibodies for BYDV-PAV (Loewe Biochemica). Initially, RT-PCR was carried out using the luteovirus universal primer pair Shu-F and Yan-R (Malmstrom and Shu, 2004) to amplify the region corresponding to ORF4. The other genomic regions from ORF1 to 3′untranslated region (UTR) were amplified using a combination of published BYDV-PAV-CN (Liu et al., 2007) and newly designed BYDV-OYV primers: P326(+)/P12(−), P115(+)/OYV5(−), OYV4(+)/OYV1(−), OYV2(+)/OYV3(−), and P47(+)/P55(−; **Table 1**). cDNA was synthesized using the corresponding reverse primer and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions, while DreamTaq DNA polymerase was used for PCR. The amplification conditions were the following: 95°C for 2 min, followed by 35 cycles of 95°C for 30 s, 48–63°C for 2 min, 72°C for 1–2 min, with a final extension at 72°C for 10 min. As a negative control, PCR was run without cDNA template. Purified amplification products were ligated into pJET1.2 vector (Invitrogen) and transformed into *Escherichia coli* DH5α competent cells. Two clones of each amplification product were sequenced at Macrogen Inc., Amsterdam. A genomic sequence was assembled from the overlapping PCR fragments using MegaAlign.

5′ RACE Reaction for BYDV-OYV Genome

Total RNA was extracted from frozen plant material using Trizol reagent (Invitrogen) according to the manufacturer's protocol. The 5′ RACE reactions were performed using the Roche 5′/3′ RACE kit (ver. 13) for Avinurme2 and Ulvi samples from Estonia. The OYV5pR1 primer (**Table 1**) directed toward the 5′end of the BYDV-OYV genome was used for primer extension in RT-reaction and later on after polyA-tailing reaction in PCR together with the oligoT-anchoring primer. Next, the OYV5pR2 primer (**Table 1**) was used for nested PCR together with the oligoT-anchoring primer. RT-PCR products were separated by gel electrophoresis. The RT-PCR product of 183 bp was purified from the gel using GeneJET Gel Extraction and DNA Cleanup kit, ligated into the pJET1.2 blunt vector (all Thermo Scientific), and

TABLE 1 | Primers used in the current study.

Primers	Nucleotide sequence 5′ to 3′	Expected position	References
P326(+)	GACTTCGAGGCNGANCTCGCT	330–350	Liu et al., 2007
P12(–)	GCTCCGTCTGTGACCGCAAT	1226–1245	Liu et al., 2007
P115(+)	GGGTTTTTAGAGGGGCTCTGT	1157–1177	Liu et al., 2007
OYV5(–)	TCACCATGTTGAAGCCGTATT	2199–2219	This study
OYV4(+)	ATGTTCGTTGAGGATAAGATGC	1973–1994	This study
OYV1(–)	AGTACGTGAGAGCTAATGTAC	2800–2820	This study
Shu-F	TACGGTAAGTGCCCAACTCC	2650–2669	Malmstrom and Shu, 2004
Yan-R	TGTTGAGGAGTCTACCTATTTG	3459–3480	Malmstrom and Shu, 2004
OYV2(+)	TGAACTCGACACTGCGTGCA	3237–3256	This study
OYV3(–)	CTACCCGAGCTTATGAACCT	4857–4876	This study
P47(+)	GCAAAGGAGTACAAGGCACAAT	4777–4798	Liu et al., 2007
P55(–)	GGATTGCTATGGTTTATGTCC	5491–5511	Liu et al., 2007
OYV5pR1	AAGTCCGTCCAAGCCTCGG	533–541	This study
OYV5pR2	CTTTGACGCTGGCTCCAATGAGC	161–183	This study
PasF	GAAGAGGGCCAAATTCTATACC	3003–3024*	This study
OyvF	CCAATTCCTCAGGGATCC	3080–3097	This study
GavF	GTTACAAGATCACAAACGTCAAG	3156–3178**	This study
PavF	CTTCACAATCAGCAGGAC	3261–3278***	This study

*Primer binding positions are shown respective to full length sequence of BYDV-OYV Avinurme2 isolate (MK012645; unless mentioned otherwise). *BYDV-PAS (MK012660), **BYDV-GAV (MK012663), and ***BYDV-PAV (MK012661).*

transformed into *E. coli* DH5α competent cells. Plasmid DNA extracted from three individual colonies was sequenced using a vector-specific primer.

Sequencing of Small RNA Libraries and Virus Identification

The total RNA of 1–7 individual plant samples collected from the same field were pooled in equal concentrations to synthesize the indexed HTS libraries using the TruSeq siRNA kit (Illumina) according to the Sample Preparation Guide 02/2013. Depending on the number of collected plant samples, 1–3 HTS libraries were prepared for each field. The libraries were sequenced on HiSeq2500 as described in Sõmera et al. (2020). The number of raw reads varied between the library samples from 4.5 to 17.2 million. *De novo* contig assembly was done using the Oases 0.2.08 software (Schulz et al., 2012) varying the k-mer value from 15 to 31. The assembled contigs were analyzed using the BLAST + version 2.2.28 against the GenBank non-redundant databases of nucleotide collection and protein sequences, respectively, by using BlastN and BlastX search with standard parameters. All HTS data libraries were mapped to reference genomes using Geneious Prime software (2019.0.4). For that, 90 luteovirus, 10 polerovirus, BYDV-SGV, and MYDV-RMV complete genome sequences publicly available in March 2018 were retrieved from NCBI GenBank (see column 1 in **Supplementary Table 1**). *De novo* assembled complete genome consensus sequences of BYDV-OYV Avinurme2 and Ulvi isolates were used as the reference sequences for other BYDV-OYV sequences.

Mapping against the closest reference generated a consensus sequence. If any gaps existed in the sequence that was not filled during re-mapping, the mapping against other isolates was checked and in case the gaps were covered with mapped reads, the gaps were filled accordingly. This combined consensus was used as a draft reference sequence to repeat the mapping to obtain the final consensus sequence. The assembled complete consensus genome sequences were annotated and deposited in GenBank. Incomplete consensus sequences were used only for the identification of the virus species. After assembly of the complete genome consensus sequences, open reading frames (ORFs) were identified using the "Annotate and Predict" function in the Geneious program. The parameters were set for the standard genetic code using a minimum of 100 codons, starting with the AUG initiation codon. For detection of the non-AUG start codon of ORF3a, the near-cognate codons AUU, ACG, AUA, or CUG were considered. The site for -1 programmed ribosomal frameshift was identified according to alignment with other luteoviruses.

Pairwise Identity Calculations and Phylogenetic Analysis

The identity calculations of nucleotide sequences of the complete genome sequences and individual protein sequences were performed using the MUSCLE multiple sequence alignment tool implemented in the Geneious Prime program. The maximum-likelihood tree of nucleotide sequences for luteoviruses was constructed using PhyML ver. 3.1, implemented in the Seaview 4.6.1 program, with a GTR nucleotide substitution mode and 1000 bootstrap replications (Gouy et al., 2010).

Recombination Detection Point Analysis of BYDV-OYV

Recombination analysis of Estonian YDV isolates was carried out using the Recombination Detection Program (RDP4; Martin

et al., 2015) with a selection of reference luteovirus sequences aligned by ClustalW in MegaAlign (DNASTAR Lasergene 11). The recombination analyses were done as described in Kamali et al. (2016).

Design of Species-Specific Primers Able to Discriminate Between BYDV-PAV, -GAV, -PAS, and -OYV

MUSCLE multiple sequence alignment of luteovirus full-length genomes retrieved from GenBank was used for the selection of unique primer binding sites for BYDV-PAV, -GAV, -PAS, and -OYV in the region of ORF3/ORF4. A mix of new species-specific forward primers PasF, OyvF, GavF, and PavF (**Table 1**) together with the universal primer Yan-R (Malmstrom and Shu, 2004) was tested for the detection of single, double, triple, or quadruple infection of BYDV-PAV, -GAV, -PAS, and -OYV. The field-collected high-throughput sequenced BYDV samples were used as a reaction template for each species. Non-infected plant material was used as a negative control. Total RNA (1 μg) was used as a template in single-template RT-PCR. Per 20 μl reaction volume, 1 μl of dNTP mix (10 mM), 1 μl of Yan-R primer (200 μM), 200 U of Maxima reverse transcriptase, and 40 U of Ribolock RNase inhibitor (both Thermo Scientific) were used. Reverse transcription was carried out at 50°C for 1 h. In PCR, 0.5 μl of synthesis product was used as a template. The detection of multiple templates (BYDV-PAV, -GAV, -PAS, and -OYV) was tested using artificial mixes of their first-strand synthesis products (0.5 μl of each). The reaction was carried out using DreamTaq DNA polymerase green mix together with the primer mix of Yan-R (1 μl of 200 μM stock), PavF, GavF, PasF, and OyvF primers (0.5 μl each of 200 μM stocks). Yan-R primer was added after 5 cycles to favor the initial amplification of different BYDV templates. The amplification conditions were the following: 95°C for 3 min, followed by 35 cycles of 95°C for 1 min, 52°C for 1 min, 72°C for 1 min, with a final extension at 72°C for 5 min. Reaction products were separated electrophoretically, purified from the gel, and sequenced to verify the viral origin.

In silico Primer Binding Tests

In silico binding tests were performed to evaluate the specificity and sensitivity of 32 primer pairs (Robertson et al., 1991; Balaji et al., 2003; Malmstrom and Shu, 2004; Nagy et al., 2006; Deb and Anderson, 2008; Zhao et al., 2010; Tao et al., 2012; Svanella-Dumas et al., 2013) commonly used to detect YDVs species, including four newly designed pairs from this study (**Supplementary Table 2**). All primers were tested against 112 BYDV genomes (see column 1 in **Supplementary Table 1**) using the "Test with Saved Primers" option in Geneious. A maximum number of 0 to 3 mismatches were allowed between the primers and the viral sequences. Two measures were performed to assess the binding efficiency of each primer pair. First, the sensitivity was calculated as a ratio between the number of sequences of one BYDV species correctly identified and the total number of BYDV sequences belonging to this species. In this analysis, BYDV-PAS isolates 064, 0109, and KS-SHKR, which yet are identified as BYDV-PAV in GenBank, were treated as BYDV-PAV. Second, the specificity was calculated as a ratio between the number of sequences correctly not identified and the total number of BYDV sequences that should not have been identified. To be close to real PCR conditions, a mismatch on the last base of the 3′ end primers is considered to be an absence of binding.

RESULTS

Virus Identification by the Analysis of HTS Data

The analysis of HTS results from the virus survey in Estonia revealed the presence of one polerovirus, CYDV-RPS, and four luteoviruses, BYDV-GAV, BYDV-PAV, BYDV-PAS, and the tentative new species BYDV-OYV, in 24 of the 47 cereal fields included in the survey. A summary of the mapping results of specific HTS libraries is shown in **Table 2**. A near-complete consensus sequence of CYDV-RPS (MK012664) with a few small gaps was obtained from one HTS library. A consensus sequence of BYDV-PAV (MK012661) was also obtained from one HTS library. BYDV-GAV was detected from three HTS libraries for samples collected in different locations and during several years. Consensus sequences covering the complete genome were obtained for the Jõgeva3 and Kumna isolates (MK012662 and MK012663, respectively), but not for the Abja isolate (57.4% genome coverage). BYDV-PAS was detected in 12 HTS libraries. The full genome was assembled for 11 samples (MK012650-MK012660), while the genome sequence of the isolate BYDV-PAS Listaku was recovered only partially (19.6% of the full genome sequence). Detection of BYDV-OYV was confirmed in seven samples collected during three different years and in several locations. Complete consensus genome sequences of BYDV-OYV were obtained for seven isolates (MK012643-MK012649). The genomes of the isolates Avinurme2 and Ulvi were assembled *de novo*. The five other consensus genomes of BYDV-OYV were obtained by mapping these two references.

Characterization of BYDV-OYV Genome

Using 5′RACE sequencing, the 5′ ends of the complete genomes assembled by HTS data analysis were confirmed (100% identity) for two Estonian isolates (Avinurme2 and Ulvi) of the new candidate species. In parallel, BYDV-OYV was identified by IC-RT-PCR using BYDV-PAV antibodies in two meadow fescue samples collected in Sweden. Using the strategy of primer walking, 5,133 nt of the Swedish isolate (MK012642) was sequenced. The obtained sequence of the Swedish isolate covered 90% of the genome length, lacking the 5′-UTR, the beginning of ORF1, and a part of 3′UTR. The complete genome sizes for the seven Estonian BYDV-OYV isolates ranged between 5,664 and 5,681 nt (MK012643- MK012649). The biggest difference for the variation in genome length between the Estonian isolates lies at the beginning of ORF5 [encoding the read-through domain; read-through domain protein (RTD)] where the isolates of Saunja and Rannu have an insertion of 18 nucleotides and the isolates Ulvi, Avinurme1, Avinurme2, and Avinurme3 have an insertion of 6 nucleotides compared to the isolates from Sweden and Õru. The Õru isolate has an additional insertion of a single

TABLE 2 | Identification of YDVs collected in Estonia during the current study.

Field no	YDV species identified	Isolate name	GenBank Acc. No.	Genome coverage depth, average	Virus specific read/total reads in the library
1	BYDV-PAS	Imavere	MK012660	822.8	193,297/3,054,788
2	BYDV-PAS	Listaku	–		
3	BYDV-PAS	Jõgeva1	MK012659	156.6	37,484/4,298,886
4	BYDV-PAS	Jõgeva2	MK012658	197.3	49,267/3,362,095
5	BYDV-PAS	Matapera	MK012657	535.0	132,644/6,000,966
8	BYDV-PAS	Puide	MK012656	158.9	35,129/3,090,120
9	BYDV-GAV	Kumna	MK012663	434.4	111,635/2,304,733
10	BYDV-GAV	Jõgeva3	MK012662	315.5	79,589/10,679,704
11	BYDV-PAS	Jõgeva4	MK012654	359.2	91,136/1,220,020
14	BYDV-PAS	Väimela	MK012655	103.4	24,760/9,172,512
17	CYDV-RPS	Olustvere1-O	MK012664	24.2	6,129/4,998,325
18	BYDV-PAV	Olustvere1-B	MK012661	225.4	53,606/6,340,305
19	BYDV-PAS	Olustvere2-B	MK012653	383.3	86,751/5,531,072
20	BYDV-PAS	Olustvere2-W	MK012652	481.4	121,789/5,339,363
23	BYDV-PAS	Rannu1	MK012651	494.0	118,588/7,259,253
26	BYDV-OYV	Rannu2	MK012649	260.0	63,991/6,056,844
30	BYDV-OYV	Õru	MK012648	81.5	18,530/8,529,024
32	BYDV-PAS	Põlva	MK012650	61.8	15,999/2,863,991
35	BYDV-GAV	Abja	–		
39	BYDV-OYV	Ulvi	MK012647	286.9	72,813/7,931,998
40	BYDV-OYV	Avinurme1	MK012646	244.7	63,453/3,070,516
41	BYDV-OYV	Avinurme2	MK012645	371.8	86,326/8,693,892
43	BYDV-OYV	Avinurme3	MK012644	237.7	57,958/6,163,353
47	BYDV-OYV	Saunja	MK012643	465.7	120,362/6,946,669

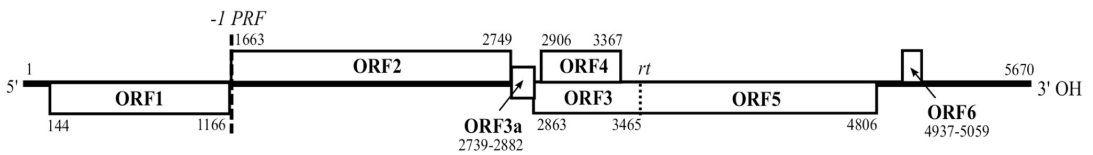


FIGURE 1 | Genome organization of BYDV-OYV. The nucleotide positions of open reading frames (ORFs) are exemplified using Avinurme2 isolate (MK012645). The functions of ORFs are assumed to be analogous to those assigned in other luteoviruses: P1 protein is encoded by ORF1, RdRp is translated via -1 programmed ribosomal frameshift (-1 PRF) and encoded by ORF2, P3a protein is encoded by ORF3a, viral coat protein (CP) is encoded by ORF3, viral cell-to-cell movement protein (MP) is encoded by ORF4, read-through domain (RTD) is encoded by ORF5 and translated via ORF3 stop codon read-through (rt), P6 is encoded by ORF6. Subgenomic RNAs (sgRNA) are needed for translation of P3a, MP, CP, or CP-RTD and P6.

nucleotide within the 3′-UTR. The Estonian BYDV-OYV isolates sampled from cereal plants shared a nucleotide identity of 93.9–98.6% with each other and showed an identity of 92.7–93.4% with the Swedish BYV-OYV isolate. The identity of the new sequences with the previously sequenced partial 502-bp CP gene sequence fragment of BYDV-OYV (Latvian isolate; AJ563410) was 95.8–97.0%. Assembly of the complete genome sequences for BYDV-OYV isolates revealed a genome organization and size characteristic of luteoviruses (**Figure 1**).

The sequence of BYDV-OYV Avinurme2 (MK012645) was further used for detailed genome annotation. The length of the BYDV-OYV Avinurme2 genome is 5,670 nt, with 5′- and 3′-UTRs of 143 and 611 nt, respectively. The beginning of the 5′-UTR is conserved between isolates of BYDV-OYV, but slightly different from other BYDVs, being most close to BYDV-PAS isolates. The 3′-UTR ends with the conserved motif CGGCAUCCC, which is also characteristic of BYDV-PAS and BYDV-PAV-CN. The coding region between the terminal UTRs is polycistronic and consists of seven predicted ORFs.

The ORF1 of BYDV-OYV (nt position 144–1166) encodes a protein P1 with a calculated molecular weight of 38.8 kDa. In case of the event of -1 programmed ribosomal frameshifting at the GGGUUUU just before the ORF1 stop codon, translation can continue from nt position 1163 in the frame of ORF2 until a stop codon at nt position 2747–2749, leading to the synthesis of a 98.8 kDa polyprotein P1-P2 constituting the viral RNA-dependent RNA polymerase (RdRp). Translation of ORF3a is predicted to depend on non-AUG initiation (Smirnova et al., 2015). Here, we identified the near-cognate codon ACG at nt position 2739–2741 and a stop codon at nt position 2880–2882.

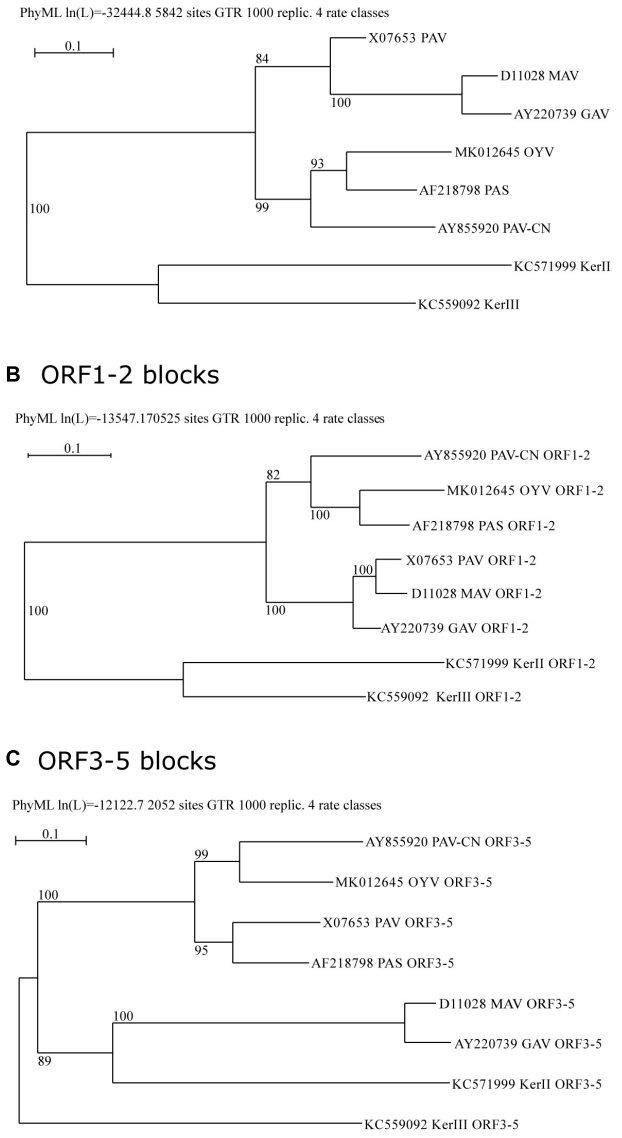


FIGURE 2 | Maximum-likelihood phylogenetic trees of BYDVs belonging to the genus *Luteovirus* based on a MUSCLE multiple sequence alignment of the complete genome **(A)**, ORF1-2 **(B)**, or ORF3-5 **(C)** nucleotide sequences, respectively. The trees were constructed using PhyML ver. 3.1, implemented in the Seaview 4.6.1 program, with a GTR nucleotide substitution model. Percentage values on the tree indicate the frequency with which the grouping occurred after bootstrapping with 1000 replicates and is shown only when >70. Scale bars indicate the distance in nucleotide substitution per site.

The calculated molecular weight for BYDV-OYV P3a is 5.3 kDa. The ORFs 3 and 4 overlap in different reading frames. BYDV-OYV ORF3 encodes a viral CP of 22 kDa; it starts at nt position 2863 and ends with an amber stop codon at nt position 3463–3465. The viral aphid transmission factor which is also called the RTD is translated from ORF5 as a polyprotein CP-RTD with a calculated molecular weight of 71.2 kDa via a read-through of the ORF3 stop codon. The sequence context (AAAUAGGUAGAC) around the ORF3 amber stop codon is conserved. BYDV-OYV ORF4 (nt pos. 2906–3367) encodes the cell-to-cell movement protein P4 with a calculated molecular mass of 17 kDa. Before the beginning of ORF6, there is an internal UTR of 131 nt. ORF6 of BYDV-OYV lies in nt position 4937–5059. The calculated molecular weight of the putative P6 protein is 4 kDa.

Phylogenetic Grouping of BYDV Species

Phylogenetic analyses of BYDV genomes supported clustering of cereal-infecting BYDVs (**Figure 2A**). Within this group, BYDV-PAV and BYDV-MAV (and its subspecies BYDV-GAV) grouped with 84% of bootstrap support, and BYDV-PAV-CN, BYDV-PAS, and BYDV-OYV grouped with 99% support whereas BYDV-OYV was closer to BYDV-PAS than to BYDV-PAV-CN. Phylogenetic analyses of ORF1-2 and ORF3-5 gene blocks (encoding P1-RdRp and CP-RTD, respectively; **Figures 2B,C**) showed conflicting results with each other and with the genome tree, a characteristic of recombination-driven species evolution. BYDV-OYV ORF1-2 clustered with that of BYDV-PAS and BYDV-PAV-CN, being closer to BYDV-PAS ORF1-2 (100% of support). BYDV-OYV ORF3-5 clustered with BYDV-PAV-CN ORF3-5 (99% of support) whereas BYDV-PAS ORF3-5 clustered with 95% of bootstrap support to that of BYDV-PAV. ORF3-5 of BYDV-MAV (and its subspecies BYDV-GAV) was distant to that of other cereal-infecting BYDVs, clustering with barley yellow dwarf kerII (BYDV-KerII) ORF3-5 (89% of support).

Pairwise Identity Between the Genomes and Respective Gene Products

Overall genome identity is higher for BYDV-OYV and BYDV-PAS (82.8%) than for BYDV-OYV and BYDV-PAV-CN (79.4%). Between BYDV-PAV-CN and BYDV-PAS, the overall genome identity is 79.7% (**Table 3**). Phylogenetic analyses of the ORF1-2 and ORF3-5 gene blocks are in concordance with their gene product identity percentages. Comparison of BYDV-OYV and BYDV-PAV-CN indicates that whereas proteins encoded by the 3′ half of their genomes (MP, CP, and RTD) share the highest identity, the proteins (P1, RdRp) encoded by the 5′ part of the BYDV-OYV genome are most similar to those of BYDV-PAS (**Table 3**). RdRp, encoded in the 5′ gene block, showed an amino acid identity of 90.5% between BYDV-OYV and BYDV-PAV-CN, but 94.9% between BYDV-OYV and BYDV-PAS, and 91.5% between BYDV-PAV-CN and BYDV-PAS. CP, encoded in the 3′ gene block, showed an identity of 88.0% between BYDV-OYV and BYDV-PAV-CN, but 71.5% between BYDV-OYV and BYDV-PAS, and 72.5% between BYDV-PAV-CN and BYDV-PAS. The short and highly conserved P3a protein (encoded from ORF3a located between the ORF1-2 and ORF3-5 gene blocks) showed the overall highest identities: 93.6% between BYDV-OYV and BYDV-PAV-CN, but 95.7% between BYDV-OYV and BYDV-PAS, and 97.9% between BYDV-PAV-CN and BYDV-PAS.

Recombination Detection in BYDV-OYV Genome

Recombination detection tests (**Figure 3**) for putative recombination events in the BYDV-OYV genome detected

TABLE 3 | Nucleotide and amino acid percentage identity calculations based on Muscle multiple sequence alignments of the genome sequences and gene products encoded by BYDV-OYV (MK012645) or BYDV-PAV-CN (AY855920) and the respective sequences of BYDV-KerII (KC571999), BYDV-KerIII (KC571992), BYDV-MAV (D11028), BYDV-GAV (AY220739), BYDV-PAS (AF218798), and BYDV-PAV (X07653).

BYDV-OYV	BYDV-KerII	BYDV-KerIII	BYDV-MAV	BYDV-GAV	BYDV-PAS	BYDV-PAV	BYDV-PAV-CN
Genome	60.9	64.1	69.1	70.5	82.8	76.1	79.4
P1	52.1	51.7	77.6	77.3	89.7	77.6	77.3
RdRp	75.8	77.0	87.9	87.1	94.9	88.1	90.5
P1-RdRp	66.5	67.6	83.9	82.9	92.9	84.0	85.4
P3a	78.7	76.6	95.7	91.5	95.7	91.5	93.6
MP	69.3	66.2	74.0	73.4	79.1	77.8	85.6
CP	63.0	58.5	66.0	66.0	71.5	75.5	88.0
CP-RTD	56.7	57.4	60.2	60.1	81.7	80.8	83.5
RTD	54.3	57.5	57.4	57.6	86.7	83.8	81.9
P6	37.8	ND	30.5	30.5	38.5	27.0	55.0
BYDV-PAV-CN	**BYDV-KerII**	**BYDV-KerIII**	**BYDV-MAV**	**BYDV-GAV**	**BYDV-PAS**	**BYDV-PAV**	**BYDV-OYV**
Genome	60.9	63.8	69.8	72.0	79.7	76.1	79.4
P1	53.7	52.3	79.1	79.9	80.5	79.4	77.3
RdRp	75.0	77.3	86.7	86.4	91.5	86.7	90.5
P1-RdRp	66.7	67.8	83.7	83.5	87.2	83.9	85.4
P3a	78.7	76.6	93.6	93.6	97.9	93.6	93.6
MP	69.3	66.2	72.1	71.4	79.1	77.8	85.6
CP	64.5	57.5	71.0	70	72.5	74.5	88.0
CP-RTD	56.2	57.6	59.7	59.9	77.6	77.3	83.5
RTD	54.9	59.3	57.6	57.4	80.5	79.2	81.9
P6	32.4	ND	45.0	42.5	38.5	28.6	55.0

intra-specific recombination in the genome of isolates Saunja, Avinurme 1, and Ulvi. When compared with representative genomes of other BYDV species, BYDV-OYV was also identified as a putative minor parent for recombinant regions in the genomes of BYDV-PAV-CN (AY855920 and EU332321) and BYDV-PAS-129 (AF218798) with two different BYDV-PAS genotypes as a putative major parent, respectively. In all these cases, detection of recombination was supported by at least 5 out of 9 methods used.

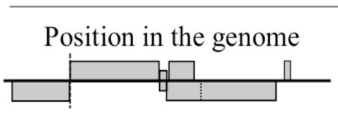
New Multiplex-RT-PCR for Discrimination of BYDV-OYV, -PAS, -PAV, and -GAV

A multiplex-RT-PCR specifically targeting BYDV-PAS, BYDV-OYV, BYDV-GAV, and BYDV-PAV found in this study was established for simultaneous monitoring. The single sets and the multiplex primer set produced RT-PCR fragments with the expected sizes: 476 bp for BYDV-PAS, 401 bp for BYDV-OYV, 317 bp for BYDV-GAV, and 193 bp for BYDV-PAV (**Figure 4**). Simultaneous detection of one, two, three, or four different targets in a single amplification was achieved, although the intensity of the bands corresponding to different viruses varied a bit if multiple templates were present in a single reaction (**Figure 4**). The reason might be a result of unbalanced template amounts as the real amount of viral template in each sample was unknown (the first-strand synthesis products of the previously used samples were artificially mixed to have different templates in one reaction). Weak non-specific signals were detected in the background when the picture was studied in inverted-color mode.

In silico Primer Analysis

The sensitivity and the specificity of the primers were estimated for 32 primer pairs, including the primer pairs designed in this study, are included in the **Supplementary Tables 1A–D**. The sensitivity reflects the ability of the primer pair to bind correctly to its target sequences, and the specificity measures the ability of the primer pair not binding to the non-target sequences.

Regarding the new candidate species BYDV-OYV, the novel "OyvF & Yan-R" primer pair shows high sensitivity and specificity. Regarding the BYDV-PAS isolates, both the primer pairs "PasF & Yan-R" and "PASF & PASR" show high specificity. The primer pair "GavF & Yan-R" shows the best complementary to BYDV-GAV isolates whereas the primer pair "MAVF & MAVR" shows the highest sensitivity and specificity to BYDV-MAV subspecies solely. The primer pairs "PavF & Yan-R" and "PAVL1 & PAVR1" show the highest sensitivity and specificity for BYDV-PAV isolates, although these primer pairs may miss and fail to detect several isolates. Theoretically, if three mismatches are allowed, "PavF & Yan-R" shows possible annealing to the BYDV-PAS isolates. However, this was not noticed in our multiplex-RT-PCR (**Figure 4**). The primer pair "F3-PAV & B3-PAV" is the only primer pair targeting BYDV-PAV-CN isolates (although not all of them) while retaining a high specificity.



Isolate	Recombination breakpoints	Major parent	Minor parent	No. of methods support the recombination	Probabilities
BYDV-OYV Saunja	863 and 1749	BYDV-OYV Rannu	BYDV-OYV Avinurme2	7/9	2.215×10^{-4} - 1.069×10^{-12}
BYDV-OYV Avinurme1	494 and 955	BYDV-OYV Avinurme2	BYDV-OYV Õru	5/9	1.433×10^{-4}- 1.237×10^{-6}
BYDV-OYV Ulvi	663 and 2260	BYDV-OYV Rannu	BYDV-OYV Õru	5/9	1.204×10^{-3} - 5.679×10^{-6}
BYDV-PAV -CN AY855920	2881 and 3466	BYDV-PAS-129 AF218798	BYDV-OYV Avinurme2	5/9	Probabilities 3.339×10^{-4}- 2.037×10^{-7}
BYDV-PAS-129 AF218798	4565 and 4928	BYDV-PAS Imavere	BYDV-OYV Avinurme1	6/9	3.621×10^{-3} - 6.251×10^{-7}
BYDV-PAV 05ZZ4 (PAVIII) EU332321	2886 and 3471	Unknown (BYDV-PAS-129 AF218798)	BYDV-OYV Avinurme2	5/9	Probabilities 3.339×10^{-4} - 2.037×10^{-7}

FIGURE 3 | Determination of putative recombination events between the isolates of BYDV-OYV; and BYDV-OYV, BYDV-PAS, and BYDV-PAV-CN using RDP4 software. Only events supported by minimum of five methods are reported. Nucleotide numbering corresponds to the aligned sequences.

Several analyzed primer pairs show potential unspecific amplification for isolates belonging to different species. For example, the primer pair "PAV (forward) & PAV (reverse)" does not discriminate BYDV-PAV, BYDV-PAS, and BYDV-MAV/GAV; the primer pair "MAVL1 & MAVR1" does not discriminate BYDV-MAV and BYDV-PAV, and "MAVF2 & YanR" recognizes BYDV-GAV instead of BYDV-MAV even if no mismatches to the target are allowed. If mismatches are allowed, the probability of unspecific recognition also shows up for several other primer pairs.

Among the universal primers designed to recognize multiple BYDV species, the primer pair "Luteo1F & YanR-New" is the most universal — theoretically, it only fails to detect BYDV-KerII and -KerIII isolates, two isolates of BYDV-PAV and one BYDV-PAV-CN isolate.

DISCUSSION

By high throughput sequencing of field samples from a country-wide survey in Estonia, we identified five species of cereal/barley YDVs – CYDV-RPS, BYDV-GAV, BYDV-PAS, BYDV-PAV, and BYDV-OYV which are described in the current study. The putative novel luteovirus, BYDV-OYV, was also identified in the samples of meadow fescue collected from Sweden and included in our analysis. The same survey identified brome mosaic virus, European wheat striate mosaic virus, oat sterile dwarf virus, and putative cereal closterovirus (Sõmera et al., 2016, 2020; Sõmera et al., unpublished). In addition, wheat dwarf virus was found in 2017 (Sõmera et al., 2019).

In the past, the novel luteovirus BYDV-OYV was detected on one occasion from a neighboring country, Latvia. During this previous study, a 502-nt fragment of the CP gene was amplified and sequenced using the luteovirus universal primer pair Lu1/Lu4. A phylogenetic relationship to a group of Chinese isolates of BYDV-PAV (BYDV-PAV-CN) was proposed but the taxonomic status of BYDV-OYV and BYDV-PAV-CN remained unclear (Bisnieks et al., 2004).

After the start of the sequencing era, it was suggested that the isolates of BYDV-PAV-like sequences be allocated to three quasi-descriptive groupings: PAV-I, PAV-II, and PAV-III corresponding to BYDV-PAV, BYDV-PAS, and BYDV-PAV-CN (Liu et al., 2007). BYDV-PAS, which was initially recognized as the severe strain of BYDV-PAV breaking the standard tolerance in oats (Chay et al., 1996), has now been recognized as a distinct species due to genome sequence divergences. Additional phylogenetic analyses

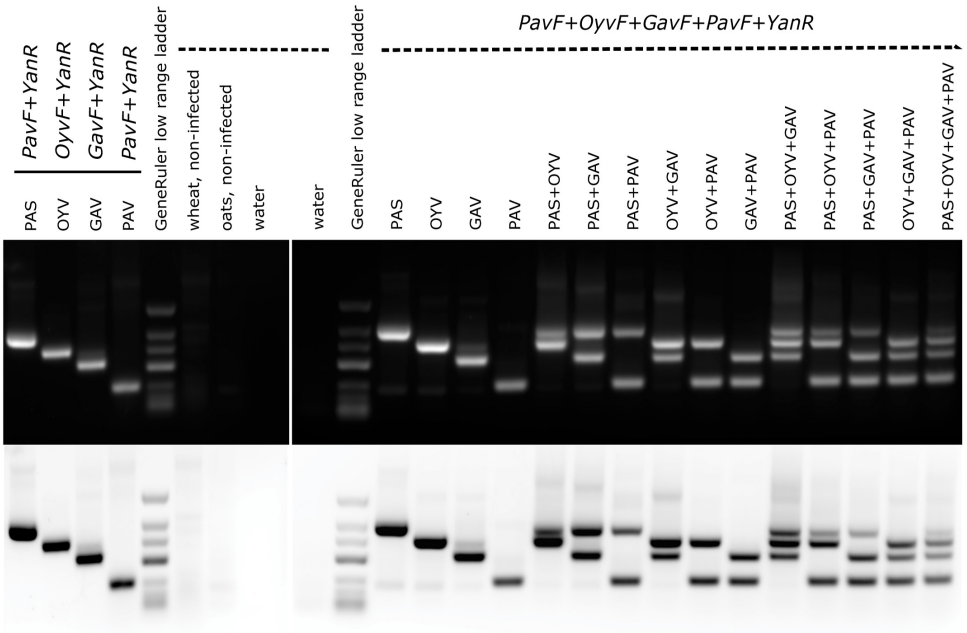


FIGURE 4 | Single and multiplex RT-PCR detection of BYDV-PAS, -OYV, -GAV, and -PAV. The expected RT-PCR product sizes are as follows: BYDV-PAS 476 bp, BYDV-OYV 401 bp, BYDV-GAV 317 bp, and BYDV-PAV 193 bp. GeneRuler low range DNA ladder (ThermoScientific) was used as a size marker (700/500/400/300/200/150/100 bp). 1 μg of total RNA extractions from field-derived plant samples was used as a template in cDNA synthesis reaction. For PCR step, artificial mixes of the first-strand synthesis products were mixed in equal quantities (0.5 μl of each template). The following isolates were used: BYDV-PAS Imavere, BYDV-OYV Avinurme2, BYDV-GAV Kumna, and BYDV-PAV Olustvere1-B.

have indicated that BYDV-PAV and BYDV-PAV-CN isolates also split into two distinct major groups (Miller et al., 2002; Boulila, 2011; Wu et al., 2011). Sequencing the complete genome of BYDV-PAV-CN confirmed that BYDV-PAV-CN fullfills the species demarcation criteria of sequence divergence established for the genus *Luteovirus*. However, the authors suggested waiting until the genome of BYDV-OYV is sequenced to determine whether these two viruses could be classified as a single species or two distinct species (Liu et al., 2007).

Extensive recombination detection analyses indicate that recombination imprints are common in BYDV genomes (Pagan and Holmes, 2010; Wu et al., 2011). The existence of recombination events further complicates the taxonomy of BYDV species. When the viral RdRp genes have been analyzed, it has been observed that the isolates of BYDV-PAV and BYDV-MAV were present in a monophyletic clade whereas BYDV-PAV-CN formed another monophyletic lineage together with the isolates of BYDV-PAS. At the same time, the phylogenies of the CP and MP genes placed BYDV-PAV and BYDV-PAS isolates in a monophyletic clade separate from BYDV-MAV and BYDV-CN-PAV (Miller et al., 2002; Hall, 2006). The characterization of the BYDV-OYV CP gene suggested that there is at least one more putative BYDV species close to BYDV-PAV-CN (Bisnieks et al., 2004). In the current study, we present for the first time the complete genome characterization of seven BYDV-OYV isolates from Estonia, and a near-complete genome sequence for one isolate from Sweden.

Our phylogenetic analyses and sequence alignment studies indicate that BYDV-PAV-CN and BYDV-OYV share the closest relationships between their 3′ gene blocks (CP-RTD/MP), but slightly looser relationships between their 5′ gene blocks (P1/RdRp), which instead show the closest relationships between BYDV-PAS and BYDV-OYV. Remarkably, a very highly conserved region (also known as ORF3a) resides between these two gene blocks. Most probably, such a highly conserved region is a suitable template for copy-choice recombination during mixed infection. The template-switching model includes dissociation of the replicase and nascent strand, followed by nascent RNA strand hybridization to the region of complementarity in the acceptor strand. It has been proposed that the subgenomic promoters at the end of the 5′gene block (within the region encoding RdRp) of luteoviruses serve as the "hot spot" sites for recombination during replication (Miller et al., 1995; Koev et al., 1999). Alternatively, a bulged stem–loop of the sgRNA promoter region in the acceptor strand may act as an interactor, which helps the replicase (or a host factor involved in the replication complex) to relocate during a template-switching process near the complementary region (Miller and Koev, 1998). In recombination detection analyses, BYDV-OYV was identified as a putative minor parent for recombinant regions in the genomes of BYDV-PAV-CN and BYDV-PAS. Agreeing with the sequence comparisons, the recombination analyses suggest that the common PAS genotype (represented by Imavere isolate) recombined with BYDV-OYV leading to a more diverse

BYDV-PAS-129 now having a part of ORF5 from BYDV-OYV. Furthermore, BYDV-PAS-129 recombined with BYDV-OYV resulting in BYDV-PAV-CN which then has ORF3/4 from BYDV-PAS-129 and the end part of ORF5 from BYDV-OYV.

According to the identity calculations, the multiple gene products of BYDV-OYV show differences of around 10% or higher when compared with the respective gene products of other BYDVs (**Table 3**). Species demarcation criteria in the genus *Luteovirus* is a >10% difference in amino acid sequence identity of any gene product from its closest relative. Therefore, considering the sequence analyses described and taking into account that both viruses have been found from several different locations and years, we propose creating two new species in the genus *Luteovirus*: BYDV-OYV and BYDV-PAV-CN.

Based on these taxonomic distinctions, the incidence results can be analyzed. In our survey, BYDV-PAS and BYDV-OYV were two YDVs whose incidence in the sampled plants was higher compared to other YDV species (BYDV-PAV, BYDV-GAV, and CYDV-RPS). Interestingly, BYDV-PAV was found only in one sample. The absence of BYDV-PAV, but the abundance of BYDV-PAS, has been reported in Alaska, which is exposed to climatic conditions similar to Northern Europe with long photoperiods and cooler temperatures during the growing season (Robertson and French, 2007). It has been speculated that this situation is related to the ability of BYDV-PAS to overwinter in perennial grasses. In the Czech Republic, where BYDV-PAS is more prevalent than BYDV-PAV, this difference has been suggested to be related to the number of aphid species capable of transmitting these viruses – BYDV-PAV is recorded as being transmitted by *Rhopalosiphum padi* and *Sitobion avenae* whereas BYDV-PAS is also transmitted by *R. maidis* and *Metopolophium dirhodium* (Jarošová et al., 2013).

Whereas BYDV-PAS was most commonly identified from winter wheat (although it was also present in some samples of spring wheat and other cereals) in our survey, BYDV-OYV was detected from oats and spring wheat. The first discovery of BYDV-OYV was also in oats (Bisnieks et al., 2004). Our finding of BYDV-OYV infection in meadow fescue in Sweden indicates one possible overwintering host. The occurrence of BYDV-OYV in spring cereals suggests that it might be transmitted from perennial grasses to cereals during the growing season whereas BYDV-PAS might be transmitted to winter cereals during the autumn flight of the aphids, and later on, to the spring cereals. Intriguingly, differences in identified host species may account for differences in dominant vector species. The aphid vector of BYDV-OYV is currently unknown and needs to be revealed.

In the case of cereal-infecting YDVs, BYDV-PAV, BYDV-MAV, and MYDV-RMV are generally accepted as being the most prevalent YDV species worldwide (Signoret and Maroquin, 1990; Zhou and Zhang, 1990; Henry and Adams, 2003; Parry et al., 2012). The raised CP-specific antisera can detect the serotypes of CYDV-RPV, MYDV-RMV, BYDV-PAV, BYDV-MAV, and BYDV-SGV, which are the five most characteristic YDVs in the United States (Gildow, 1990). Similarly, the most widely used multiplex-RT-PCR protocol enables one to discriminate against these viruses (Malmstrom and Shu, 2004). The sequencing of virus amplicons has revealed that BYDV-PAS isolates could have been incorrectly identified as BYDV-PAV (Robertson and French, 2007). Therefore, the original protocol was improved to discriminate BYDV-PAS by adding a restriction digestion step (Kundu et al., 2009) or an extra pair of primers for BYDV-PAS (Laney et al., 2018). Even if sequenced, false identification of new BYDV-PAS isolates as BYDV-PAV can easily occur if only BYDV-PAV sequences are retrieved from GenBank and included in phylogenetic analyses as a subset of the genome and CP gene sequences of BYDV-PAS isolates still exist under the name of BYDV-PAV there (see Najar et al., 2017).

The existing B/CYDV antisera fail to detect or do not allow proper discrimination between some species and may lead to biased epidemiological analysis. Earlier studies have noticed that BYDV-PAV antiserum does not discriminate BYDV-PAV and BYDV-PAS (Chay et al., 1996). Likewise, BYDV-OYV was detected by BYDV-PAV antiserum (Bisnieks et al., 2004), and BYDV-KerII and BYDV-KerIII were detected by BYDV-PAV and BYDV-MAV antisera (Svanella-Dumas et al., 2013). BYDV-GPV, a tentative member of the genus *Polerovirus* that has been known to not react with antisera raised against MAV, PAV, SGV, RPV, or RMV and has exclusively been detected in China until recently (Zhang et al., 2009), was suddenly reported to occur in the Czech Republic according to HTS data analysis (Singh et al., 2020).

The occurrence of other species might also be underestimated. For example, another YDV species found in our study, CYDV-RPS, can also be masked by cross-reaction of CYDV-RPV antiserum (no CYDV-RPS specific antiserum has been raised) or amplified by CYDV-RPV primers. According to currently existing data, CYDV-RPS seems to be very rare in comparison to CYDV-RPV although both are transmitted by *R. padi*. There are only a few earlier findings of CYDV-RPS, all based on sequencing – the first ones from Mexico (Miller et al., 2002) and Iran (Rastgou et al., 2005) were detected by RT-PCR using CYDV-RPV primers, and all the recent ones come from HTS analyses performed in the United Kingdom (Pallett et al., 2010), the United States (Malmstrom et al., 2017), the Czech Republic (Singh et al., 2020), and Estonia (this study). Similar to the previous example, there is a lack of specific antisera for BYDV-GAV and it can be misidentified as BYDV-MAV due to cross-reaction of MAV-antiserum (Zhou and Zhang, 1990). For a long time, BYDV-GAV was supposed to be exclusively spread in China (Liu et al., 2007), but recent sequencing-related findings confirm its presence in Poland (Trzmeil, 2017) and in Estonia (this study). Therefore, further findings of BYDV-GAV may be expected in other regions in Eurasia as well when the detection will be verified by sequencing. It has been recorded that BYDV-GAV found from China is transmitted by *S. avenae* and *Schizaphis graminum* whereas the master species BYDV-MAV collected from North America is known not to be transmitted by *S. graminum* (Rochow, 1982; Liu et al., 2007). The biological characterization of European BYDV-GAV isolates remains to be carried out. Finally, recent HTS-data based identifications of new cereal-infecting poleroviruses, MaYMV, and BVG related to MYDV-RMV (Chen et al., 2016; Zhao et al., 2016), lead to the question whether these viruses could react with MYDV-RMV antibodies and might have been interpreted as MYDV-RMV in previous serotype-specific MYDV-RMV identifications.

The use of antibody detection or multiplex RT-PCR for simultaneous amplification of different virus templates is helpful to reduce diagnostic costs (several multiplex-RT-PCR protocols are available for YDVs, see: Malmstrom and Shu, 2004; Deb and Anderson, 2008; Tao et al., 2012). However, these multiplex tests have been designed to discriminate the species characteristic for the geographic region of sampling and are based on a limited set of YDV genetic diversity in the databases. Therefore, they should be used with care as sequences of other YDV species can be amplified and misidentified by the primers used in these protocols, which remains unknown if the PCR product is not sequenced. In addition, other YDV species different enough in primer annealing positions may remain unidentified as suggested by our *in silico* primer binding test and the similar analysis carried out by Laney et al. (2018). The potential to detect a subset of BYDV species as well as the specificity and sensitivity of commonly used primer pairs varied greatly (**Supplementary Table 1**). Therefore, it can be concluded that the ability of a diagnostician to detect the virus(es) infecting a sample relies on the techniques used and their specificity.

HTS technologies can provide full virological indexing of samples while being as sensitive as RT-PCR (Santala and Valkonen, 2018), although they still represent a much higher cost per sample. Nevertheless, their use in virus surveys, alone or in combination with targeted RT-PCR, will most probably grow in the future. Detection of new virus species and genotypes by HTS indicates a growing need to adapt the detection primers continuously and make the process efficient and quick (Katsiani et al., 2018). With a significant impact on primer development, HTS technologies are helping to target the existing diversity of the viral population within ecosystems and are taking steps toward an improved understanding of virus epidemiology and evolution, including that of cereal and barley YDVs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MK012642–MK012664.

AUTHOR CONTRIBUTIONS

MS designed the study, prepared the HTS libraries and analyzed the data, performed the multiplex-RT-PCR experiment, wrote the manuscript, and prepared the figures and tables. SM advised on HTS data analysis and took part in planning and finalizing the manuscript. LT performed *in silico* primer analysis and discussed the data. PS performed the field sampling and collected metadata about the sampled sites. KS sequenced the Swedish isolate of BYDV-OYV. AK supervised KS, performed recombination detection analysis, discussed the interpretation of the data, and took part in finalizing the manuscript. All authors agreed to be accountable for the content of the work.

FUNDING

This article was based upon collaboration from COST Action FA1407 (DIVAS), supported by COST (European Cooperation in Science and Technology). Research visit of MS to Swedish University of Agricultural Sciences was performed thanks to a Swedish Institute scholarship. Lundströms Stiftelse offered financial support for the sequence determination of the Swedish isolate. The Estonian field survey was performed as part of the Estonian Biotechnology Program "Breeding for disease resistance in plants" (2012–2015).

ACKNOWLEDGMENTS

MS is grateful to Erkki Truve (deceased in April 2020) for his collegial support and discussions. Anna Akkerman is appreciated for technical assistance with the Estonian samples, and Ingrid Eriksson is appreciated for technical assistance with analyses of the Swedish samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.673218/full#supplementary-material

REFERENCES

Balaji, B., Bucholtz, D. B., and Anderson, J. M. (2003). Barley yellow dwarf virus and cereal yellow dwarf virus quantification by real-time polymerase chain reaction in resistant and susceptible plants. *Virology* 93, 1386–1392. doi: 10.1094/phyto.2003.93.11.1386

Bisnieks, M., Kvarnheden, A., Sigvald, R., and Valkonen, J. P. T. (2004). Molecular diversity of the coat protein-encoding region of *Barley yellow dwarf virus-PAV* and *Barley yellow dwarf virus-MAV* from Latvia and Sweden. *Arch. Virol.* 149, 843–853. doi: 10.1007/s00705-003-0242-2

Boulila, M. (2011). Selective constraints, molecular recombination structure and phylogenetic reconstruction of isometric plant RNA viruses of the families *Luteoviridae* and *Tymoviridae*. *Biochemie* 93, 242–253. doi: 10.1016/j.biochi.2010.09.017

Chay, C. A., Smith, D. M., Vaughan, R., and Gray, S. M. (1996). Diversity among isolates within the PAV serotype of barley yellow dwarf virus. *Phytopathology* 86, 370–377. doi: 10.1094/phyto-86-370

Chen, S., Jiang, G., Wu, J., Liu, Y., Qian, Y., and Zhou, X. (2016). Characterization of a novel polerovirus infecting maize in China. *Viruses* 8:e120.

D'Arcy, C. J. (1995). "Symptomology and host range of barley yellow dwarf," in *Barley yellow dwarf: 40 years of progress*, eds C. J. D. D'Arcy and P. A. Burnett (St. Paul, MN: The American Phytopathological Society), 9–28.

Deb, M., and Anderson, J. M. (2008). Development of a multiplexed PCR detection method for barley and cereal yellow dwarf viruses, wheat spindle streak virus, wheat streak mosaic virus and soil-borne wheat mosaic virus. *J. Virol. Methods* 148, 17–24. doi: 10.1016/j.jviromet.2007.10.015

Gildow, F. E. (1990). "Current status of barley yellow dwarf in the United States: A regional situation report," in *World perspectives on barley yellow dwarf*, ed. P. A. Burnett (Mexico: DCAS), 11–20.

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259

Halbert, S., and Voegtlin, D. (1995). "Biology and taxonomy of vectors of barley yellow dwarf viruses," in *Barley Yellow Dwarf: 40 Years of Progress*, eds C. J. D'Arcy and P. A. Burnett (St. Paul, MN: American Phytopathological Society), 217–258.

Hall, G. (2006). Selective constraint and genetic differentiation in geographically distant barley yellow dwarf virus populations. *J. Gen. Virol.* 87, 3067–3075. doi: 10.1099/vir.0.81834-0

Henry, M., and Adams, M. J. (2003). "Other cereals," in *Virus and Virus-like Diseases of Major Crops in Developing Countries*, eds G. Loebenstein and G. Thottapilly (Dordrecht: Springer Science+Business Media), 337–354.

Jarošová, J., Chrpová, J., Šíp, V., and Kundu, J. K. (2013). A comparative study of the barley yellow dwarf virus species PAV and PAS: distribution, virus accumulation and host resistance. *Plant Pathol.* 62, 436–443. doi: 10.1111/j.1365-3059.2012.02644.x

Kamali, M., Heydarnejad, J., Massumi, H., Kvarnheden, A., Kraberger, S., and Varsani, A. (2016). Molecular diversity of turncurtoviruses in Iran. *Arch. Virol.* 161, 551–561. doi: 10.1007/s00705-015-2686-6

Katsiani, A., Maliogka, V. I., Katis, N., Svanella-Dumas, L., Olmos, A., Ruiz-García, A. B., et al. (2018). High-throughput sequencing reveals further diversity of little cherry virus 1 with implications for diagnostics. *Viruses* 10:E385. doi: 10.3390/v10070385

Koev, G., Mohan, B. R., and Miller, W. A. (1999). Primary and secondary structural elements required for synthesis of barley yellow dwarf virus subgenomic RNA1. *J. Virol.* 73, 2876–2885. doi: 10.1128/jvi.73.4.2876-2885.1999

Kundu, J. K., Jarosova, J., Gadiou, S., and Cervena, G. (2009). Discrimination of three BYDV species by one-step RT-PCR-RFLP and sequence based methods in cereal plants from the Czech Republic. *Cereal Res. Commun.* 37, 541–550. doi: 10.1556/crc.37.2009.4.7

Laney, A. G., Acosta-Leal, R., and Rotenberg, D. (2018). Optimized yellow dwarf virus multiplex PCR assay reveals a common occurrence of *Barley yellow dwarf virus*-PAS in Kansas winter wheat. *Plant Health Prog.* 19, 37–43. doi: 10.1094/php-09-17-0056-rs

Liu, F., Wang, X., Liu, Y., Xie, J., Gray, S. M., Zhou, G., et al. (2007). A Chinese isolate of barley yellow dwarf virus-PAV represents a third distinct species within the PAV serotype. *Arch. Virol.* 152, 1365–1373. doi: 10.1007/s00705-007-0947-8

Malmstrom, C. M., Bigelow, P., Trêbicki, P., Busch, A. K., Friel, C., Cole, E., et al. (2017). Crop-associated virus reduces the rooting depth of non-crop perennial native grass more than non-crop-associated virus with known viral suppressor of RNA silencing (VSR). *Virus Res.* 241, 172–184. doi: 10.1016/j.virusres.2017.07.006

Malmstrom, C., and Shu, R. (2004). Multiplexed RT-PCR for streamlined detection and separation of barley and cereal yellow dwarf viruses. *J. Virol. Methods* 120, 69–78. doi: 10.1016/j.jviromet.2004.04.005

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolut.* 1:vev003.

Martin, R. R., Keese, P. K., Young, M. J., Waterhouse, P. M., and Gerlach, W. L. (1990). Evolution and molecular biology of luteoviruses. *Annu. Rev. Phytopathol.* 28, 341–363. doi: 10.1146/annurev.py.28.090190.002013

Miller, W. A., and Koev, G. (1998). Getting a handle on RNA virus recombination. *Trends Microbiol.* 6, 421–423. doi: 10.1016/s0966-842x(98)01384-5

Miller, W. A., Dinesh-Kumar, S. P., and Paul, C. P. (1995). Luteovirus gene expression. *Crit. Rev. Plant Sci.* 14, 179–211. doi: 10.1080/713608119

Miller, W. A., Liu, S., and Beckett, R. (2002). Barley yellow dwarf virus: *Luteoviridae* or *Tombusviridae*? *Mol. Plant Pathol.* 3, 177–183. doi: 10.1046/j.1364-3703.2002.00112.x

Nagy, A. A., Sharaf, A. N., Soliman, M. H., Shalaby, A.-B. A., and Youssef, A. (2006). Molecular characterization of barley yellow dwarf virus coat protein gene in wheat and aphids. *Arab. J. Biotech.* 10, 207–218.

Najar, A., Hamdi, I., and Varsani, A. (2017). Barley yellow dwarf virus in barley crops in Tunisia: prevalence and molecular characterization. *Phytopathol. Mediterr.* 56, 111–118.

Pagan, I., and Holmes, E. C. (2010). Long-term evolution of the *Luteoviridae*: time scale and mode of virus speciation. *J. Virol.* 84, 6177–6187. doi: 10.1128/jvi.02160-09

Pallett, D. W., Ho, T., Cooper, I., and Wang, H. (2010). Detection of *Cereal yellow dwarf virus* using small interfering RNAs and enhanced rate with *Cocksfoot streak virus* in wild cocksfoot grass (*Dactylis glomerata*). *J. Virol. Methods* 168, 223–227. doi: 10.1016/j.jviromet.2010.06.003

Parry, H. R., Macfadyen, S., and Kirticos, D. J. (2012). The geographical distribution of yellow dwarf viruses and their aphid vectors in Australian grasslands and wheat. *Australas. Plant Pathol.* 41, 375–387. doi: 10.1007/s13313-012-0133-7

Rastgou, M., Khatabi, B., Kvarnheden, A., and Izadpanah, K. (2005). Relationships of *Barley yellow dwarf virus*-PAV and *Cereal yellow dwarf virus*-RPV from Iran with viruses of the family *Luteoviridae*. *Eur. J. Plant Pathol.* 113, 321–326. doi: 10.1007/s10658-005-1231-y

Robertson, N. L., and French, R. (2007). Genetic structure in natural populations of barley/cereal yellow dwarf virus isolates from Alaska. *Arch. Virol.* 152, 891–902. doi: 10.1007/s00705-006-0913-x

Robertson, N. L., French, R., and Gray, S. M. (1991). Use of group specific primers and the polymerase chain reaction for the detection and identification of luteoviruses. *J. Gen. Virol.* 72, 1473–1477. doi: 10.1099/0022-1317-72-6-1473

Rochow, W. F. (1982). Identification of barley yellow dwarf viruses: comparison of biological and serological methods. *Plant Dis. Rep.* 1982, 381–384. doi: 10.1094/pd-66-381

Santala, J., and Valkonen, J. P. T. (2018). Sensitivity of small RNA-based detection of plant viruses. *Front. Microbiol.* 9:939. doi: 10.3389/fmicb.2018.00939

Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 8, 1086–1092. doi: 10.1093/bioinformatics/bts094

Signoret, P. A., and Maroquin, C. (1990). "The barley yellow dwarf virus situation in Western Europe," in *World perspectives on Barley yellow dwarf*, ed. P. A. Burnett (Mexico: DCAS), 42–44.

Singh, K., Jarošová, J., Fousek, J., Chen, H., and Kundu, J. K. (2020). Virome identification in wheat in the Czech Republic using small RNA deep sequencing. *J. Integr. Agric.* 19, 1825–1833. doi: 10.1016/S2095-3119(19)62805-4

Smirnova, E., Firth, A. E., Miller, W. A., Scheidecker, D., Brault, V., Reinbold, C., et al. (2015). Discovery of a small non-AUG-initiated ORF in poleroviruses and luteoviruses that is required for long-distance movement. *PLoS Pathog.* 11:e1004868. doi: 10.1371/journal.ppat.1004868

Sõmera, M., Gantsovski, M., Truve, E., and Sooväli, P. (2016). First report of brome mosaic virus in wheat in Estonia. *Plant Dis.* 100:2175. doi: 10.1094/PDIS-04-16-0426-PDN

Sõmera, M., Kvarnheden, A., Desbiez, C., Blystad, D.-R., Sooväli, P., Kundu, J. K., et al. (2020). Sixty years after the first description: genome sequence and biological characterization of European wheat striate mosaic virus infecting cereal crops. *Phytopathology* 110, 68–79. doi: 10.1094/PHYTO-07-19-0258-FI

Sõmera, M., Truve, E., and Sooväli, P. (2019). First report of wheat dwarf virus in winter wheat in Estonia. *Plant Dis.* 103:1797. doi: 10.1094/PDIS-11-18-2059-PDN

Svanella-Dumas, L., Candresse, T., Hullé, M., and Marais, A. (2013). Distribution of *Barley yellow dwarf virus*-PAV in the Sub-Antarctic Kerguelen Islands and characterization of two new *Luteovirus* species. *PLoS One* 8:e67231. doi: 10.1371/journal.pone.0067231

Tao, Y., Man, J., and Wu, Y. (2012). Development of a multiplex polymerase chain reaction for simultaneous detection of wheat viruses and a phytoplasma in China. *Arch. Virol.* 157, 1261–1267. doi: 10.1007/s00705-012-1294-y

Trzmeil, K. (2017). Identification of barley yellow dwarf viruses in Poland. *J. Plant Pathol.* 99, 493–497.

Van den Eynde, R., Van Leeuwen, T., and Haesaert, G. (2020). Identifying drivers of spatio−temporal dynamics in barley yellow dwarf virus epidemiology as a critical factor in disease control. *Pest Manag. Sci.* 76, 2548–2556. doi: 10.1002/ps.5851

Wu, B., Blanchard-Letort, A., Liu, Y., Zhou, G., Wang, X., and Elena, F. S. (2011). Dynamics of molecular evolution and phylogeography of *Barley yellow dwarf virus*-PAV. *PLoS One* 6:e16896. doi: 10.1371/journal.pone.0016896

Zhang, P., Liu, Y., Liu, W., Cao, M., Massart, S., and Wang, X. (2017). Identification, characterization and full-length sequence analysis of a novel polerovirus associated with wheat leaf yellowing disease. *Front. Microbiol.* 8:1689. doi: 10.3389/fmicb.2017.01689

Zhang, W., Cheng, Z., Xu, L., Wu, M., Zhou, G., and Li, S. (2009). The complete nucleotide sequence of the barley yellow dwarf GPV isolate from China shows that it is a new member of the genus *Polerovirus*. *Arch. Virol.* 154, 1125–1128. doi: 10.1007/s00705-009-0415-8

Zhao, F., Lim, S., Yoo, R. H., Igori, D., Kim, S. M., Kwak do, Y., et al. (2016). The complete genomic sequence of a tentative new polerovirus identified in barley in South Korea. *Arch. Virol.* 161, 2047–2050. doi: 10.1007/s00705-016-2881-0

Zhao, K., Liu, Y., and Wang, X. (2010). Reverse transcription loop-mediated isothermal amplification of DNA for detection of barley yellow dwarf viruses in China. *J. Virol. Methods* 169, 211–214. doi: 10.1016/j.jviromet.2010.06.020

Zhou, G., and Zhang, S. X. (1990). "Identification of the variants of Barley yellow dwarf virus in China," in *World perspectives on barley yellow dwarf*, ed. P. A. Burnett (Mexico: CIMMYT), 290–292.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fmicb.2021.684599

Completion of Maize Stripe Virus Genome Sequence and Analysis of Diverse Isolates

*Stephen Bolus[1], Kathryn S. Braithwaite[2], Samuel C. Grinstead[1], Irazema Fuentes-Bueno[1], Robert Beiriger[3], Bryce W. Falk[4] and Dimitre Mollov[1]**

[1] National Germplasm Resources Laboratory, United States Department of Agriculture-Agricultural Research Service, Beltsville Agricultural Research Center, Beltsville, MD, United States, [2] Sugar Research Australia Limited, Indooroopilly, QLD, Australia, [3] Institute of Food and Agricultural Sciences, University of Florida, Belle Glade, FL, United States, [4] Department of Plant Pathology, University of California, Davis, Davis, CA, United States

Maize stripe virus is a pathogen of corn and sorghum in subtropical and tropical regions worldwide. We used high-throughput sequencing to obtain the complete nucleotide sequence for the reference genome of maize stripe virus and to sequence the genomes of ten additional isolates collected from the United States or Papua New Guinea. Genetically, maize stripe virus is most closely related to rice stripe virus. We completed and characterized the RNA1 sequence for maize stripe virus, which revealed a large open reading frame encoding a putative protein with ovarian tumor-like cysteine protease, endonuclease, and RNA-dependent RNA polymerase domains. Phylogenetic and amino acid identity analyses among geographically diverse isolates revealed evidence for reassortment in RNA3 that was correlated with the absence of RNA5. This study yielded a complete and updated genetic description of the tenuivirus maize stripe virus and provided insight into potential mechanisms underpinning its diversity.

Keywords: maize stripe, tenuivirus, diversity, reassortment, *Zea mays*, *Rottboellia cochinchinensis*

OPEN ACCESS

Edited by:
Rajarshi Kumar Gaur, Deen Dayal Upadhyay Gorakhpur University, India

Reviewed by:
Anders Kvarnheden, Swedish University of Agricultural Sciences, Sweden
Siew Pheng Lim, Denka Life Innovation Research (DLIR), Singapore

***Correspondence:**
Dimitre Mollov
dimitre.mollov@usda.gov

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *23 March 2021*
Accepted: *26 April 2021*
Published: *14 June 2021*

Citation:
Bolus S, Braithwaite KS, Grinstead SC, Fuentes-Bueno I, Beiriger R, Falk BW and Mollov D (2021) Completion of Maize Stripe Virus Genome Sequence and Analysis of Diverse Isolates. Front. Microbiol. 12:684599. doi: 10.3389/fmicb.2021.684599

INTRODUCTION

Rice, maize, and sorghum are staple food crops. Diverse plant pathogens can threaten global food security and agricultural economies by infecting these vital crop plants and reducing their marketable yield. *Maize stripe virus* is a tenuivirus species that induces stippling symptoms between leaf veins on corn (*Zea mays* L.), which later can coalesce into continuous chlorotic stripes. Furthermore, infection of young plants often leads to stunting and dramatic "hoja blanca" or white leaf symptoms (Falk and Tsai, 1998). The first scientific reports of maize stripe virus (MSpV) were from Hawaii, Cuba, Trinidad, Mauritius, and East Africa (Storey, 1936). Serological testing of MSpV isolates from the United States (Florida), Venezuela, Peru, Australia, India, Mauritius, Réunion, Thailand, and Taiwan showed that they were all related (Gingery et al., 1979; Greber, 1981; Peterschmitt et al., 1987, 1991; De Doyle et al., 1992; Chen et al., 1993; Sdoodee et al., 1997). Besides infecting corn plants, MSpV isolates have caused disease on sorghum [*Sorghum bicolor* (L.) Moench] in India (Peterschmitt et al., 1991; Srinivas et al., 2014) and itchgrass [*Rottboellia cochinchinensis* (Lour.) Clayton] in the United States (Florida) (Gingery et al., 1981). The host specificity and geographical distribution of MSpV are largely explained by that of its

vector, the corn planthopper *Peregrinus maidis* Ashmead, which transmits MSpV in a circulative-propagative manner (Tsai and Zitter, 1982; Nault and Gordon, 1988; Falk and Tsai, 1998; Singh and Seetharama, 2008). Corn planthoppers are also capable of transmitting MSpV transovarially (Tsai and Zitter, 1982).

MSpV is serologically related to the tenuivirus species *Rice stripe virus* (Gingery et al., 1983), which is vectored by the small brown planthopper *Laodelphax striatellus* Fallén. There are reports of rice stripe virus (RSV) infecting maize (Gingery et al., 1983; Bradfute and Tsai, 1990), although its infamy comes from its epidemics on *japonica* cultivars of rice (*Oryza sativa* L.) in Eastern Asia (Wang et al., 2008; Otuka et al., 2010). Both MSpV and RSV are grouped in the genus *Tenuivirus* in the family *Phenuiviridae* (Abudurexiti et al., 2019). Tenuiviruses and the vertebrate-infecting viruses in the genus *Phlebovirus* share conserved complementary RNA end sequences and commonalities in their nucleoprotein, RNA-dependent RNA polymerase (RdRp), and glycoprotein sequences (Falk and Tsai, 1998). However, in contrast to the enveloped virions of phleboviruses, tenuiviruses (from "tenuis," meaning slender in Latin) are distinguished by their non-enveloped, thread-like ribonucleoprotein particles (Ramírez and Haenni, 1994; Falk and Tsai, 1998). Tenuivirus genomes also differ from those of phelboviruses in that they often have four or five negative and ambisense RNAs (Ramírez and Haenni, 1994; Falk and Tsai, 1998).

The genome of a Florida (United States of America) isolate of MSpV was composed of five RNAs (Falk and Tsai, 1984), and the complete sequences for RNAs 2–5 were determined (Huiet et al., 1991, 1992, 1993; Estabrook et al., 1996). In a separate effort, a partial sequence of RNA1 from an isolate of MSpV from Réunion (France) was determined (Mahmoud et al., 2007). These efforts revealed that RNA1 most likely encodes an RdRp with similarity to that of RSV (Mahmoud et al., 2007). RNA2 is ambisense and encodes p2, a putative membrane-associated protein, on the viral RNA strand and pc2, a putative glyco-polyprotein, on the viral complementary RNA strand (Estabrook et al., 1996). RNA3 and RNA4 are also ambisense with RNA3 encoding p3 and pc3, the nucleocapsid protein (Huiet et al., 1991), and RNA4 encoding p4, the major non-capsid protein, and pc4 (Huiet et al., 1990, 1992). The major non-capsid protein accumulates at very high amounts *in planta*, forming inclusion bodies and needle-shaped crystals that are visible by light microscopy (Bradfute and Tsai, 1990; Falk and Tsai, 1998). RNA5 only encodes pc5, a highly basic, hydrophilic protein of unknown function (Huiet et al., 1993). The intergenic regions in the ambisense RNAs are thought to be important for transcription termination and contain a conserved inverted repeat sequence motif that may form a stem-loop structure (Zhang et al., 2007).

There have been notable advancements in the molecular characterization of several proteins encoded by RSV, the type member of the genus *Tenuivirus*. The p2 and p3 proteins were shown to be silencing suppressors *in planta* (Xiong et al., 2009; Du et al., 2011). The p3 protein also functioned as a silencing suppressor for another tenuivirus, *Rice hoja blanca virus* (Bucher et al., 2003). The glyco-polyprotein pc2 was identified as a helper component for RSV, allowing it to overcome the midgut barriers of its insect vector (Lu G. et al., 2019). In addition, pc4 was recognized as the *in planta* movement protein for RSV (Xiong et al., 2008; Fu et al., 2018).

In this paper, we report the first complete genome sequence for an isolate of MSpV. Using high-throughput sequencing (HTS), we sequenced three additional isolates from *Z. mays* collected in the United States of America. We also identified and sequenced seven isolates of MSpV from *Z. mays* and *R. cochinchinensis* plants collected in Papua New Guinea. We compared these 11 sequenced isolates with other tenuiviruses, other MSpV isolates, and each other to explore patterns underlying the genetic diversity of MSpV.

MATERIALS AND METHODS

Plant Material and RNA Extraction

We obtained an RNA sample originating from the genomic sequencing and characterization work previously performed with a Florida, United States of America (USA) isolate of MSpV (Falk and Tsai, 1984; Huiet et al., 1991, 1992, 1993; Estabrook et al., 1996). We refer to this isolate as MSpV21. In 2019, symptomatic leaves were collected from *Z. mays* (maize) in Palm Beach County, Florida, United States. Symptomatic leaves were also collected in 2019 from *Z. mays* and *R. cochinchinensis* (itchgrass) plants from the Ramu Valley in the Madang Province of Papua New Guinea (PNG) as part of joint Sugar Research Australia and Ramu Agri Industries Limited (RAIL) sugarcane disease surveys. Collected leaves from PNG were stored in tubes containing anhydrous granular calcium chloride (Merck, Darmstadt, Germany) as the drying agent, were treated with 25 or 50 kGy gamma irradiation in Australia, and were forwarded to the United States for further processing. RNA was extracted from the leaf samples using either KingFisher Pure RNA Plant Kit (Thermo Fisher Scientific, Waltham, MA, United States) or RNeasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturers' instructions.

High Throughput Sequencing

DNase treatment, ribosomal RNA depletion, cDNA synthesis, and library preparation were outsourced (SeqMatic, Fremont, CA, United States). Libraries were sequenced on an Illumina NextSeq 500 platform as 75 single end reads. HTS data were analyzed using CLC Workbench 11–20 (Qiagen).

To quantify the number of reads mapping to each RNA of every MSpV isolate, the reads per kilobase per million reads (RPKM) measurements were calculated by taking the total number of reads mapping to each isolate RNA, dividing by the nucleotide length of the RNA and the total number of sample reads, and finally multiplying by 10^9 (Wagner et al., 2012). Read mapping was performed using CLC Workbench 11–20 (Qiagen).

Genome Completion for Isolate MSpV21

To confirm the 5′ and 3′ terminal sequences of RNAs 1–5 from MSpV isolate MSpV21, cDNA was first synthesized from RNA using SuperScript III First-Strand Synthesis System for RT-PCR (Thermo Fisher Scientific) and a universal tenuivirus

5′ and 3′ ends primer Tenui (De Miranda et al., 1994) or a genome-specific primer (**Supplementary Table 1**). The specific RNA end regions for each RNA molecule were then amplified from cDNA using GoTaq Green Master Mix and protocol, Tenui primer, and genome specific primers (**Supplementary Table 1**). The PCR products thus obtained were ligated to pGEM-T Easy Vector and cloned in competent *Escherichi coli* JM109 cells using the manufacturer's kit and protocol (Promega, Madison, WI, United States). At least three clones for each end were selected and sequenced using M13 F and M13 R primers (MCLAB, South San Francisco, CA, United States). Final RNA genome alignments were made using Geneious v. 9 (Biomatters, Auckland, New Zealand) and CLC Workbench 11–20 (Qiagen) software.

Genome Annotation and Analysis

The assembled genome sequences for all 11 isolates were submitted to the National Center for Biotechnology Information (NCBI)'s GenBank database (**Table 1**). NCBI's Conserved Domain-Search tool was used to identify the conserved domains present in the pc1 sequence of MSpV21 (Lu S. et al., 2019), and NCBI's Open Reading Frame Finder "ORFfinder" was used to identify the coding regions of all the MSpV isolates. The Basic Local Alignment Search Tool (BLAST) from NCBI was used to search for related nucleotide and amino acid sequences and to determine their corresponding percent identities.

Recombination and Phylogenetic Analyses

Alignments of RNAs and encoded proteins were made using the ClustalW method in Molecular Evolutionary Genetics Analysis (MEGA) X under default settings (Kumar et al., 2018). When appropriate, RNA alignments were trimmed at the ends, since the terminal sequences were not determined for all the isolates. Recombination Detection Program v.4.101 (RDP4) (Martin et al., 2015) was used to identify any possible recombinant regions in the individual RNA alignments and was also used to identify any possible RNA reassortments using a concatenated RNA sequence alignment as input. A full exploratory recombination scan was performed after selecting options of linear sequences, 0.05 *P*-value, and Bonferroni correction and selecting the recombination detection methods of RDP, GENECONV, Chimaera, MaxChi, BootScan, SiScan, and 3Seq (Martin et al., 2015). Areas of potential recombination or reassortment were reported only if they were identified by more than four of these selected detection methods under the described significance criteria.

To construct the percent identity matrices, pairwise distances were computed using the Poisson correction model under default settings in MEGA X (Kumar et al., 2018) using selected amino acid alignments as input. Pairwise distances were then converted into percent identities using the following formula: percent identity = 100 – (pairwise distance*100).

To make the phylogenetic trees, selected amino acid and nucleotide alignments were first subjected to model testing in MEGA X (Kumar et al., 2018). Based on the model testing results, the following models were used for the corresponding phylogenies: LG + G + F for RdRp, T92 + G for RNA3 in **Figure 3B**, GTR + I for RNA1, T92 + G + I for RNA2, HKY + G for RNA3 in **Figure 4C**, and T92 + I for RNA4. Maximum likelihood phylogenetic trees were constructed using the previously described parameters with 1,000 bootstrap replications and the partial deletion option selected.

RESULTS

MSpV Genome Completion and Characterization

MSpV21 Isolate Genome Completion

Using HTS and completing the ends using Sanger sequencing, we determined the complete genome sequence of MSpV21, an isolate of MSpV that had previously been sequenced, except for RNA1 (Falk and Tsai, 1984; Huiet et al., 1991, 1992, 1993; Estabrook et al., 1996). Our sequences for RNAs 2–5 of MSpV21 were 99–100% identical to those previously deposited in NCBI (**Supplementary Table 2**). Using ORFfinder (NCBI), we identified the coding regions in our MSpV21 isolate and compared its encoded proteins to those previously deposited in NCBI (**Supplementary Table 3**). As with the nucleotide sequences, the amino acid sequences were 99–100% identical to the previously deposited sequences (**Supplementary Table 3**). Excluding other MSpV sequences, BLAST nucleotide and protein searches revealed that MSpV21 RNAs 1–4 and encoding protein sequences were most like corresponding sequences from RSV. RNA5 and pc5 are not present in the RSV genome. These sequences were most like those deposited for tenuivirus *Echinochloa hoja blanca virus* (**Supplementary Tables 4, 5**).

Tenuiviruses have conserved and complementary end sequences, possibly explaining the circular forms of ribonucleoproteins observed by electron microscopy (Ramírez and Haenni, 1994; Falk and Tsai, 1998). We compiled and aligned the 5′ and 3′ termini of complete, genomic RNAs from our MSpV21, 1704-01, 1704-03, and 2002-07 isolates and compared them to MSpV reference sequences in NCBI (**Supplementary Figure 1**). The expected conservation and complementarity of end sequences for each MSpV RNA segment was apparent (**Supplementary Figure 1**). A few exceptions are noted. For RNA1, the sixth nucleotide position from the 3′ end was variable, with 1704-01 and 1704-03 sequences having an A and MSpV21 having a U (**Supplementary Figure 1**). At the same aligned position for RNA5, there was a U for our MSpV21 sequence, whereas the reference isolate and our 2002-07 sequences had an A at this position (**Supplementary Figure 1**). We attribute the differences observed at this alignment position to real and/or artifactual genetic variability, since the universal Tenui primer (De Miranda et al., 1994) was used to complete the ends (**Supplementary Table 1**).

RNA1 Characterization

After obtaining the first complete sequence for RNA1 from an isolate of MSpV, we proceeded to characterize the 9,011 nucleotides long RNA1 from MSpV21. ORFfinder (NCBI)

identified a long, open reading frame encoding a protein of 2,919 amino acids in the viral complementary strand (**Figure 1A**). We refer to this putative protein as pc1. RNA1 and pc1 from MSpV are very similar to those of RSV (Toriyama et al., 1994; **Supplementary Tables 4, 5**). The Conserved Domain-Search tool (NCBI) identified several conserved domains in the pc1 sequence of MSpV21, including ovarian tumor-like cysteine protease (OTU), N-terminus bunyavirus endonuclease (Endo), domain of unknown function found in viruses (DUF3770), and bunyavirus RNA-dependent RNA polymerase (RdRp) (**Figure 1B**). Further bioinformatic and manual inspection of the OTU domain motifs revealed that the putative OTU in MSpV shared all the amino acids that were previously shown to be conserved and essential for the deubiquitinating enzyme function of OTU domain from RSV (Makarova et al., 2000; Zhao et al., 2020). Closer inspection of the Endo domain also revealed that the H. . .D. . .PD. . .ExT. . .K. . .Y motif was conserved between RSV and MSpV, including the essential amino acids for endonuclease activity identified in RSV (Zhao S. et al., 2019; **Figure 1C**). Analysis of MSpV21's RdRp domain revealed that the conserved motifs (pre-A/F, A, H, B, C, D, and E) of bunyaviruses" RdRps (Amroun et al., 2017) were present (**Figure 1C**). For motif G, only the conserved R was present in RSV and MSpV instead of the usual RY (Amroun et al., 2017). Combined, these results indicate that the RdRp of MSpV is very similar to RSV and could be expected to function similarly.

MSpV Isolate Diversity Analysis

Read Mapping to RNAs

Given our HTS methodology and the fact that MSpV is an RNA virus, we could not clearly differentiate between genomic and transcriptomic reads for our MSpV isolates. Nonetheless, we sought to compare the total reads mapping to each RNA across MSpV isolates as expressed in their reads per kilobase per million reads (RPKM) measurements (**Figure 2**). Of note, RNA5 was only robustly detected in MSpV21 and 2002-07 isolates (**Figure 2**). Although no uniform pattern of RNA abundance stood out across these sampled isolates, the RPKM measurements between RNAs in any one sample usually did not differ by more than 3.5-fold (**Figure 2**). These results are like those reported for RSV, where there was at most a 15-fold genomic RNA difference between the four RNA segments *in planta* as measured by absolute real-time quantitative PCR (Zhao W. et al., 2019). The relative abundance of genomic RNAs also varied across time of infection (1–20 days after inoculation) in that study (Zhao W. et al., 2019), possibly explaining the lack of global RNA abundance patterns observed in our isolates.

Recombination/Reassortment Analysis

RDP4 (Martin et al., 2015) was used to detect areas of possible recombination in alignments of individual RNAs from all 11 isolates of MSpV sequenced in this study. No areas of recombination that met our significance criteria were detected in the individual RNA 1, 2, 3, and 4 alignments (data not presented). RNA5 was not analyzed for recombination due to the presence of only two sequences from our isolates. To probe for RNA segment reassortment in our 11 isolate sequences, we concatenated RNAs 1–4 for each isolate and aligned them as input for the RDP4 program. RDP4 did detect a region in both 2002-04 (12,354–14,758) and 2002-10 (12,355–14,487) isolates that largely overlapped with the nucleotide positions (12,326–14,638 and 12,326–14,727) of their respective RNA3 sequences in the alignment (**Table 2**). In fact, the 99% confidence intervals for both regions encompassed the start and stop positions for each isolates' RNA3 (**Table 2**). Six out of the seven selected recombination detection methods identified the "recombinant" region at high confidence (probability range 10^{-13}–10^{-62}) (**Table 2**). This finding suggests that 2002-04 and 2002-10 isolates could be derived from a reassortment event with RNAs 1, 2, and 4 coming from a parent isolate like MSpV21 and RNA3 from an isolate more like 1704-04.

Phylogenetic Relationships of MSpV Isolates

With other tenuiviruses

We assembled a maximum likelihood phylogenetic tree to compare the recently determined MSpV RdRp amino acid

TABLE 1 | Depository information for maize stripe virus isolates.

Maize stripe virus isolate	Host plant	Country collected	RNA1*	RNA2*	RNA3*	RNA4*	RNA5*
MSpV21	Zm	USA	MW328593	MW328594	MW328595	MW328596	MW328597
1704-01	Rc	PNG	MW491852	MW491853	MW491854	MW491855	N/A
1704-02	Rc	PNG	MW491856	MW491857	MW491858	MW491859	N/A
1704-03	Zm	PNG	MW491860	MW491861	MW491862	MW491863	N/A
1704-04	Rc	PNG	MW491864	MW491865	MW491866	MW491867	N/A
1909-05	Zm	PNG	MW491868	MW491869	MW491870	MW491871	N/A
1909-06	Zm	PNG	MW491872	MW491873	MW491874	MW491875	N/A
1909-07	Zm	PNG	MW491876	MW491877	MW491878	MW491879	N/A
2002-04	Zm	USA	MW491839	MW491840	MW491841	MW491842	N/A
2002-07	Zm	USA	MW491843	MW491844	MW491845	MW491846	MW491847
2002-10	Zm	USA	MW491848	MW491849	MW491850	MW491851	N/A

Rc, Rottboellia cochinchinensis; Zm, Zea mays; PNG, Papua New Guinea; USA, United States of America.
**National Center for Biotechnology Information accession numbers.*
N/A, not applicable.

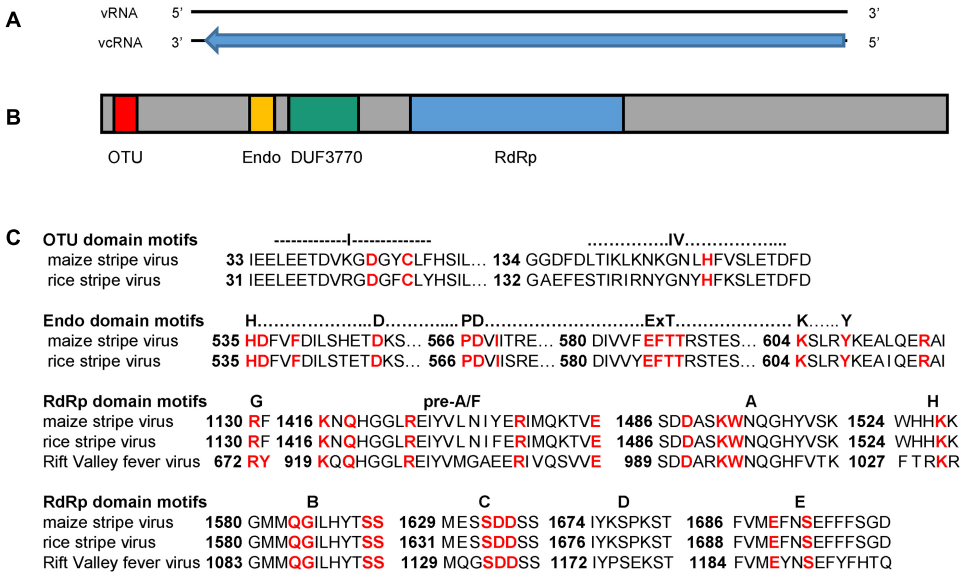


FIGURE 1 | (A) Schematic diagram of MSpV21 vcRNA1 at 9,011 nt in length (black bar) with contiguous coding region for pc1 covering nucleotide positions 50–8809 (blue arrow). **(B)** Schematic diagram of MSpV21 pc1 amino acid sequence with conserved domains (colored boxes) identified by the National Center for Biotechnology Information (NCBI) Conserved Domain Search tool (Lu S. et al., 2019) indicated. OTU = Ovarian tumor-like cysteine protease (Accession cl19932/E-value 6.03e-03) (amino acid positions 42-123); Endo = N-terminus bunyavirus endonuclease (Accession cl20011/E-value 2.76e-07) (amino acid positions 512-599); DUF3770 = domain of unknown function found in viruses (Accession cl13978/E-value 9.60e-57) (amino acid positions 648-889); RdRp = Bunyavirus RNA-dependent RNA polymerase (Accession = cl20265/E-value = 0e + 00) (amino acid positions 1,066–1,802) **(C)** Alignment between pc1 sequences of maize stripe virus isolate MSV21 and rice stripe virus (NCBI Reference Sequence NP_620522.1) with focus on the domain regions with conserved and/or essential amino acids marked in red. The RdRp sequence for the phlebovirus Rift Valley fever virus (NCBI Accession ACE78348.1) was also included in the RdRp domain motif analyses.

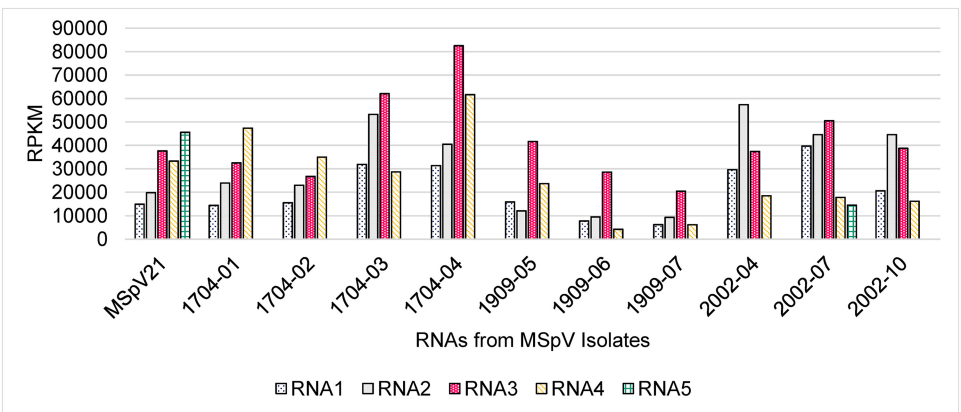


FIGURE 2 | The reads per kilobase per million reads (RPKM) measurements for the maize stripe virus RNA segments as detected by high-throughput sequencing. MSpV, maize stripe virus.

sequences with RdRp amino acid sequences from assigned and unassigned tenuiviruses (**Figure 3A**). The tree was based on 2,017 amino acid positions. The RdRp sequences from MSpV isolates form a clade and are most closely related to the representative RdRp sequence of RSV. The RdRp sequences from melon chlorotic spot virus and Ramu stunt virus formed a clade separate from MSpV, RSV, rice hoja blanca virus, European wheat striate mosaic virus, and rice grassy stunt virus (**Figure 3A**).

With other MSpV isolates

We sought to compare our MSpV isolate sequences to those that are publicly available. The most abundant, complete MSpV RNA sequence that is deposited in GenBank (NCBI) is RNA3. We, therefore, assembled, aligned, and built a maximum likelihood phylogeny based on RNA3 sequences from our MSpV isolates and those that had been deposited in GenBank (NCBI). The resulting tree revealed that our MSpV isolates from PNG form a monophyletic group (**Figure 3B**). RNA3 sequences from MSpV

TABLE 2 | Potential reassortment regions identified among maize stripe virus isolates using recombination detection program v.4.101 (RDP4).

Identified "recombinant" isolate	Suggested major parent isolate	Suggested minor parent isolate	"Recombinant" region identified	Number of RDP4 methods identifying region	Probability range
2002-04	MSpV21	1704-04	12,354–14,758*	6/7	10^{-13}–10^{-62}
2002-10	MSpV21	1704-04	12,355–14,487*	6/7	10^{-13}–10^{-62}

**In the concatenated RNA1-4 sequence alignment, nucleotide positions 12,326–14,638 and 12,326–14,727 corresponded to RNA3 for 2002-04 and 2002-10, respectively. To be noted: the beginning and ending breakpoint 99% confidence intervals of the "recombinant" region for both isolates encompassed RNA3 start and stop nucleotide positions.*

isolates from *S. bicolor* in India (Srinivas et al., 2014) also formed a monophyletic group (**Figure 3B**). However, RNA3 sequences from United States split into two separate groups, with 2002-04 and 2002-10 isolates forming a clade with PNG and India isolates and with 2002-07 and MSpV21 forming a separate clade with an isolate of MSpV from Réunion (Mahmoud et al., 2007; **Figure 3B**). The RNA3 sequence from the Réunion isolate was previously shown to be highly related to the reference MSpV isolate from Florida, United States (MSpV21) (Mahmoud et al., 2007).

With each other

We were interested in whether phylograms of individual RNA sequences from each of the MSpV isolates sequenced as part of this study would exhibit the same topologies across the conserved RNAs 1–4. The resulting maximum likelihood phylogenetic trees revealed that for RNAs 1, 2, and 4, the PNG and United States isolates form two distinct, clades (**Figures 4A,B,D**). However, for RNA3, 2002-04 and 2002-10 United States isolates group apart from other United States isolates and with PNG isolates (**Figure 4C**), supporting an observation made using comparable parameters in **Figure 3B**. The data suggest that RNA3 from 2002-04 and 2002-10 may have resulted from an ancestral reassortment event, mirroring the RDP4 analysis results (**Table 2**). Across the phylogenies, there was no distinct grouping based on host plant (*R. cochinchinensis* and *Z. mays*). In other words, a homogenous virus population appears to infect both plants in PNG.

Protein Sequence Identities Among Isolates

We were interested whether the differences observed between isolates in the RNA phylogenies would translate to differences observed at the protein level between isolates. Therefore, we made percent identity matrices for the conserved proteins encoded by RNAs 1–4 of our MSpV isolates (**Figure 5**). High percent identities (99.5–99.9%) of pc1 between isolates from the same geographic origin were observed, whereas pc1 differed (98.0–98.3% identical) when comparing isolates from distinct regions (**Figure 5A**). There was also high identity (97.3–100%) among United States and PNG isolates and lower identity between them (94.2–95.2% identical) for pc2. However, p2 percent identity values revealed a third group as 2002-04 and 2002-10 shared lower identity values to both 2002-07 and MSpV21 United States isolates (94.9–95.9%) and PNG isolates (92.2–93.3%) (**Figure 5B**). The percent identity matrices for p3 and pc3 reflect phylogenetic tree groupings in **Figure 4C** with sequences from United States isolates 2002-04 and 2002-10 being closer related to PNG isolates (97.1–99.0% identical) than 2002-07 and MSpV21 United States isolates (92.6–95.1% identical) (**Figure 5C**). Amino acid percent identities overall were high for p4 (lowest 98.3%), and excluding p4 sequences from United States isolate MSpV21 and PNG isolate 1909-07, PNG and United States isolates formed two distinct groups with 100% intragroup identity. The percent identity matrix for pc4 largely reflected geographical origins of the isolates except for the PNG isolate 1909-07, which was more similar (99.6% identical) to pc4 sequences from United States isolates 2002-04 and 2004-10 than those of other United States and PNG isolates (98.9–99.3% identical) (**Figure 5D**). The pc5 protein sequences from MSpV21 and 2002-07 were compared using BLAST (NCBI). These sequences were 97.6% identical (100% query coverage, *E*-value 0.0). Overall, the data from the protein percent identity analyses support the results from the RNA phylogenetic trees.

DISCUSSION

We have completed the genome sequence of the reference United States (Florida) isolate of MSpV (MSpV21) and used HTS to determine the genomes of ten additional MSpV isolates from PNG and the United States. There were minor discrepancies between our HTS-derived sequence for the reference MSpV isolate and those previously deposited in NCBI for RNAs 2–5. We largely attribute these differences to the fact that our HTS sequences are derived from consensus sequences of hundreds of thousands of mapped reads and the reference isolate sequences previously deposited in NCBI were mostly derived from several cDNA clones (Huiet et al., 1991, 1992, 1993; Estabrook et al., 1996). In this paper, we also describe, to our knowledge, the first report of MSpV in PNG. We sequenced seven isolates of MSpV from PNG infecting both *R. cochinchinensis* and *Z. mays*. We did not observe evidence of genomic separation by host plant for PNG MSpV isolates, supporting the notion that itchgrass may serve as a reservoir for MSpV as long postulated (Gingery et al., 1981).

The conserved, terminal sequences of MSpV may interact with distinct regions of the RdRp in a pre-initiation configuration, as was shown for La Crosse orthobunyavirus (Gerlach et al., 2015; Amroun et al., 2017). Consensus terminal sequences for RNAs 2–5 of our MSpV genomic RNA sequences matched those for the published MSpV reference isolate (Huiet et al., 1991, 1992, 1993; Estabrook et al., 1996). The first and last

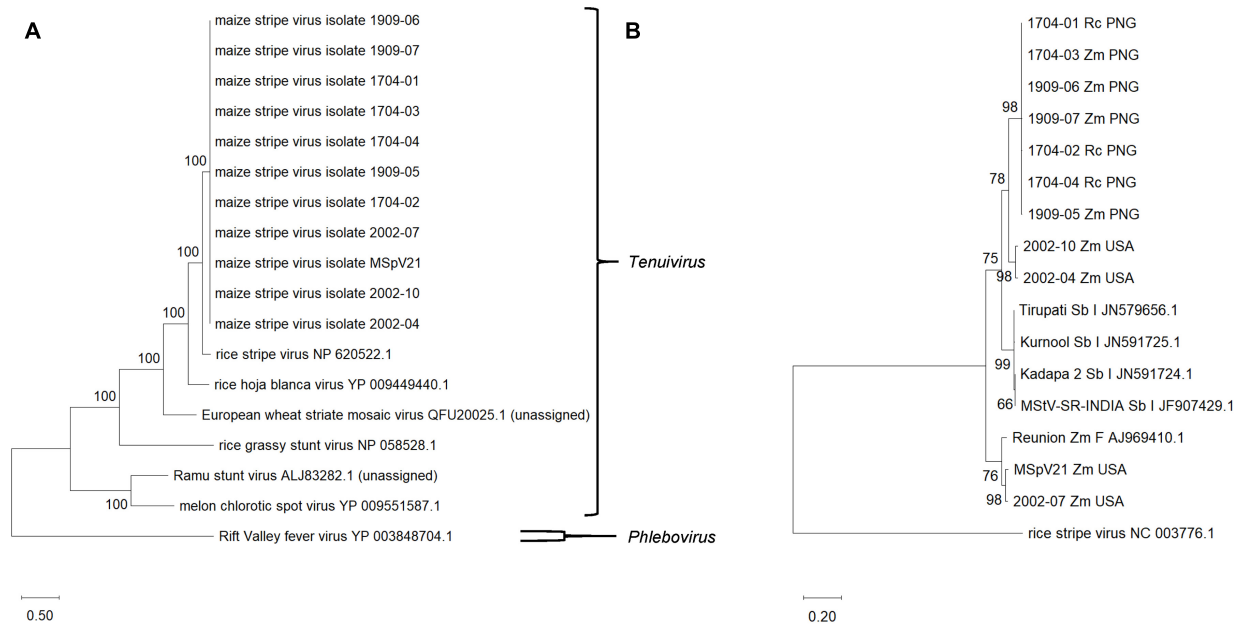


FIGURE 3 | (A) Maximum likelihood phylogenetic tree based on RNA-dependent RNA polymerase amino acid sequences from selected tenuiviruses and tenui-like viruses. Rift Valley fever virus, a phlebovirus, was selected as an outgroup. **(B)** Maximum likelihood phylogenetic tree of RNA3 for maize stripe virus isolates. RNA 3 from rice stripe virus was included as an outgroup. The National Center for Biotechnology Information accession and reference sequence numbers are provided to the right of the appropriate names. The numbers at the nodes are bootstrap values, and the scale bar represents the number of substitutions per site. Rc, *Rottboellia cochinchinensis*; Zm, *Zea mays*; Sb, *Sorghum bicolor*; PNG, Papua New Guinea; USA, United States of America; I, India; F, France.

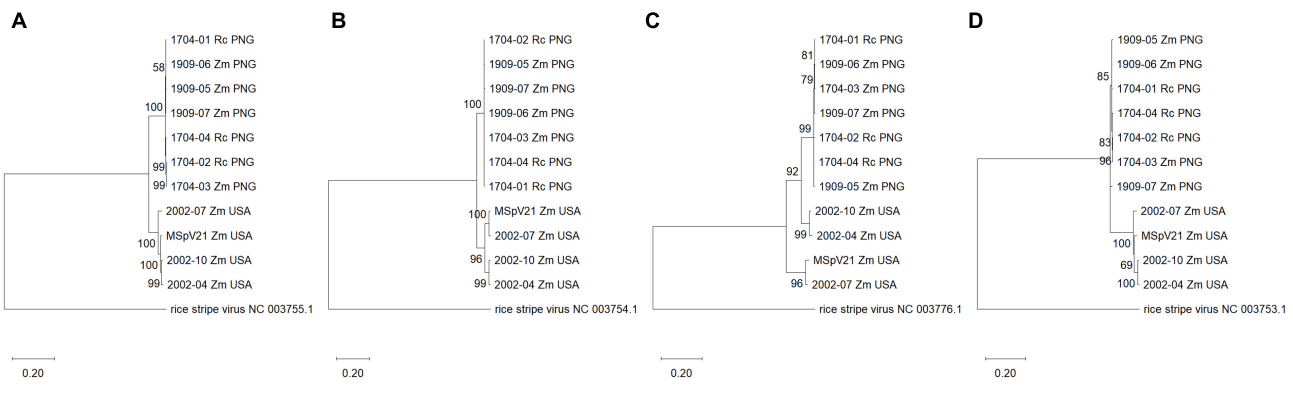


FIGURE 4 | Maximum likelihood phylogenetic trees of each RNA segment for isolates of maize stripe virus. **(A)** RNA1; **(B)** RNA2; **(C)** RNA3; **(D)** RNA4. The respective RNA segments from rice stripe virus were included as outgroups, and the corresponding National Center for Biotechnology Information reference sequence numbers are provided for these. The scale refers to the number of substitutions per site, and the numbers at the nodes are bootstrap values. Rc, *Rottboellia cochinchinensis*; Zm, *Zea mays*; PNG, Papua New Guinea; USA, United States of America.

10 nucleotides are largely conserved across RNA segments as also observed for RSV (Takahashi et al., 1990). In addition, the terminal 11–20 nucleotides were conserved within MSpV RNA segments but varied across RNA segments (Takahashi et al., 1990). The genomic MSpV RNA1 consensus 5′ terminal 5′-ACACAAAGUCCAGAGGAAAC-3′ and 3′ terminal 5′-UUUUUCCUCUGACUAUGUGU-3′ sequences are the same as those published for RNA 1 from RSV (Takahashi et al., 1990), except the nucleotide at position 20 of the 5′ end is A for RSV and C for MSpV. Overall, MSpV RNA and protein sequences are most closely related to those of RSV, except RNA5 and pc5, which are absent in sequenced genomes of RSV and are most closely related to those from Echinochloa hoja blanca virus.

We determined and described the first complete RNA1 sequence for a MSpV isolate. Characterization of the genomic sequence for RNA1 of MSpV21 revealed a sequence for pc1 in the viral complementary strand that was similar to that of pc1 from RSV. Besides having a domain with the conserved motifs of bunyavirus RdRps, the pc1 for MSpV also had predicted OTU and Endo domains. Investigation of the motifs in these domains

A

	1	2	3	4	5	6	7	8	9	10	11
1. MSpV21 Zm USA											
2. 1704-01 Rc PNG	98.1										
3. 1704-02 Rc PNG	98.1	99.7									
4. 1704-03 Zm PNG	98.0	99.7	99.8								
5. 1704-04 Rc PNG	98.0	99.7	99.7	99.7							
6. 2002-10 Zm USA	99.7	98.3	98.3	98.2	98.2						
7. 2002-07 Zm USA	99.5	98.3	98.2	98.2	98.2	99.6					
8. 2002-04 Zm USA	99.6	98.2	98.2	98.1	98.1	99.9	99.5				
9. 1909-05 Zm PNG	98.0	99.7	99.6	99.7	99.7	98.2	98.2	98.1			
10. 1909-06 Zm PNG	98.0	99.7	99.6	99.6	99.6	98.2	98.2	98.2	99.6		
11. 1909-07 Zm PNG	98.1	99.8	99.7	99.8	99.7	98.3	98.3	98.2	99.8	99.7	

B

	1	2	3	4	5	6	7	8	9	10	11
1. MSpV21 Zm USA		94.8	94.7	94.4	94.8	98.1	98.8	98.3	94.6	94.7	94.7
2. 1704-01 Rc PNG	94.4		99.9	99.6	100	94.9	94.6	95.1	99.8	99.9	99.9
3. 1704-02 Rc PNG	94.9	99.5		99.5	99.9	94.8	94.4	94.9	99.6	99.8	99.8
4. 1704-03 Zm PNG	94.9	99.5	100		99.6	94.6	94.2	94.7	99.4	99.5	99.5
5. 1704-04 Rc PNG	94.4	99.0	99.5	99.5		94.9	94.6	95.1	99.8	99.9	99.9
6. 2002-10 Zm USA	94.9	92.8	93.3	93.3	92.8		97.3	99.0	94.7	95.1	94.8
7. 2002-07 Zm USA	99.0	94.4	94.9	94.9	94.4	95.9		97.6	94.3	94.4	94.4
8. 2002-04 Zm USA	94.9	92.8	93.3	93.3	92.8	99.5	95.9		94.8	95.2	94.9
9. 1909-05 Zm PNG	94.9	99.0	99.5	99.5	99.0	93.3	94.9	93.3		99.6	99.6
10. 1909-06 Zm PNG	93.8	98.5	99.0	99.0	98.5	92.2	93.8	92.2	98.5		99.8
11. 1909-07 Zm PNG	94.9	99.5	100	100	99.5	93.3	94.9	93.3	99.5	99.0	

C

	1	2	3	4	5	6	7	8	9	10	11
1. MSpV21 Zm USA		93.5	93.5	93.5	93.5	95.1	99.7	95.1	93.5	93.5	93.5
2. 1704-01 Rc PNG	93.7		100	100	100	97.1	93.1	97.1	100	100	100
3. 1704-02 Rc PNG	93.7	100		100	100	97.1	93.1	97.1	100	100	100
4. 1704-03 Zm PNG	93.2	99.5	99.5		100	97.1	93.1	97.1	100	100	100
5. 1704-04 Rc PNG	93.7	100	100	99.5		97.1	93.1	97.1	100	100	100
6. 2002-10 Zm USA	92.6	99.0	99.0	98.5	99.0		94.8	99.4	97.1	97.1	97.1
7. 2002-07 Zm USA	99.0	94.8	94.8	94.3	94.8	93.7		94.8	93.1	93.1	93.1
8. 2002-04 Zm USA	92.6	99.0	99.0	98.5	99.0	100	93.7		97.1	97.1	97.1
9. 1909-05 Zm PNG	93.7	100	100	99.5	100	99.0	94.8	99.0		100	100
10. 1909-06 Zm PNG	93.7	100	100	99.5	100	99.0	94.8	99.0	100		100
11. 1909-07 Zm PNG	93.7	100	100	99.5	100	99.0	94.8	99.0	100	100	

D

	1	2	3	4	5	6	7	8	9	10	11
1. MSpV21 Zm USA		98.9	98.9	98.9	98.9	99.3	99.3	99.3	98.9	98.6	98.9
2. 1704-01 Rc PNG	98.9		100	100	100	98.9	98.2	98.9	100	99.6	99.3
3. 1704-02 Rc PNG	98.9	100		100	100	98.9	98.2	98.9	100	99.6	99.3
4. 1704-03 Zm PNG	98.9	100	100		100	98.9	98.2	98.9	100	99.6	99.3
5. 1704-04 Rc PNG	98.9	100	100	100		98.9	98.2	98.9	100	99.6	99.3
6. 2002-10 Zm USA	99.4	99.4	99.4	99.4	99.4		99.3	100	98.9	98.6	99.6
7. 2002-07 Zm USA	99.4	99.4	99.4	99.4	99.4	100		99.3	98.2	97.9	98.9
8. 2002-04 Zm USA	99.4	99.4	99.4	99.4	99.4	100	100		98.9	98.6	99.6
9. 1909-05 Zm PNG	98.9	100	100	100	100	99.4	99.4	99.4		99.6	99.3
10. 1909-06 Zm PNG	98.9	100	100	100	100	99.4	99.4	99.4	100		98.9
11. 1909-07 Zm PNG	98.3	99.4	99.4	99.4	99.4	98.9	98.9	98.9	99.4	99.4	

FIGURE 5 | Percent identity matrices of predicted protein sequences from each sequenced isolate of maize stripe virus. **(A)** pc1 from RNA1; **(B)** p2 (bottom) and pc2 (top) from RNA2; **(C)** p3 (bottom) and pc3 (top) from RNA3; **(D)** p4 (bottom) and pc4 (top) from RNA4. Rc, *Rottboellia cochinchinensis*; Zm, *Zea mays*; PNG, Papua New Guinea; USA, United States of America.

revealed that MSpV has the conserved elements identified in those from RSV and would, therefore, be expected to function similarly (Makarova et al., 2000; Zhao S. et al., 2019; Zhao et al., 2020). The OTU domain in RSV was shown to have deubiquitinating enzyme activity and is suspected to be involved in the autoproteolytic cleavage of pc1 (Zhao et al., 2020). The Endo domain is thought to function in cap-snatching, a function described in some bunyaviruses, where the viral RdRp cleaves capped mRNAs from the host and uses them to prime transcription of its own genes (Amroun et al., 2017; Zhao S. et al., 2019). Evidence of cap-snatching for RSV and MSpV has been previously described (Falk and Tsai, 1998; Liu et al., 2018; Lin et al., 2020).

Recombination appears to be rare in negative-sense, single-stranded RNA viruses; although for those with segmented genomes like influenza A, genetic exchange can still occur through reassortment (Simon-Loriere and Holmes, 2011). We did not find any strong signatures of recombination by RDP4 in individual alignments of our MSpV RNA segments (data not presented). However, in a concatenated RNA1-4 alignment, RDP4 did identify regions corresponding to RNA3 for 2002-04 and 2002-10 MSpV United States isolates that were suggestive of reassortment. This finding was supported by phylogenetic grouping of individual RNA segments of MSpV isolates, where MSpV isolates grouped with high bootstrap support by geographic origin for RNA segments 1, 2, and 4, but for RNA3, 2002-04 and 2002-10, MSpV United States isolates grouped separate from other United States isolates and with PNG isolates. We consider these data as strong evidence that RNA3 from 2002-04 and 2002-10 isolates are derived from an ancestral reassortment event with RNAs 1, 2, and 4 coming from a parent isolate like MSpV21 and RNA3 from an isolate like 1704-04. Differential groupings by RNA segment were also attributed to reassortment for European wheat striate mosaic virus isolates from Northern Europe (Sõmera et al., 2020) and for RSV isolates from Korea (Jonson et al., 2009a,b, 2011).

We herein report the first complete genomes of MSpV isolates that lack RNA5. By comparing RNA3 sequences of MSpV from across the world, there is distinct clustering based on geographic origin and presence/absence of RNA5. The Réunion isolate contained RNA5 based on observed RNA migration sizes (Mahmoud et al., 2007). The Kurnool isolate from India (Srinivas et al., 2014) had a RNA5 sequence deposited in GenBank (NCBI) under the accession number JN626912.1. Previous analysis of a MSpV isolate infecting sorghum in India also revealed the presence of RNA5 by migration size (Peterschmitt et al., 1991). Therefore, three groups are evident by the RNA3 phylogenetic tree: (1) The PNG and United States isolates lacking RNA5, (2) the India isolates infecting sorghum, and (3) the Réunion and United States isolates containing RNA5.

Since the function of pc5 from RNA5 has not been established in tenuiviruses, it is difficult to speculate how some isolates of MSpV accommodate its absence. RSV lacks pc5 and can infect maize (Gingery et al., 1983; Bradfute and Tsai, 1990); therefore, we hypothesize that pc5 does not have a deterministic role in maize infection. It may, however, influence vector infection and transmission efficiency. Differences in MSpV transmission efficiency by its vector *P. maidis* were already noted in a previous study, where *P. maidis* from United States (Hawaii) transmitted MSpV isolates from Costa Rica and Nigeria more efficiently than an isolate from the United States (Florida) (Ammar et al., 1995). Our data indicate that between United States isolates with and without pc5, there were mostly changes in the coding regions of p2, p3, and pc3. Although the exact mechanisms of function for p2 have not been completely characterized, it may function *in planta* for RSV as a weak silencing suppressor by

binding to a rice suppressor of gene silencing and targeting a silencing amplification pathway specific to plants (Du et al., 2011). RSV p2 may also promote systemic movement of RSV *in planta* by interacting with fibrillarin (Zheng et al., 2015). The p3 protein from RSV appears to have a more general silencing suppression function through binding of dsRNA (Shen et al., 2010). Indeed, silencing suppression was demonstrated for p3 of rice hoja blanca virus in both plant and insect cells (Hemmes et al., 2007). The pc3 protein is the nucleoprotein, and it has been shown to be expressed in *P. maidis* (Falk et al., 1987). The pc3 protein also colocalized with proteins that play essential roles in the transmission efficiency and transovarial transmission of RSV by its insect vector (Huo et al., 2014; Liu et al., 2015). Based on existing literature on p3 and pc3 and our data on the RNA3 phylogenetic grouping by presence/absence of RNA5, it is tempting to speculate that an ancestral reassortment event of RNA3 in 2002-04 and 2002-10 isolates helped to accommodate their loss of RNA5. Further sampling and sequencing of MSpV isolates worldwide and the molecular characterization of pc5 are needed to clarify the relationship between MSpV core RNAs1-4 and RNA5.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) listed in **Table 1** can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MW328593; https://www.ncbi.nlm.nih.gov/genbank/, MW328594; https://www.ncbi.nlm.nih.gov/genbank/, MW328595; https://www.ncbi.nlm.nih.gov/genbank/, MW 328596; and https://www.ncbi.nlm.nih.gov/genbank/, MW328597.

AUTHOR CONTRIBUTIONS

DM and BF: conceptualization. SB, SG, and IF-B: methodology. SG: software. SB and IF-B: validation. SB, SG, and DM: formal analysis and data curation. SB, SG, DM, IF-B, BF, KB, and RB: investigation. DM, BF, KB, and RB: resources. SB and DM: writing—original draft preparation. SB, DM, BF, SG, KB, RB, and IF-B: writing—review and editing. SB, DM, and BF: visualization. DM: supervision, project administration, and funding acquisition. All authors read and agreed to the published version of the manuscript.

FUNDING

This work was supported through U.S. Department of Agriculture Funding, Agricultural Research Service, Research Project 8042-22000-302-00-D.

ACKNOWLEDGMENTS

We would like to thank Leka Tom and Lastus Kuniata, Ramu Agri Industries Limited, for organizing the surveys and assisting with sample collection in PNG.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.684599/full#supplementary-material

REFERENCES

Abudurexiti, A., Adkins, S., Alioto, D., Alkhovsky, S. V., Avšič-Županc, T., Ballinger, M. J., et al. (2019). Taxonomy of the order Bunyavirales: update 2019. *Arch. Virol.* 164, 1949–1965. doi: 10.1007/s00705-019-04253-6

Ammar, E. D., Gingery, R. E., and Madden, L. V. (1995). Transmission efficiency of three isolates of maize stripe tenuivirus in relation to virus titre in the planthopper vector. *Plant Pathol.* 44, 239–243. doi: 10.1111/j.1365-3059.1995.tb02774.x

Amroun, A., Priet, S., De Lamballerie, X., and Quérat, G. (2017). Bunyaviridae RdRps: structure, motifs, and RNA synthesis machinery. *Crit. Rev. Microbiol.* 43, 753–778. doi: 10.1080/1040841X.2017.1307805

Bradfute, O., and Tsai, J. (1990). Rapid identification of maize stripe virus. *Phytopathology* 80, 715–719.

Bucher, E., Sijen, T., De Haan, P., Goldbach, R., and Prins, M. (2003). Negative-strand tospoviruses and tenuiviruses carry a gene for a suppressor of gene silencing at analogous genomic positions. *J. Virol.* 77, 1329–1336. doi: 10.1128/jvi.77.2.1329-1336.2003

Chen, C., Tsai, J., Chiu, R., and Chen, M. (1993). Purification, characterization, and serological analysis of maize stripe virus in Taiwan. *Plant Dis.* 77, 367–372.

De Doyle, M. M. R., Autrey, L. J. C., and Jones, P. (1992). Purification, characterization and serological properties of two virus isolates associated with the maize stripe disease in Mauritius. *Plant Pathol.* 41, 325–334. doi: 10.1111/j.1365-3059.1992.tb02354.x

De Miranda, J., Hernandez, M., Hull, R., and Espinoza, A. M. (1994). Sequence analysis of rice hoja blanca virus RNA 3. *J. Gen. Virol.* 75, 2127–2132. doi: 10.1099/0022-1317-75-8-2127

Du, Z., Xiao, D., Wu, J., Jia, D., Yuan, Z., Liu, Y., et al. (2011). p2 of Rice stripe virus (RSV) interacts with OsSGS3 and is a silencing suppressor. *Mol. Plant Pathol.* 12, 808–814. doi: 10.1111/j.1364-3703.2011.00716.x

Estabrook, E. M., Suyenaga, K., Tsai, J. H., and Falk, B. W. (1996). Maize stripe tenuivirus RNA2 transcripts in plant and insect hosts and analysis of pvc2, a protein similar to the Phlebovirus virion membrane glycoproteins. *Virus Genes* 12, 239–247. doi: 10.1007/BF00284644

Falk, B., and Tsai, J. (1984). Identification of single-and double-stranded RNAs associated with maize stripe virus. *Phytopathology* 74, 909–915.

Falk, B. W., and Tsai, J. H. (1998). Biology and molecular biology of viruses in the genus Tenuivirus. *Annu. Rev. Phytopathol.* 36, 139–163. doi: 10.1146/annurev.phyto.36.1.139

Falk, B. W., Tsai, J. H., and Lommel, S. A. (1987). Differences in levels of detection for the maize stripe virus capsid and major non-capsid proteins in plant and insect hosts. *J. General Virol.* 68, 1801–1811. doi: 10.1099/0022-1317-68-7-1801

Fu, S., Xu, Y., Li, C., Li, Y., Wu, J., and Zhou, X. (2018). Rice stripe virus interferes with s-acylation of remorin and induces its autophagic degradation to facilitate virus infection. *Mol. Plant* 11, 269–287. doi: 10.1016/j.molp.2017.11.011

Gerlach, P., Malet, H., Cusack, S., and Reguera, J. (2015). Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell* 161, 1267–1279. doi: 10.1016/j.cell.2015.05.006

Gingery, R., Re, G., and Rj, L. (1979). Occurrence of maize stripe virus in the United States and Venezuela. *Plant Dis. Reporter* 63, 341–343.

Gingery, R. E., Nault, L. R., and Bradfute, O. E. (1981). Maize stripe virus: characteristics of a member of a new virus class. *Virology* 112, 99–108. doi: 10.1016/0042-6822(81)90616-4

Gingery, R. E., Nault, L. R., and Yamashita, S. (1983). Relationship between maize stripe virus and rice stripe virus. *J. Gen. Virol.* 64, 1765–1770. doi: 10.1099/0022-1317-64-8-1765

Greber, R. (1981). Maize stripe disease in Australia. *Australian J. Agric. Res.* 32, 27–36. doi: 10.1071/AR9810027

Hemmes, H., Lakatos, L., Goldbach, R., Burgyán, J., and Prins, M. (2007). The NS3 protein of Rice hoja blanca tenuivirus suppresses RNA silencing in plant and insect hosts by efficiently binding both siRNAs and miRNAs. *RNA* 13, 1079–1089. doi: 10.1261/rna.444007

Huiet, L., Klaassen, V., Tsai, J. H., and Falk, B. W. (1990). Identification and sequence analysis of the maize stripe virus major noncapsid protein gene. *Virology* 179, 862–866. doi: 10.1016/0042-6822(90)90156-L

Huiet, L., Klaassen, V., Tsai, J. H., and Falk, B. W. (1991). Nucleotide sequence and RNA hybridization analyses reveal an ambisense coding strategy for maize stripe virus RNA3. *Virology* 182, 47–53. doi: 10.1016/0042-6822(91)90646-S

Huiet, L., Tsai, J. H., and Falk, B. W. (1992). Complete sequence of maize stripe virus RNA4 and mapping of its subgenomic RNAs. *J. Gen. Virol.* 73, 1603–1607. doi: 10.1099/0022-1317-73-7-1603

Huiet, L., Tsai, J. H., and Falk, B. W. (1993). Maize stripe virus RNA5 is of negative polarity and encodes a highly basic protein. *J. Gen. Virol.* 74, 549–554. doi: 10.1099/0022-1317-74-4-549

Huo, Y., Liu, W., Zhang, F., Chen, X., Li, L., Liu, Q., et al. (2014). Transovarial transmission of a plant virus is mediated by vitellogenin of its insect vector. *PLoS Pathogens* 10:e1003949. doi: 10.1371/journal.ppat.1003949

Jonson, M. G., Choi, H.-S., Kim, J.-S., Choi, I.-R., and Kim, K.-H. (2009a). Complete genome sequence of the RNAs 3 and 4 segments of Rice stripe virus isolates in Korea and their phylogenetic relationships with Japan and China isolates. *Plant Pathol. J.* 25, 142–150.

Jonson, M. G., Choi, H.-S., Kim, J.-S., Choi, I.-R., and Kim, K.-H. (2009b). Sequence and phylogenetic analysis of the RNA1 and RNA2 segments of Korean Rice stripe virus isolates and comparison with those of China and Japan. *Arch. Virol.* 154:1705. doi: 10.1007/s00705-009-0493-7

Jonson, M. G., Lian, S., Choi, H.-S., Lee, G.-S., Kim, C.-S., and Kim, K.-H. (2011). Genetic reassortment of Rice stripe virus RNA segments detected by RT-PCR restriction enzyme analysis-based method. *Plant Pathol. J.* 27, 148–155.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lin, W., Wu, R., Qiu, P., Jing, J., Yang, Y., Wang, J., et al. (2020). A convenient in vivo cap donor delivery system to investigate the cap snatching of plant bunyaviruses. *Virology* 539, 114–120. doi: 10.1016/j.virol.2019.10.017

Liu, W., Gray, S., Huo, Y., Li, L., Wei, T., and Wang, X. (2015). Proteomic analysis of interaction between a plant virus and its vector insect reveals new functions of hemipteran cuticular protein. *Mol. Cellular Proteomics* 14, 2229–2242. doi: 10.1074/mcp.M114.046763

Liu, X., Jin, J., Qiu, P., Gao, F., Lin, W., Xie, G., et al. (2018). Rice stripe tenuivirus has a greater tendency to use the prime-and-realign mechanism in transcription of genomic than in transcription of antigenomic template RNAs. *J. Virol.* 92, e01414–e01417. doi: 10.1128/jvi.01414-17

Lu, G., Li, S., Zhou, C., Qian, X., Xiang, Q., Yang, T., et al. (2019). Tenuivirus utilizes its glycoprotein as a helper component to overcome insect midgut barriers for its circulative and propagative transmission. *PLoS Pathogens* 15:e1007655. doi: 10.1371/journal.ppat.1007655

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2019). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Mahmoud, A., Royer, M., Granier, M., E-D, A., and Peterschmitt, M. (2007). High genetic identity between RNA 3 segments of an Old World isolate and a New World isolate of Maize stripe virus. *Arch. Virol.* 152, 1583–1586. doi: 10.1007/s00705-007-0981-6

Makarova, K. S., Aravind, L., and Koonin, E. V. (2000). A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem. Sci.* 25, 50–52. doi: 10.1016/S0968-0004(99)01530-3

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

Nault, L., and Gordon, D. (1988). Multiplication of maize stripe virus in *Peregrinus maidis*. *Phytopathology* 78, 991–995.

Otuka, A., Matsumura, M., Sanada-Morimura, S., Takeuchi, H., Watanabe, T., Ohtsu, R., et al. (2010). The 2008 overseas mass migration of the small brown planthopper, *Laodelphax striatellus*, and subsequent outbreak of rice stripe disease in western Japan. *Appl. Entomol. Zool.* 45, 259–266. doi: 10.1303/aez.2010.259

Peterschmitt, M., Chatenet, M., and Baudin, P. (1987). Application de la méthode ELISA au diagnostic des viroses du maïs. *L'Agronomie Tropicale (1975)* 42, 131–138.

Peterschmitt, M., Ratna, A. S., Sacks, W. R., Reddy, D. V. R., and Mughogho, L. K. (1991). Occurrence of an isolate of maize stripe virus on sorghum in India. *Ann. Appl. Biol.* 118, 57–70. doi: 10.1111/j.1744-7348.1991.tb06085.x

Ramírez, B.-C., and Haenni, A.-L. (1994). Molecular biology of tenuiviruses, a remarkable group of plant viruses. *J. Gen. Virol.* 75, 467–475. doi: 10.1099/0022-1317-75-3-467

Sdoodee, R., Teakle, D. S., and Louie, R. (1997). Preliminary identification of maize stripe tenuivirus in Thailand. *Plant Dis.* 81, 228–228. doi: 10.1094/pdis.1997.81.2.228b

Shen, M., Xu, Y., Jia, R., Zhou, X., and Ye, K. (2010). Size-independent and noncooperative recognition of dsRNA by the rice stripe virus RNA silencing suppressor NS3. *J. Mol. Biol.* 404, 665–679. doi: 10.1016/j.jmb.2010.10.007

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

Singh, B. U., and Seetharama, N. (2008). Host plant interactions of the corn planthopper, *Peregrinus maidis* Ashm. (Homoptera: delphacidae) in maize and sorghum agroecosystems. *Arthropod-Plant Interactions* 2, 163–196. doi: 10.1007/s11829-007-9026-z

Sõmera, M., Kvarnheden, A., Desbiez, C., Blystad, D.-R., Sooväli, P., Kundu, J. K., et al. (2020). Sixty years after the first description: genome sequence and biological characterization of European wheat striate mosaic virus infecting cereal crops. *Phytopathology* 110, 68–79. doi: 10.1094/phyto-07-19-0258-fi

Srinivas, K. P., Sreekanth Reddy, M., Subba Reddy, C. R. V., Hema, M., and Sreenivasulu, P. (2014). Sequence analysis of RNA3 of Maize stripe virus associated with stripe disease of sorghum (Sorghum bicolor) in India. *Phytopathologia Mediterranea* 53, 188–193.

Storey, H. H. (1936). Virus diseases of east African plants. *East Afr. Agric. J.* 1, 333–337. doi: 10.1080/03670074.1936.11663679

Takahashi, M., Toriyama, S., Kikuchi, Y., Hayakawa, T., and Ishihama, A. (1990). Complementarity between the 5'- and 3'-terminal sequences of rice stripe virus RNAs. *J. Gen. Virol.* 71, 2817–2821. doi: 10.1099/0022-1317-71-12-2817

Toriyama, S., Takahashi, M., Sano, Y., Shimizu, T., and Ishihama, A. (1994). Nucleotide sequence of RNA 1, the largest genomic segment of rice stripe virus, the prototype of the tenuiviruses. *J. Gen. Virol.* 75, 3569–3579. doi: 10.1099/0022-1317-75-12-3569

Tsai, J. H., and Zitter, T. A. (1982). Characteristics of maize stripe virus transmission by the corn delphacid. *J. Econ. Entomol.* 75, 397–400. doi: 10.1093/jee/75.3.397

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281–285.

Wang, H.-D., Chen, J.-P., Zhang, H.-M., Sun, X.-L., Zhu, J.-L., Wang, A.-G., et al. (2008). Recent rice stripe virus epidemics in zhejiang province, China, and experiments on sowing date, disease–yield loss relationships, and seedling susceptibility. *Plant Dis.* 92, 1190–1196. doi: 10.1094/pdis-92-8-1190

Xiong, R., Wu, J., Zhou, Y., and Zhou, X. (2008). Identification of a movement protein of the *Tenuivirus* rice stripe virus. *J. Virol.* 82, 12304–12311. doi: 10.1128/jvi.01696-08

Xiong, R., Wu, J., Zhou, Y., and Zhou, X. (2009). Characterization and subcellular localization of an RNA silencing suppressor encoded by Rice stripe tenuivirus. *Virology* 387, 29–40. doi: 10.1016/j.virol.2009.01.045

Zhang, H. M., Yang, J., Sun, H. R., Xin, X., Wang, H. D., Chen, J. P., et al. (2007). Genomic analysis of rice stripe virus Zhejiang isolate shows the presence of an OTU-like domain in the RNA1 protein and a novel sequence motif conserved within the intergenic regions of ambisense segments of tenuiviruses. *Arch. Virol.* 152, 1917–1923. doi: 10.1007/s00705-007-1013-2

Zhao, S., Gu, X., Li, J., and Liang, C. (2020). The N-terminal cysteine protease domain of rice stripe tenuivirus Pc1 possesses deubiquitinating enzyme activity. *Virus Genes* 57, 117–120. doi: 10.1007/s11262-020-01807-8

Zhao, S., Xu, G., He, G., Peng, Y., and Liang, C. (2019). Characterization of an endonuclease in rice stripe tenuivirus Pc1 in vitro. *Virus Res.* 260, 33–37. doi: 10.1016/j.virusres.2018.11.006

Zhao, W., Wang, Q., Xu, Z., Liu, R., and Cui, F. (2019). Distinct replication and gene expression strategies of the *Rice Stripe virus* in vector insects and host plants. *J. Gen. Virol.* 100, 877–888. doi: 10.1099/jgv.0.001255

Zheng, L., Du, Z., Lin, C., Mao, Q., Wu, K., Wu, J., et al. (2015). Rice stripe tenuivirus p2 may recruit or manipulate nucleolar functions through an interaction with fibrillarin to promote virus systemic movement. *Mol. Plant Pathol.* 16, 921–930. doi: 10.1111/mpp.12220

Conflict of Interest: KB was employed by Sugar Research Australia Limited, Indooroopilly, QLD, Australia.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 June 2021
doi: 10.3389/fmicb.2021.702210

Spatial Virome Analysis of *Zanthoxylum armatum* Trees Affected With the Flower Yellowing Disease

Mengji Cao[1†], Song Zhang[1†], Ruiling Liao[1], Xiaoru Wang[1], Zhiyou Xuan[1], Binhui Zhan[2], Zhiqi Li[3], Jie Zhang[4], Xinnian Du[5], Zhengsen Tang[5], Shifang Li[2,6*] and Yan Zhou[1*]*

[1] National Citrus Engineering Research Center, Citrus Research Institute, Southwest University, Chongqing, China, [2] State Key Laboratory of Biology for Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China, [3] Jiangjin Agricultural Technology Extension Station, Chongqing, China, [4] Bishan Modern Agricultural Development Promotion Center, Chongqing, China, [5] Zhaotong Forestry and Grassland Pest Monitoring and Testing Center, Yunnan, China, [6] Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

OPEN ACCESS

Edited by:
Xiaofei Cheng,
Northeast Agricultural University, China

Reviewed by:
Won Kyong Cho,
Seoul National University, South Korea
Hongguang Cui,
Hainan University, China

***Correspondence:**
Mengji Cao
caomengji@cric.cn
Shifang Li
sfli@ippcaas.cn
Yan Zhou
zhouyan@cric.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: 29 April 2021
Accepted: 07 June 2021
Published: 28 June 2021

Citation:
Cao M, Zhang S, Liao R, Wang X, Xuan Z, Zhan B, Li Z, Zhang J, Du X, Tang Z, Li S and Zhou Y (2021) Spatial Virome Analysis of Zanthoxylum armatum Trees Affected With the Flower Yellowing Disease. Front. Microbiol. 12:702210. doi: 10.3389/fmicb.2021.702210

Zanthoxylum armatum is an important woody crop with multiple applications in pharmaceutics, cosmetics, and food industries. With continuous increases in the plantation area, integrated pest management is required for scale production when diseases caused by biotic factors such as pests and pathogens have become new problems, one of which is the infectious flower yellowing disease (FYD). Here, isolates of a new illarvirus (3) and a new nepovirus-associated subviral satellite RNA (12) were identified in *Z. armatum*, in addition to 38 new isolates of four previously reported RNA viruses. Sequence variation can be observed in viral/subviral quasispecies and among predominant isolates from the same or different samples and geographic origins. Intriguingly, RNA sequencing of different diseased trees invariably showed an extraordinary pattern of particularly high reads accumulation of the green Sichuan pepper-nepovirus (GSPNeV) and the satellite RNA in symptomatic tissues. In addition, we also examined small RNAs of the satellite RNA, which show similar patterns to those of coinfecting viruses. This study provides further evidence to support association of the FYD with viral/subviral infections and deepens our understanding of the diversity and molecular characteristics of the viruses and satellite, as well as their interactions with the host.

Keywords: *Zanthoxylum* viruses, *Ilarvirus*, satellite RNA, RNA-seq, small RNAs

INTRODUCTION

Zanthoxylum, one of the most economically important genera in the family Rutaceae, comprises deciduous, spiny shrub species (Ekka et al., 2020). Some species, the so-called Chinese prickly ash, are well-known in China, especially in southwestern regions, for large-scale cultivation for production of spices, medicines, and essential/edible oils (Zhang et al., 2017; Ma et al., 2020). *Zanthoxylum armatum* (Zhuye huajiao in Chinese), also known as green Sichuan (Szechwan) pepper, is a commercial species that

generates green fruits that can be used as a seasoning for its special aroma and numbing flavor (Xu et al., 2019). There is one representative production area and trade center in Chongqing Municipality, that is, Jiangjin District, where *Z. armatum* var. *novemfolius* as a native cultivar with an annual output value of nearly half a billion dollars has been widely planted. *Z. armatum* orchards are generally monocultural with one or few commercial cultivars planted locally, which, despite convenient to manage, are prone to be affected by disease outbreaks. In recent years, a virus-like disease, the flower yellowing disease (FYD), that leads to a severe disorder of the floral organs, has emerged as the main restriction factor for the industrial development of *Z. armatum* (Zhang et al., 2020). The FYD is typically characterized by symptoms of pistil abortion, stamen yellowing and intumescence, usually with yellowing and stunting of the foliage, so as to be extremely destructive to fructification of the affected plants (Cao et al., 2019). This disease may occur first on a few branches and then extends progressively to the whole plant, causing irreversible tree decline and eventually death, thus resulting in huge economic losses. Similar problems also arise in other *Zanthoxylum* species, including *Z. bungeanum*. However, Koch's postulates are yet to be fulfilled to explain the etiology of whether a pathogen is involved.

With its unbias and high sensitivity, high-throughput sequencing (HTS) has been extensively applied for identification and detection of viruses present in various organisms, such as the research on all plant viruses in samples from single or multiple plant species under specific conditions (Coetzee et al., 2010; Massart et al., 2014; Bernardo et al., 2018; Maclot et al., 2020). The global plant virome has revealed extensive viral diversity, allowing us to gain insight into the roles these viruses play in ecosystems across scales, from simple cytozoic parasites to the essential impetus of plant evolution (Dolja et al., 2020). For instance, plant viruses can easily have adverse effects on agroecosystem due to their pathogenicity to sensitive hosts, and there is an epidemic risk in genetically uniform crops in the presence of biological vectors as vehicles (Gilbertson et al., 2015).

With the aid of HTS, we have previously shown diverse viruses in *Zanthoxylum* species that are transmissible among trees in agricultural environments and possibly pathogenic, and thus they are potential threats to the sustainability of the industry (Cao et al., 2019). The FYD is possibly caused by a nepovirus, green Sichuan pepper-nepovirus (GSPNeV), which is frequently accompanied by one or several other viruses in the plant. It seems that GSPNeV is present mainly in symptomatic branches in a tree before it disturbs asymptomatic branches and causes symptoms, based on the results of reverse-transcription polymerase chain reaction (RT-PCR). In the present work, this phenomenon was further studied by HTS analysis, and this showed a large number of reads derived from GSPNeV in symptomatic branches but very few reads in asymptomatic branches, thus supporting an uneven viral/subviral distribution and likely suggesting its association with the disease pattern of non-uniform symptoms. Moreover, a deep analysis regarding co-evolution and co-transmission of the viral/subviral RNAs and the dynamics of their sequence variation was also conducted.

MATERIALS AND METHODS

Sample Collection

A survey was first conducted in orchards to identify *Zanthoxylum* trees affected by the flower yellowing disease (FYD) with symptomatic and asymptomatic branches on a single plant. Six samples (foliar and floral tissues mixed for each) from symptomatic and asymptomatic branches of three selected partial diseased *Z. armatum* trees, one from Yunnan Province and two from Chongqing Municipality, were collected for RNA sequencing (RNA-seq) and subsequent comparative analyses of viral/subviral species and accumulation. In addition, 14 samples, each from a completely diseased tree or tree without symptoms as a negative control were incorporated for viral/subviral diversity and variation analyses. In total, 20 samples were separately sequenced for etiology study of the FYD, in addition to a diseased tree (marked here as CQ1-D) with RNA-seq data from a previous study (Cao et al., 2019). Among these 21 samples, six were from Zhaotong City in Yunnan (YN) Province (YN1-DS, YN1-DA, YN2-D, YN3-H, YN4-H, YN5-H), while the others were from Chongqing (CQ) Municipality, one in Bishan District (CQ4-D), two in Jiangjin District (CQ1-D and CQ2-D), five in Tongnan District (CQ3-H, CQ5-D, CQ6-H, CQ7-DS, CQ7-DA), and seven in Changshou District (CQ8-D, CQ9-D, CQ10-D, CQ11-D, CQ12-H, CQ13-DS, CQ13-DA). Among these, symptomatic samples from completely diseased trees were indicated with D (diseased), symptomatic samples from partial diseased trees were indicated with DS (diseased-symptomatic) while from the same trees asymptomatic samples indicated with DA (diseased-asymptomatic), and samples from trees without any symptoms were tentatively marked with H (healthy). The sample information is shown in **Table 1**.

High-Throughput Sequencing and Data Processing

The collected tissues of each sample were used for total RNA extraction by the EASY spin Plus Complex Plant RNA Kit (Aidlab, Beijing, China). The RNA was tested using a Nanodrop (Thermo Fisher Scientific, Waltham, United States) and agarose-gel electrophoresis to ensure a high quality. After ribosome RNA depletion by the RiboZero Magnetic Kit (Epicenter, Madison, United States), a library was built with the TruSeq RNA Sample Prep Kit (Illumina, San Diego, United States). The treated RNA was then sequenced by the Mega Genomics Company (Beijing, China) using an Illumina HiSeq X-ten platform with 150 bp layout in paired-end read length. Total small RNA (sRNA) was extracted with the EASYspin Plant microRNA Extract Kit (Aidlab, Beijing, China), constructed as a library using the TruSeq Small RNA Sample Prep Kit (Illumina) and sequenced on an Illumina Hiseq2500 platform with a length of 50 bp per reading (Mega Genomics). A series of built-in programs in the CLC Genomics Workbench 11 (Qiagen, Hilden, Germany) were employed to process generated RNA-seq and sRNA-seq raw data in the

TABLE 1 | Sampling information, sequencing data size, and proportion (%) of viral/subviral RNA reads in total reads.

Sample	**[a]YN1-DS**	**YN1-DA**	**YN2-D**	**YN3-H**	**YN4-H**	**YN5-H**	**CQ1-D**	**CQ2-D**	**CQ3-H**	**CQ4-D**	**CQ5-D**	**CQ6-H**	**CQ7-DS**	**CQ7-DA**
Location			**[b]ZT**				**JJ**		**TN**	**BS**			**TN**	
SRA accessions	SRR14663392	SRR14663391	SRR14663380	SRR14663378	SRR14663377	SRR14663376	SRR14663375	SRR14663374	SRR14663373	SRR14663372	SRR14663390	SRR14663389	SRR14663388	SRR14663387
Total reads	61,008,212	60,539,870	67,129,108	64,411,510	76,455,572	56,140,062	58,077,354	57,580,796	56,724,560	53,849,114	54,642,862	57,458,864	49,756,462	64,684,982
GSPNeV_RNA1	8.32%	0	4.46%	0	0	0	3.16%	4.53%	0	5.75%	4.30%	0	3.69%e	0
GSPNeV_RNA2	9.24%	0	4.39%	0	0	0	4.02%	5.14%	0	8.62%	7.33%	0	5.89%	0
SatGSPNeV	13.34%	0	3.62%	0	0	0	8.91%	8.76%	0	20.33%	13.42%	0	11.72%	0
GSPIdV_RNA1	0.21%	0	0.006%	0	0	0.007%	1.57%	0.7%	0	0.95%	1.72%	0	2.89%	1.1%
GSPIdV_RNA2	0.14%	0	0.02%	0	0	0.05%	3.73%	2.49%	0	2.19%	4.64%	0	3.49%	2.91%
GSPNuV_RNA	0	0	0	0	0	0	0.06%	0	0	0	0	0	0	0
GSPEV_RNA	0	0	0.94%	0.004%	0.005%	0.07%	0.4%	0.33%	0.60%	0.03%	0.01%	0.06%	0	0.01%
GSPIlV_RNA1	0.08%	0	0.09%	0	0	0	0	0	0	0	0	0	0.73%	0
GSPIlV_RNA2	0.07%	0	0.08%	0	0	0	0	0	0	0	0	0	0.7%	0
GSPIlV_RNA3	0.28%	0	0.42%	0	0	0	0	0	0	0	0	0	3.74%	0.0008%

Sample	**CQ8-D**	**CQ9-D**	**CQ10-D**	**CQ11-D**	**CQ12-H**	**CQ13-DS**	**CQ13-DA**
Location				**CS**			
SRA accessions	SRR14663386	SRR14663385	SRR14663384	SRR14663383	SRR14663382	SRR14663381	SRR14663379
Total reads	68,344,842	62,910,592	70,511,028	80,552,614	58,534,686	71,598,316	84,446,876
GSPNeV_RNA1	7.12%	5.47%	5.3%	5.32%	0	6.36%	0
GSPNeV_RNA2	12.74%	9.61%	11.8%	9.91%	0	12.4%	0.0002%
SatGSPNeV	12.12%	18.35%	16.55%	0	16.89%	0.0002%	
GSPIdV_RNA1	0.52%	0.9%	1.6%	0.69%	0	1.45%	9.52%
GSPIdV_RNA2	1.02%	2.7%	3.22%	2.02%	0	6.05%	1.11%
GSPNuV_RNA	0	0	0	0	0	0	0
GSPEV_RNA	6.36%	2.2%	6.04%	6.62%	1.09%	0.35%	0.21%
GSPIlV_RNA1	0	0	0	0	0	0	0
GSPIlV_RNA2	0	0	0	0	0	0	0
GSPIlV_RNA3	0	0	0	0	0	0	0

[a]Yunnan (YN) Province, Chongqing (CQ) Municipality; D, symptomatic sample from diseased (D) tree with systematic symptoms; DS, symptomatic (S) sample from diseased (D) tree with partial symptoms; DA, asymptomatic (A) sample from the same diseased (D) tree with partial symptoms; H, sample from tree without symptoms that appears to be healthy (H).

[b]Zhaotong (ZT) City, Jiangjin (JJ) District, Tongnan (TN) District, Bishan (BS) District, Changshou (CS) District.

following steps: (1) remove adaptors and low-quality reads, and the host-related reads by mapping to draft genomes of citrus within Rutaceae family (*Zanthoxylum* genomes are publicly unavailable) used as references (Xu et al., 2013); (2) *de novo* assemble the remaining reads into contigs using *De Novo* Assembly program with default parameters of *de Bruijn* graph word size 20, and minimum contig length 200 bp; and (3) annotate the contigs using a local BLASTx program with the viral sequence database (taxid:10239) downloaded from the National Center for Biotechnology Information (NCBI) used as a target.

Verification of Viral/Subviral Sequence

To exclude the possibility of chimeric viral/subviral sequences that result from assembly program algorithm error and false virus-positive results caused by contamination during HTS, contigs of the new ilarvirus from CQ7-DS and the satellite from CQ1-D were verified by RT-PCR using the leaf RNA extract, viral/subviral contig-specific primers (**Supplementary Table 1**), and a One-step RT-PCR Kit (Takara, Otsu, Japan). Specifically, the terminal sequences of the satellite were obtained using a commercial RACE Kit (Invitrogen, Carlsbad, United States). The amplicons were purified through the Gel Extraction Kit (Omega Bio-Tek, Norcross, United States), cloned with the pEasy-T1 Vector System (TransGen, Beijing, China), and sequenced by the TsingKe company (Beijing, China)—five clones per viral/subviral PCR amplicon. The resulting sequences were merged using the SeqMan (DNASTAR, Madison, United States), and assembled sequences were submitted to NCBI-GenBank and assigned specific accession numbers (see section "Data Availability Statement"). Subsequently, these sequences and available sequences from databases were used as reference genomes, onto which the corresponded viral/subviral reads of each sequencing data were mapped to obtain consensus sequences between mapped reads and the referenced sequences. The consensus sequences were compared with corresponding viral/subviral contigs generated from the assembly of independent reads of all data to ensure sequence fidelity, and were also submitted to GenBank. This comparison analysis was aimed at eliminating the interference of viral/subviral variation and recombination with the assembly process; it requires a high nt identity (99%) between two sequences; if not satisfied, it is necessary to confirm the consensus sequence using molecular cloning and sequencing.

Virome and Sequence Analyses

A heatmap of the read numbers of viral/subviral RNAs in different samples (clustering by the squared Euclidean distance) was drawn by Heat Map with Dendrogram extension in Origin 2017 software (OriginLab, Northampton, United States), and a flower plot of common (center) or unique (petals) viral/subviral RNAs across symptomatic or symptomatic/healthy samples was drawn by R (ver. 4.0.2) using draw.ellipse and draw.circle functions according to previous research (Sugawara et al., 2013). Open reading frames (ORFs) in viral/subviral sequences were predicted using the NCBI ORF finder[1], and conserved protein domains were searched with the Conserved Domain Database (CDD) web tool[2]. Sequence comparison was conducted in the CLC Genomics Workbench 11 to identify conserved sites among viral/subviral sequences.

Phylogenic and Population Analyses

Viral/subviral nucleotide (nt) and amino acid (aa) sequences from this study and the GenBank databank were aligned using MAFFT 7 (Katoh and Standley, 2013); poorly aligned regions were trimmed using trimAl (Capella-Gutiérrez et al., 2009), and the remaining regions were imported into MEGA-X (Kumar et al., 2018) to select the best-suited model under the Bayesian information criterion. The maximum-likelihood phylogenetic relationships (500 bootstrap replications) were inferred, and the four rate categories of a discrete Gamma distribution were applied if necessary (Yang, 1996). RDP4 (Martin et al., 2015), SDT 1.2 (Muhire et al., 2014), the default Alignments tool, and Low Variant Detectors in the CLC Genomics Workbench 11 were used in viral/subviral population comparison for recombination, sequence identity matrix, sequence alignment (the identity plot), and variation analysis (the latter two were used to study evolutionary hotspots), respectively. A histogram of the sequence variation was visualized with the ggplot2 package 3.3.2 in R (Wickham, 2011). A variation heatmap of quasispecies was drawn with the pheatmap package 1.0.12 in R (Kolde and Kolde, 2015). The network was drawn by Cytoscape 3.8.0 to indicate virus phylogenetic-geographical relationships (Shannon et al., 2003), where greater phylogenetic incongruity between two viral/subviral RNAs and their more specific geographic origins increased the edge thickness of their nodes (a thick line indicates high confidence of viral/subviral RNA co-evolution or co-transmission), and an attribute circle layout was adopted. The penalty rules used to indicate the thickness are as follows: (i) default penalty between two nodes = 1; (ii) for one more phylogenetic incongruity or specific geographic origin, the penalty is + 1; (iii) the lower the penalty, the greater the thickness.

Small RNA Analysis

The small RNA data previously used were reanalyzed to extract sRNA characteristics of the new satellite RNA that is likely associated with the nepovirus (Cao et al., 2019). Reads were mapped to the satellite sequence using the CLC Genomics Workbench 11. Size distribution and the 5′-nucleotide (5′-nt) preference of satellite sRNAs were counted and schematized using Origin 2017. The distribution of sRNAs in the positive (pos) and negative (neg) strands of the satellite RNA was visualized by the ggplot2 package in R. For sRNA size distribution of the satellite, the Kolmogorov–Smirnov (K–S) method in Origin 2017 was used to test the normality, and then the distribution was fitted with a Gaussian function. The viral/subviral RNA-clustering sRNA size and 5′-nt heatmaps were Z-score normalized and plotted using the heatmap.2 function (gplots package 3.1.0) in R (Warnes et al., 2016).

[1] https://www.ncbi.nlm.nih.gov/orffinder/

[2] https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

RESULTS

Virus Identification and Sequence Confirmation

After data processing, a total of 49,756,462 (6.95 Gb)–84,446,876 (11.8 Gb) clean reads were obtained for individual sequencing of 20 samples except that of CQ1-D, which was previously published (**Table 1**). The reads of each sample were assembled and annotated independently to detect known viruses and any potential new viruses. This allowed identification of contigs related to the known green Sichuan pepper-nepovirus (GSPNeV), -idaeovirus (GSPIdV, note that its abbreviation in previous study is GSPIV), -enamovirus (GSPEV), and -nucleorhabdovirus (GSPNuV), and the new contigs homologous to the *Nepovirus* genus-associated satellite RNA and viruses in the genus *Ilarvirus* (family *Bromoviridae*), based on the *e*-values (> 5e–04 for the former, and > 6e–98 for the latter) from BLASTx analysis. After verification by RT-PCR, cloning, and sequencing of complete nucleotide sequences of the satellite and most of the ilarvirus genome and based on preliminarily taxonomic analyses, we tentatively named the satellite and ilarvirus "green Sichuan pepper-nepovirus large satellite RNA" (satGSPNeV) and "-ilarvirus" (GSPIlV), respectively. With the use of GSPNeV, satGSPNeV, GSPIdV, and GSPEV from the CQ1-D sample and GSPIlV from the CQ7-DS sample as reference genomes for read mapping, viral/subviral consensus sequences between the references and mapped reads from other samples were obtained. Since viral/subviral nucleotide sequences derived from independent assembly of all sample data were highly identical to the consensus sequences (>99%), these consensus sequences were considered predominant and accurate and were directly used for other analyses. If high sequence heterogeneity is present in a viral/subviral quasispecies that can affect sequence assembly, contigs from assembly and consensus sequences from read mapping with different parameters (threshold values: similarity = 0.4–0.8, fraction = 0.4–0.8) will be obviously distinct from one another, and our tests showed that the viral/subviral sequences resulting from different conditions were highly consistent.

Viral/Subviral Constituents and Accumulation in Samples

Overall, reads from GSPNeV and satGSPNeV were more abundant in diseased samples, in which they accounted for 12–36% of the total reads relative to the reads from all other viruses (<11% for GSPIdV, GSPEV, GSPNuV, GSPIlV), but in those samples for which symptoms were not obvious, the read proportions for GSPNeV and satGSPNeV were < 0.003% (**Figures 1A,B** and **Table 1**). A more specific comparison of three biological repeats (**Figure 1C** and **Table 1**) between symptomatic and asymptomatic branch samples from the same trees showed a striking accumulation of GSPNeV and satGSPNeV rather than other viruses in symptomatic samples, not in asymptomatic samples (almost no reads), and this result was confirmed by RT-PCR: GSPNeV and satGSPNeV were only detected in symptomatic samples. Symptomatic and asymptomatic samples were independently clustered by read numbers of viral/subviral RNAs (**Figure 1D**), indicating a special pattern of viral/subviral infections in symptomatic samples, where GSPNeV and satGSPNeV were major factors while other viruses were minor, because although GSPIlV occurred in all diseased trees, it was also found in asymptomatic trees with a moderate read abundance (**Figures 1E,F** and **Table 1**). Correlations of GSPNeV with satGSPNeV and the FYD shown by HTS analyses were confirmed by RT-PCR and gel electrophoresis (**Supplementary Figure 1**). Collectively, these data suggested GSPNeV and satGSPNeV are closely associated with the FYD across two type of spaces, namely, different branches of a single tree and different tree geographical positions.

Sequence and Phylogenetic Analyses

The GSPIlV from the CQ7-DS sample (viral isolate IlCQ7-DS; isolate name = the first two letters of the virus genus name or the "satellite" plus the sample name) was partially sequenced, and its nearly complete genomes are tripartite, exhibiting canonical genomic organizations found in the genus *Ilarvirus*, especially the subgroup III (G3) members (Noda et al., 2017; Bratsch et al., 2019), that is, typical monocistronic RNA1, RNA2, and bicistronic RNA3 (**Figure 2A**). Extensive identical sequences were observed at the 3′ genomic ends (**Supplementary Figure 2**). RNA1 (∼ 3.4 kb) contains a large ORF1 (nt 37–3,150), potentially coding for a replicase (Rep, 1,037 aa) based on methyltransferase (Mtr, aa 51–387, pfam01660) and helicase (Hel, aa 745–1,002, pfam01443) domains detected by CDD search. RNA2 (∼ 2.3 kb) with ORF2 encodes a putative protein (765 aa) with an RNA-dependent RNA polymerase (RdRp, aa 271–707, pfam00978) domain. RNA3 harbors two ORFs, ORF3a (nt 170–994) and ORF3b (nt 998–1,696), which were predicted to encode movement protein (MP, 274 aa; domain cl03270 at aa 7–262), and coat protein (CP, 232 aa; domain cl03355 at 29–227), respectively. In BALSTx analyses based on the NCBI Viruses database (taxid:10239), GSPIlV showed greater sequence homology to ilarviruses belonging to G3 species (score rank < 40), compared with other viruses (score rank > 40). We conducted sequence comparison analysis and found evidently higher nt/aa sequence identities that GSPIlV shared with G3 species than with others (**Supplementary Figure 3**). In addition, GSPIlV was phylogenetically closest to the G3 ilarviruses, independent of the proteins used for the analyses (**Figures 2B–E**). We propose that GSPIlV could be a member of a new species in G3 of the genus *Ilarvirus*.

The full-length sequence of the satGSPNeV isolate from sample CQ1-D (subviral isolate SaCQ1-D) was obtained, which comprises 2,247 nts excluding the poly(A) tail. Its genomic 5′ and 3′ termini are to some extent conserved compared with GSPNeV RNAs (**Supplementary Figure 4**). Two ORFs, one (nt 713–928) encompassed in the other (nt 13–1,110), were predicted in the sequence (**Figure 2F**), but the putative protein products (365-aa and 71-aa, respectively) from them were functionally unknown. Interestingly, compared with other related satellite RNAs associated with other viruses, satGSPNeV has an extraordinary long 3′ untranslated region (1,141-nt), similar to two genomic RNAs of GSPNeV. The nt sequence

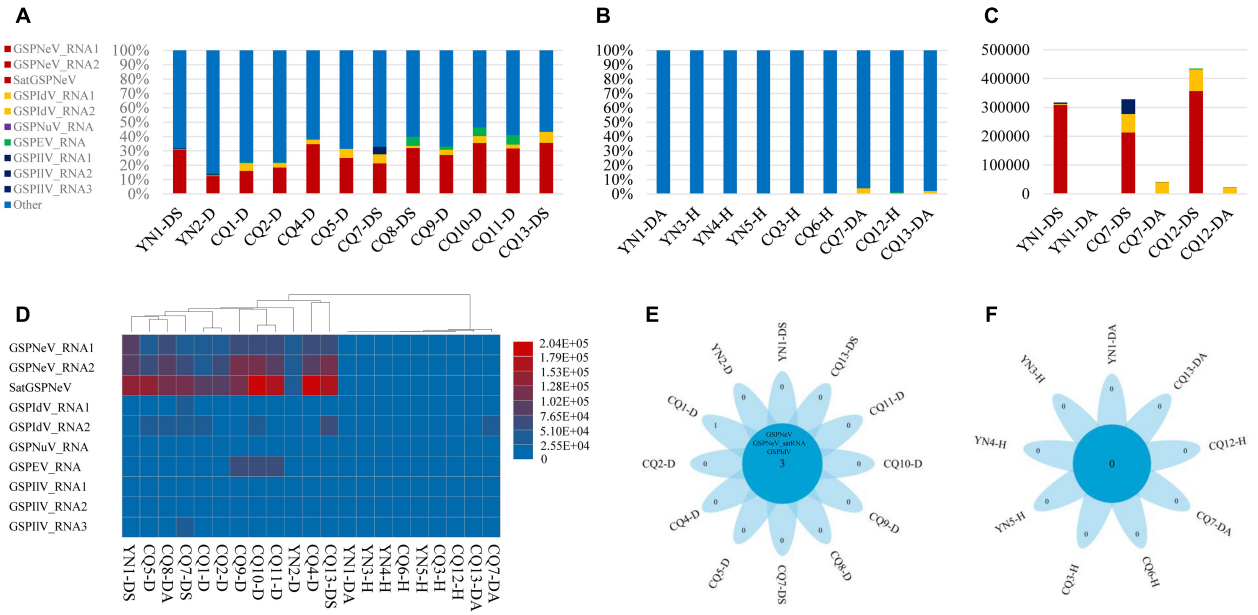


FIGURE 1 | Virome analysis of *Z. armatum* trees affected with the flower yellowing disease and the trees without obvious symptoms. Proportion of reads associated with the viral and subviral RNAs, relative to total reads of each symptomatic sample **(A)**, and asymptomatic/healthy sample **(B)**, and reads of evolution-associated RNAs, i.e., two RNAs of green Sichuan pepper-nepovirus (GSPNeV) with green Sichuan pepper-nepovirus large linear satellite RNA (satGSPNeV), two RNAs of green Sichuan pepper-idaeovirus (GSPIdV) or three RNAs of green Sichuan pepper-ilarvirus (GSPIIV), were counted together. Viral/subviral read number comparison between symptomatic and asymptomatic branch samples from the same tree, with three independent partial diseased trees analyzed **(C)**. In figures **(A–C)**, the colors of bars indicate different viruses, and satGSPNeV is shown the same color with GSPNeV. Heatmap of read numbers of viral and subviral RNAs from different samples, with the clustering of samples **(D)**; the deeper the colors of bars (from blue to red), the higher the numbers of reads specific to each viral/subviral RNA. Flower plots indicate the numbers of common virus/satellite across the symptomatic **(E)** or asymptomatic/healthy **(F)** samples that are shown in the central blue circles or of specific virus/satellite in each sample that are shown in the petals; two viruses and one satellite were detected in all symptomatic samples that have been marked with D or DS, but no common virus/satellite was detected in asymptomatic/healthy samples that have been marked with DA or H. Green Sichuan pepper-enamovirus (GSPEV), and -nucleorhabdovirus (GSPNuV).

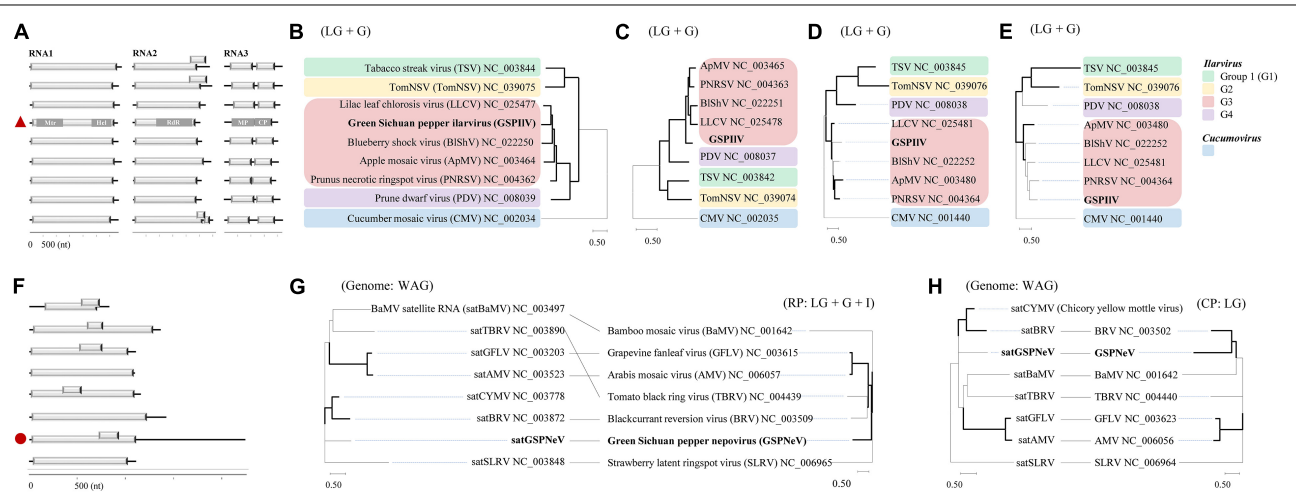


FIGURE 2 | Sequence and phylogenetic characteristics of green Sichuan pepper-ilarvirus (GSPIIV) and green Sichuan pepper-nepovirus large linear satellite RNA (satGSPNeV). Genomic comparisons of GSPIIV (the red triangle) and ilarviruses **(A)**. Mtr, methyltransferase; Hel, helicase; MP, movement protein; CP, coat protein. Phylogenetic relationships of GSPIIV and ilarviruses inferred from alignments of replicase (RP, **B**), RNA-dependent RNA polymerase **(C)**, MP **(D)**, and CP **(E)** sequences; cucumber mosaic virus (CMV) was used as an outgroup. RNA structure of satGSPNeV (**F**, the red circle). Phylogenetic trees and co-evolution analyses of satellite RNAs (**G,H**, left) and their helper viruses (**G,H**, right) based on satellite nucleotide sequences and helper virus RPs (**G**, right) and (**H**, right). The models in MEGA-X selected for phylogenies are indicated. The nodes with > 50% bootstrap supports are shown by thick lines.

identity shared with other viral satellite RNAs was not significant (< 29%), in addition to the aa sequence identity for the larger protein (< 23%). Despite this, BLASTp analysis of this protein suggested that it is homologous to those of blackcurrant reversion virus satellite RNA (*e* = 1.85E−10) and chicory yellow mottle virus large satellite RNA (*e* = 8.71E−23). The phylogenetic analysis with the complete nucleotide sequences also placed satGSPNeV with these two satellite RNAs (**Figures 2G,H**, on left). When related nepoviral RNA1 and RNA2 polyproteins were additionally included for co-evolution analysis, more phylogenetic incongruities were observed between Rep-encoding RNA1-polyprotein and the satellites (**Figures 2G,H**, on right); therefore, it appears that the coat protein-encoding RNA2s of their helper viruses are more evolutionarily associated with them. In fact, the other satellite RNAs analyzed may have a CP origin due to their close phylogenetic relationships with some viral CP (Alazem and Lin, 2017). However, this hypothesis does not have any molecular support. Based on this evidence, satGSPNeV should be a new *Nepovirus*-associated large satellite RNA.

Phylo-Genetic-Geographic Analysis

Based on the viral/subviral consensus sequences (or preponderate sequences) of quasispecies obtained from the read mapping of available samples, phylogenetic dendrograms of the viral/subviral RNAs from different isolates were constructed to gain insight into their co-evolution (single viral-species/satellite) and co-transmission (two more viral-species/satellite) relationships by weighting both phylogenetic and geographic discrepancies between the trees (**Figure 3**). The populations from five independent locations were analyzed (**Figure 3A**). For GSPNeV RNAs and satGSPNeV, the consensus sequences from different locations were, in a similar phylogenetic architecture, clearly separately clustered by the locations (**Figures 3B,C**), thus suggesting their strong associations and simple population structures without noteworthy long-distance movements and cross infections. A lower degree of association was observed between GSPIdV RNAs or among GSPIdV, GSPNeV RNAs, and satGSPNeV (**Figures 3B–F**), probably because of genomic reassortments or mixed infections of GSPIdV. In contrast, the evolutionary statuses of GSPEV were complicated in contrast to these viruses and the satellite (**Figure 3G**), where there are likely to be independent genetic and geographic development trajectories since the phylogenetic clustering and range of infected samples were largely distinct. GSPIlV also displayed population separation by geographic positions in spite of a small number of infected samples available (**Figures 3H–J**). However, the much smaller number of sequenced samples (only one) that GSPNuV infected made analysis impossible.

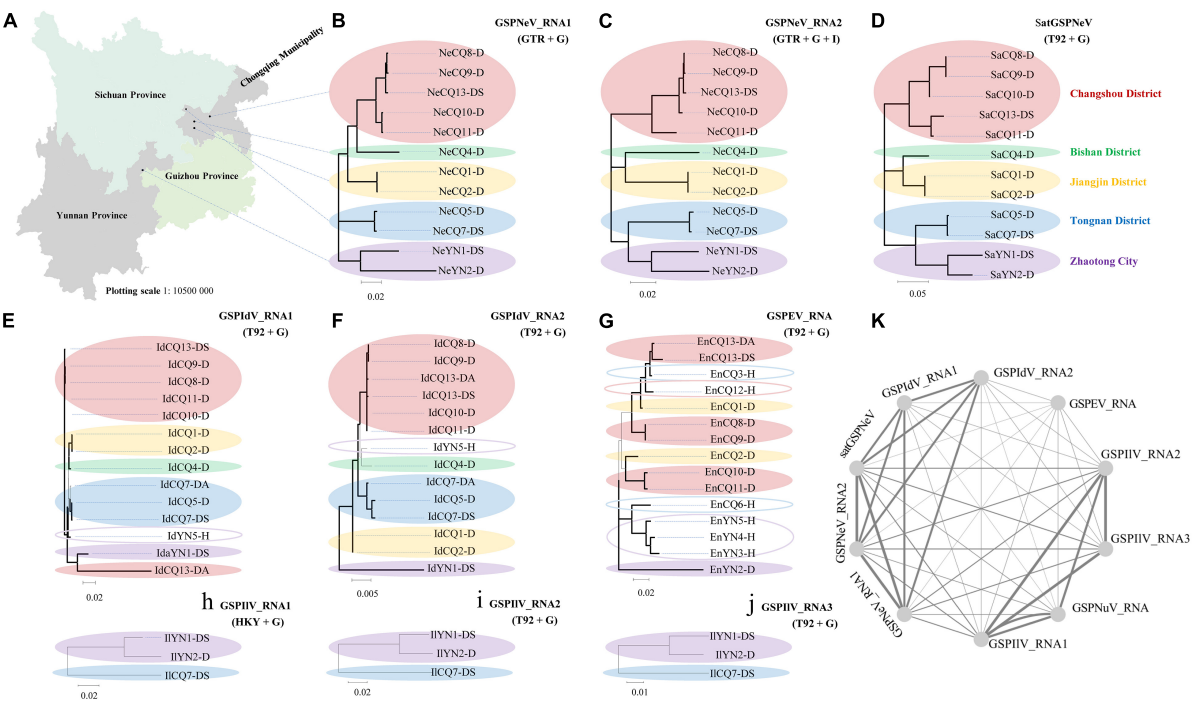


FIGURE 3 | Phylogenetic and geographical analysis of available viral/subviral isolates. A map showing the sampling sites, which are shown with black circles **(A)**. Co-evolution and co-transmission relationship constructions of green Sichuan pepper-nepovirus—GSPNeV **(B,C)**, -nepovirus large linear satellite RNA—satGSPNeV **(D)**, -idaeovirus—GSPIdV **(E–G)**, -enamovirus—GSPEV **(H)**, -nucleorhabdovirus—GSPNuV (analysis not available for only one isolate), and -ilarvirus—GSPIlV **(H–J)**, based on the phylogenetic and geographical differences. The tree structures and geographical distributions of GSPNeV RNAs and satGSPNeV were used as references for comparisons and penalty point calculations. Isolates with the same locations are indicated by solid ellipses with the same color, and hollow ellipses represent isolates with unique distributions, compared with GSPNeV RNAs and satGSPNeV. Network visualization of the relationships by adding up penalties between two RNAs **(K)**; each node represents a different RNA. The models recommended by MEGA-X for phylogenetic analyses are indicated. The nodes with significant bootstrap values (>50%) are highlighted with thick lines.

The results resembling those of the phylogenetic analyses were obtained from the sequence identity matrixes, where the 0–14.2% nt differences were shown between the consensus sequences of different viral/subviral RNAs (**Supplementary Figure 5**). Finally, by summing the penalties from phylogenetic and geographic incompatibility, a network was built that formed two main clustering groups, one for GSPNeV, GSPIdV RNAs, and satGSPNeV, and another for GSPIlV RNAs and GSPNuV, and a distant node for GSPEV, signifying the presence of both historically evolutionary similarities and differences among the viruses and satellite (**Figure 3K**).

Sequence Variation Site Pattern and Recombination Analyses

Viral/subviral isolates from CQ1-D and CQ7-DS samples were used as reference genomes, based on which nucleotide sequence variation sites from comparisons with other available conceptional isolates (the consensus sequences) were counted and added together (**Figure 4**, the histograms of each graphical illustration), regardless of definite variation type (AUCG) or length (insertions and deletions). Within-quasispecies variations were also analyzed, using the viral/subviral isolates from each sample as a reference (**Figure 4**, below the consensus), as well as the Spearman correlation coefficient (r) between the consensus sequences and quasispecies variations. Generally, the range of viral/subviral variation sites from the consensus sequences was larger than that of the quasispecies, and the only exception was GSPEV, which suggests that it has a relatively faster variation speed under similar environmental conditions. Regardless, the r for variation site comparisons of the consensus sequences and quasispecies was low (< 0.5), which, besides a smaller size of the genetic pool in quasispecies, is also potentially due to different natural selections specific to each quasispecies before a structurally stable population. Even so, Spearman correlation analysis regarding trends in the numbers of variation sites from the consensus sequences, quasispecies, and their common variation sites (**Figure 4**, the Venn diagrams) showed that these items of different viral/subviral RNAs had strong correlations, with $r > 0.77$ (**Supplementary Figure 6**). With increasing numbers, the genome-wide distribution of variation sites from both the consensuses and quasispecies became more even, but sequence identity analyses suggested the presence of some evolutionary hotspots in the viral/subviral RNAs (**Figure 4**). Thus, there was great variation complexity in some regionally specific sites. Variation site heatmaps of the RNAs from different quasispecies showed that the different isolates were sometimes grouped by geographic positions, but most were irregularly clustered (**Figure 4**). Viral sequence variation is generally affected by multifactors other than single factors (García-Arenal et al., 2001; Domingo and Perales, 2019), such as environments and hosts, let alone different stages of viral populations; therefore, these clusters are not

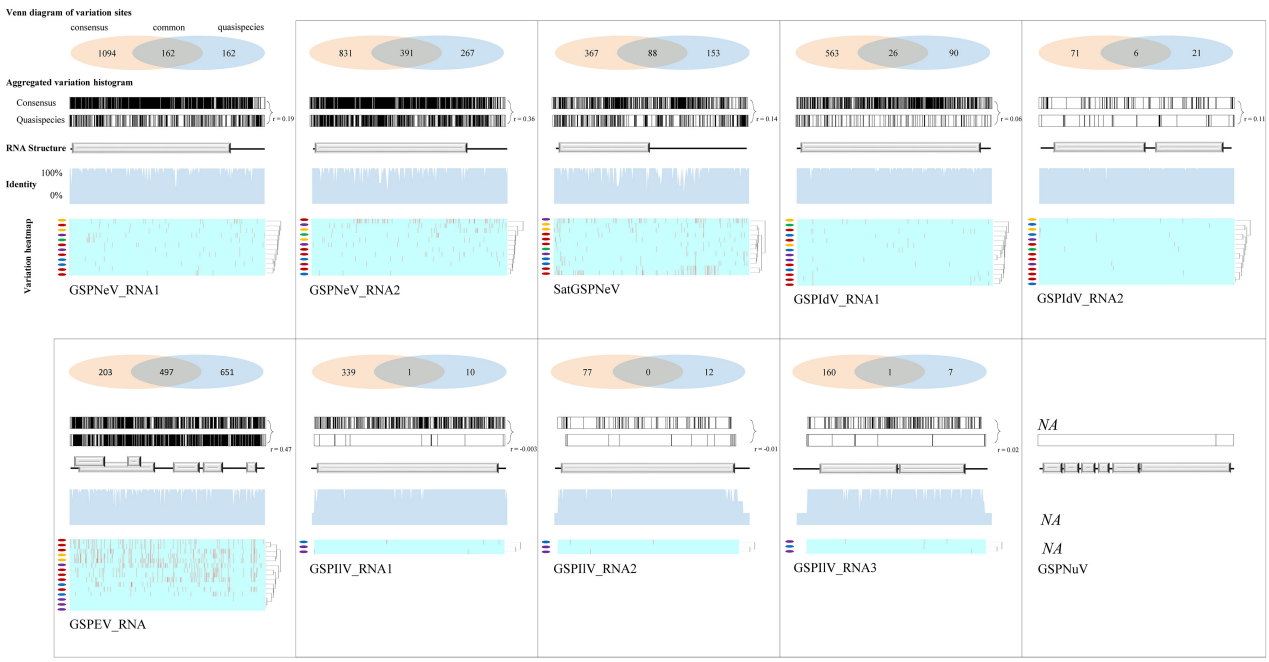


FIGURE 4 | Variation sites pattern analyses of viral/subviral RNAs. Green Sichuan pepper-nepovirus (GSPNeV), -idaeovirus (GSPIdV), -enamovirus (GSPEV), -nucleorhabdovirus (GSPNuV), -nepovirus large linear satellite RNA (satGSPNeV), and -ilarvirus (GSPIlV). For each RNA, as each diagram shows, variation sites among the consensus sequences from the read mapping of different samples to reference isolates from CQ1-D and CQ7-DS or aggregated from comparisons of each consensus sequence with their own quasispecies were calculated (the Venn diagram) and visualized (the variation histogram). The common variation sites between the consensus sequences and quasispecies were also determined, and the counted numbers shown for the consensuses and quasispecies are unique. The plots for sequence identity among the consensus sequences are shown (the identity histogram). The variation heatmap at the bottom shows the clustering of the samples by variation site distributions of their quasispecies, and the quasispecies from the same geographical locations are indicated with the same colored circles.

abnormal. Recombination analyses were also performed for the available consensus sequences of the viruses (**Table 2**), and the results showed that GSPNeV/satGSPNeV and GSPEV with more variation sites in the quasispecies had more recombination events based on the program supports (> 2) and threshold *e*-values (< E–5).

Small RNA Profiles of the Satellite

Small RNAs of satGSPNeV spread almost evenly in the genome (SaCQ1-D) except for several hotspots (**Figure 5A**). Among satGSPNeV sRNAs, the most abundant were accumulated in the size range of 21 and 22 nt (**Figure 5B**), and in the 5′-nt of U and C (**Figure 5C**), similar to the characteristics for other viruses previously reported (Cao et al., 2019). The K–S tests showed that the size distributions, in the positive, negative or both strands, were in accordance with the normal (Gaussian) distribution when the predetermined significance level was 0.05, despite the shapes being different (**Figure 5B**). Clustering of available viruses and satellite in the CQ1-D, irrespective of the size (**Figure 5D**) or 5′-nt (**Figure 5E**), showed that satGSPNeV was closely associated with the GSPNeV RNA1 that encodes RNA-directed RNA polymerase (RdRp). Because intracellular replication of satellites relies intrinsically on proteins from their helper viruses (Palukaitis, 2017), it is not surprising that satellites, accompanied by helper viruses, may enter the same processes by host endogenous RNA silencing pathways onto the helper viruses, thereby resulting in similar small RNA patterns. Under similar selection pressures exerted by *Z. armatum* RNA silencing, accumulation of GSPNeV and satGSPNeV in the sample CQ1-D was still abundant, which may represent complex defense/counter-defense interactions between the hosts and viruses/satellite (Liu et al., 2017).

DISCUSSION

Over the last ten years, the FYD has become a restrictive factor in *Z. armatum* production in China after an endemic outbreak turned invasive and prevalent in the southwestern regions, including Chongqing Municipality, Yunnan, and Sichuan provinces, the main producing areas. A previous study revealed four different RNA viruses in one diseased tree (named CQ1-D in this study), and that GSPNeV was mainly associated with the FYD through field investigations (Cao et al., 2019). Here, to understand viral diversity of the FYD-affected trees by taking different sampling tissues (symptomatic vs. asymptomatic) and sites (five) into account, we adopted a more accurate HTS technique than conventional RT-PCR, whose detection range is limited for previous identified viral sequences. We conducted a large-scale survey, as 21 samples in total were analyzed. After a series of systematic data processes and molecular experiments, we newly identified a large, linear, single-stranded satellite RNA associated with GSPNeV, the satGSPNeV, and an ilarvirus, i.e., GSPIlV. They were taxonomically new based on comprehensive analyses and should be classified as a new species or satellite RNA under their own taxa, viral genus *Ilarvirus* and the subviral category, respectively (Briddon et al., 2012; Bujarski et al., 2019). SatGSPNeV was omitted in the previous analyses of CQ1-D (Cao et al., 2019), likely because of our annotation mistakes (contigs from the assembly of highly heterogeneous reads will sometimes be too short to be recognizable). As the current evidence showed, both GSPNeV and satGSPNeV were FYD-associated, while GSPIlV and the other viruses were not connected with the FYD and their symptoms on *Z. armatum* were still unclear due to the limited number of infected samples investigated. The relationships between these non-FYD-related viruses and the hosts, i.e., competition or mutualism, remain unknown.

In agroecosystems, plant viruses sometimes are destructive to crop industries because of the presence of sensitive cultivars (Moreno et al., 2008; Rybicki, 2015; Lefeuvre et al., 2019). In fact, observation of GSPNeV and satGSPNeV with high read proportions of the total reads sequenced from RNAs of symptomatic samples suggested their intense replication in the host. It is not difficult to imagine that they will compete with hosts for nutrient resources and may interfere with host normal physiological activity and cause visible symptoms (Ammara et al., 2017). Higher viral titers may also help the biological vectors perform more efficient viral acquisition and transmission (Gray et al., 1991), but not necessarily (Linak et al., 2020). Statically significant analyses disclosed the strong ability of GSPNeV and satGSPNeV to coinfect *Z. armatum*.

Genomic variation site analyses of the viruses and satellite allowed the assessment of their adaptability to different environments. Long- and short-term variation sites predicted from the comparisons of different consensus sequences and quasispecies, respectively, were largely discrepant in spite of a similar tendency in quantitative changes, suggesting the presence of circumstance-specific variations in the quasispecies. More strong evidence was obtained directly from observations of variable site patterns of the quasispecies from different locations. Nevertheless, we have to note that there are selective processes during viral evolution, so short-term variation sites are time-sensitive, changeable at different stages, and may not be ultimately preserved (García-Arenal et al., 2003). However, this can explain the aforementioned site discrepancies between the consensus sequences and quasispecies. When we calculated the nt-average variation sites, GSPNeV, satGSPNeV, and GSPEV had superior genetic variability in both quantity and evenness

TABLE 2 | Recombination events prediction for the viral/subviral RNAs.

Viral/subviral sequence	Significant event	Lowest method support (level < E–5)
GSPNeV_RNA1	9	2
GSPNeV_RNA2	4	2
SatGSPNeV	5	3
GSPIdV_RNA1	0	NA
GSPIdV_RNA2	0	NA
GSPNuV_RNA	NA	NA
GSPEV_RNA	9	2
GSPIlV_RNA1	0	NA
GSPIlV_RNA2	0	NA
GSPIlV_RNA3	0	NA

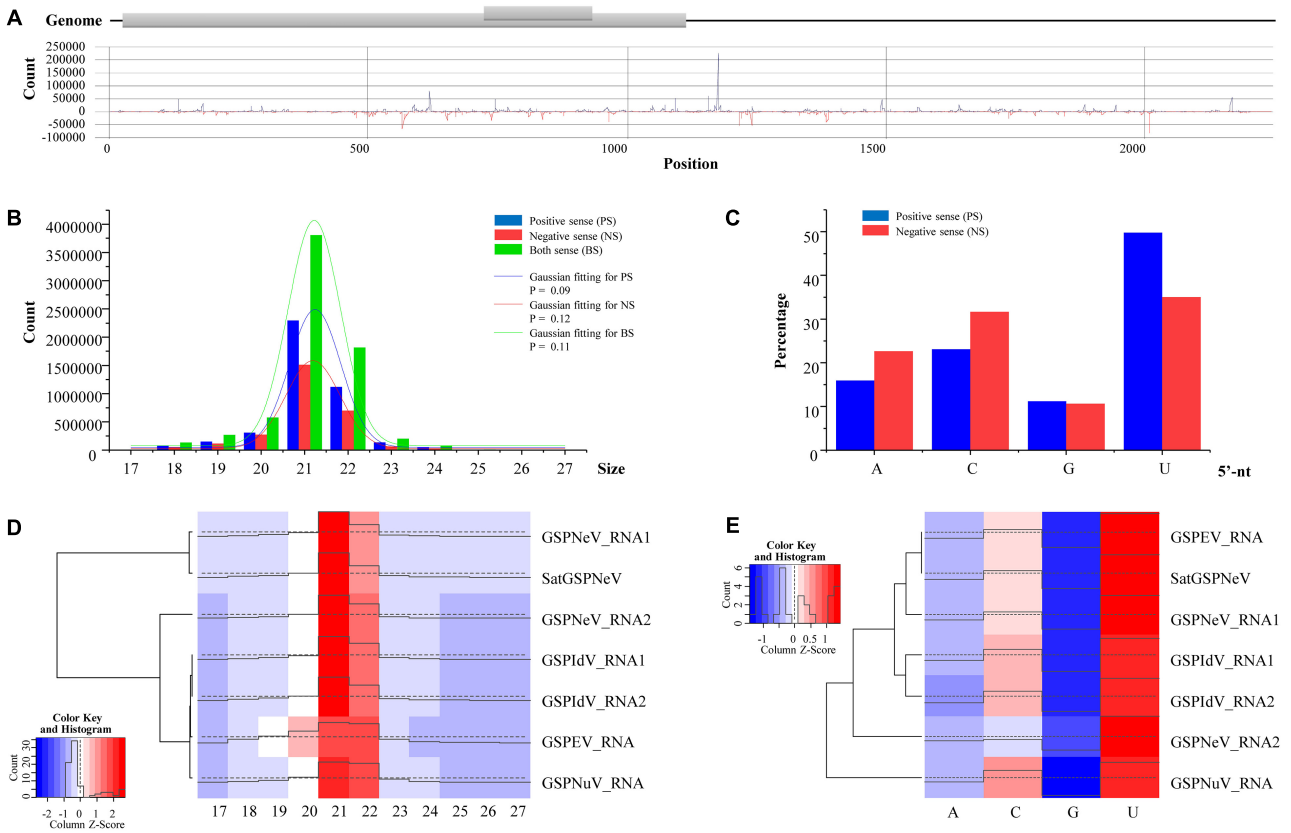


FIGURE 5 | Small RNA (sRNA) analysis of green Sichuan pepper-nepovirus large linear satellite RNA (satGSPNeV). Distribution of sRNAs derived from positive and negative strands of satGSPNeV in its sequences **(A)**. The size distribution of satGSPNeV-derived sRNAs in positive, negative, or both strands is in line with the normal distribution **(B)**. The 5′-nucleotide bias of satGSPNeV sRNAs **(C)**. Small RNA size distribution **(D)** and 5′-nucleotide bias **(E)** patterns of satGSPNeV are similar to those of other coinfecting RNA viruses; the color and histogram annotations are associated with abundance of specific sRNAs, with higher the histogram and deeper the red (or lighter the blue) representing higher the numbers of sRNAs.

(more genomic regions) relative to other viruses, which may contribute to their competitive survival in hosts. More viral variation sites may increase the possibility of viral strains with better transmissibility and pathogenicity (severe vs. mild strains) (Niblett et al., 2000), despite a trade-off between this and the increasing deleterious mutations (Carrasco et al., 2007). As for genomic recombination, sequence variation brings about heterogeneous sequences that may recombine during replications in single cells, and this therefore makes the recombinations between divergent sequences predictable, even though recombinations will negatively regulate the divergences (Lai, 1992). In our data, higher sequence variations were associated with more genomic recombinations, probably because variable populations require recombinations for purging their deleterious mutations (Xiao et al., 2016). However, in this sense, the lower sequence variation and negative recombination prediction results of the GSPIdV RNAs do not mean the absence of recombinations; it is possible that these could not be predicted using a computer.

Cutting off of the transmission routes of viruses, if they do exist, is a simple but practical method for management and control before viruses lead to greater losses (Jones, 2006). Apart from genetic similarity at the both genomic ends, perfect consistency among GSPNeV RNAs and satGSPNeV in both phylogenetic and geographical relationships of the different isolates suggested their tight binding in both transmission and evolution, similar to other nepoviruses and their satellite RNAs (Alazem and Lin, 2017). Such a distinct distribution of different representative sequences, namely, homogeneous in each population and a lack of viral cross infections, may also suggest their weak natural movement ability and the involvement of human activities in transmission, such as seedling or plant material transportation. A similar situation was observed for GSPIdV RNAs, which, in previous work, were rarely detected in seedlings and fields (Cao et al., 2019). One possibility for the cooccurrence of GSPIdV with GSPNeV and satGSPNeV is that GSPIdV may benefit from this coinfection for easier invasion in the hosts, as synergism exists widely in plant viruses (Pruss et al., 1997; Mascia and Gallitelli, 2016). In contrast, it appears that the transmission of GSPNEV is more efficient than that of other viruses and satellites and naturally random, because more areas and heterogeneous sequences were involved and interrelated.

In summary, based on HTS-dependent viromic analyses, we discovered a new satellite RNA associated with GSPNeV and

the FYD, as well as a new ilarvirus, and we studied their regional viral/subviral diversification, as well as that of other RNA viruses. This information can be useful for sustainable development of *Z. armatum* and helpful for understanding viral/subviral evolution. Then, our future works will have a shift of emphasis toward assessing biological properties of GSPNeV and satGSPNeV by construction of their infectious cDNA clones in planta. Meanwhile, parallel work is urgently needed, regarding elucidating the viral/subviral transmissions in fields, to restrain the dispersals.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI GenBank repository, accession numbers MH323432–MH323437 and MW962309–MW962391, and in the NCBI SRA repository, BioProject PRJNA732832, BioSample SAMN19341721–SAMN19341741, SRA accession numbers SRR14663372–SRR14663392.

AUTHOR CONTRIBUTIONS

MC conceived and designed the experiments. SZ, RL, XW, ZX, BZ, ZL, JZ, XD, ZT, SL, and YZ collected the samples and conducted the experiments. MC and SZ analyzed the data. MC, SZ, SL, and YZ discussed the results, drafted, and revised the manuscript. All authors read and approved the final draft of the manuscript.

FUNDING

This research was supported by the Central Public-Interest Scientific Institution Basal Research Fund (S2021XM11), Innovation Program for Chongqing's Overseas Returnees (cx2019013), and 111 Project (B18044).

ACKNOWLEDGMENTS

We would like to thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.702210/full#supplementary-material

Supplementary Figure 1 | Detection of GSPNeV RNAs and satGSPNeV in an orchard affected by the flower yellowing disease. 1, 2, 4, 6, 8, 9, 11, symptomatic samples; 3, 5, 7, 10, 12, samples without symptoms; +, virus-positive control; H, healthy control; 0, water control.

Supplementary Figure 2 | Comparisons of partial genomic terminal sequences from RNAs of green Sichuan pepper-ilarvirus (GSPIlV) suggested sequence conservation and the presence of complete coding sequences in the obtained sequences.

Supplementary Figure 3 | Sequence comparisons of genomic RNAs of Sichuan pepper-ilarvirus (GSPIlV) with those of selected related viruses. Cucumber mosaic virus from the genus *Cucumovirus* was used as an outgroup. Three groups (G1, G2, G2) as subgroups under the genus *Ilarvirus* are indicated.

Supplementary Figure 4 | Comparisons among GSPNeV RNAs and satGSPNeV showed some conserved sequences of the genomic ends.

Supplementary Figure 5 | Matrixes of sequence identities between RNAs of different viral/subviral isolates obtained in this study.

Supplementary Figure 6 | Correlations of numbers of variation sites from the consensus sequences and quasispecies of different viral/subviral RNAs and the common variation sites between the consensus sequences and quasispecies.

Supplementary Table 1 | List of primers used in this study.

REFERENCES

Alazem, M., and Lin, N.-S. (2017). "Large satellite RNAs," in *Viroids and Satellites*, eds A. Hadidi, R. Flores, J. W. Randles, and P. Palukaitis (London: Elsevier Academic Press), 639–648. doi: 10.1016/b978-0-12-801498-1.00059-0

Ammara, U., Al-Sadi, A. M., Al-Shihi, A., and Amin, I. (2017). Real-time qPCR assay for the TYLCV titer in relation to symptoms-based disease severity scales. *Int. J. Agric. Biol.* 19, 145–151. doi: 10.17957/ijab/15.0256

Bernardo, P., Charles-Dominique, T., Barakat, M., Ortet, P., Fernandez, E., Filloux, D., et al. (2018). Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME J.* 12, 173–184. doi: 10.1038/ismej.2017.155

Bratsch, S. A., Grinstead, S., Creswell, T. C., Ruhl, G. E., and Mollov, D. (2019). Characterization of tomato necrotic spot virus, a subgroup 1 ilarvirus causing necrotic foliar, stem, and fruit symptoms in tomatoes in the United States. *Plant Dis.* 103, 1391–1396. doi: 10.1094/pdis-11-18-2112-re

Briddon, R., Ghabrial, S., Lin, N., Palukaitis, P., Scholthof, K., and Vetten, H. (2012). "Satellites and other virus-dependent nucleic acids," in *Virus Taxonomy-Ninth Report of the International Committee on Taxonomy of Viruses*, eds A. M. Q. King, M. J. Adams, E. B. Carstens, and E. J. Lefkowitz (London: Elsevier Academic Press), 1209–1219.

Bujarski, J., Gallitelli, D., García-Arenal, F., Pallás, V., Palukaitis, P., Reddy, M. K., et al. (2019). ICTV virus taxonomy profile: Bromoviridaee. *J Gen Virol* 100, 1206–1207.

Cao, M., Zhang, S., Li, M., Liu, Y., Dong, P., Li, S., et al. (2019). Discovery of four novel viruses associated with flower yellowing disease of green Sichuan pepper (*Zanthoxylum armatum*) by virome analysis. *Viruses* 11:696. doi: 10.3390/v11080696

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Carrasco, P., de la Iglesia, F., and Elena, S. F. (2007). Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch* virus. *J. Virol.* 81, 12979–12984. doi: 10.1128/jvi.00524-07

Coetzee, B., Freeborough, M.-J., Maree, H. J., Celton, J.-M., Rees, D. J. G., and Burger, J. T. (2010). Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology* 400, 157–163. doi: 10.1016/j.virol.2010.01.023

Dolja, V. V., Krupovic, M., and Koonin, E. V. (2020). Deep roots and splendid boughs of the global plant virome. *Annu. Rev. Phytopathol.* 58, 23–53. doi: 10.1146/annurev-phyto-030320-041346

Domingo, E., and Perales, C. (2019). Viral quasispecies. *PLoS Genet.* 15:e1008271. doi: 10.1371/journal.pgen.1008271

Ekka, G., Jadhav, S. K., and Quraishi, A. (2020). "An overview of genus *Zanthoxylum* with special reference to its herbal significance and application," in *Herbs and Spices*, eds M. Akram and R. S. Ahmad (Rijeka: IntechOpen Limited), 1–17.

García-Arenal, F., Fraile, A., and Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* 39, 157–186.

García-Arenal, F., Fraile, A., and Malpica, J. M. (2003). Variation and evolution of plant virus populations. *Int. Microbiol.* 6, 225–232. doi: 10.1007/s10123-003-0142-z

Gilbertson, R. L., Batuman, O., Webster, C. G., and Adkins, S. (2015). Role of the insect supervectors *Bemisia tabaci* and *Frankliniella occidentalis* in the emergence and global spread of plant viruses. *Annu. Rev. Virol.* 2, 67–93.

Gray, S. M., Power, A. G., Smith, D. M., Seaman, A. J., and Altman, N. S. (1991). Aphid transmission of barley yellow dwarf virus: acquisition access periods and virus concentration requirements. *Phytopathology* 81, 539–545. doi: 10.1094/phyto-81-539

Jones, R. A. (2006). Control of plant virus diseases. *Adv. Virus Res.* 67, 205–244. doi: 10.1016/s0065-3527(06)67006-1

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kolde, R., and Kolde, M. R. (2015). Package 'pheatmap'. *R Package* 1:790.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lai, M. (1992). RNA recombination in animal and plant viruses. *Microbiol. Mol. Biol. Rev.* 56, 61–79. doi: 10.1128/mmbr.56.1.61-79.1992

Lefeuvre, P., Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P., and Varsani, A. (2019). Evolution and ecology of plant viruses. *Nat. Rev. Microbiol.* 17, 632–644.

Linak, J. A., Jacobson, A. L., Sit, T. L., and Kennedy, G. G. (2020). Relationships of virus titers and transmission rates among sympatric and allopatric virus isolates and thrips vectors support local adaptation. *Sci. Rep.* 10:7649.

Liu, S.-R., Zhou, J.-J., Hu, C.-G., Wei, C.-L., and Zhang, J.-Z. (2017). MicroRNA-mediated gene silencing in plant defense and viral counter-defense. *Front. Microbiol.* 8:1801. doi: 10.3389/fmicb.2017.01801

Ma, Y., Li, J., Tian, M., Liu, Y., and Wei, A. (2020). Authentication of Chinese prickly ash by ITS2 sequencing and the influence of environmental factors on pericarp quality traits. *Ind. Crop Prod.* 155:112770. doi: 10.1016/j.indcrop.2020.112770

Maclot, F., Candresse, T., Filloux, D., Malmstrom, C. M., Roumagnac, P., van der Vlugt, R., et al. (2020). Illuminating an ecological blackbox: using high throughput sequencing to characterize the plant virome across scales. *Front. Microbiol.* 11:578064. doi: 10.3389/fmicb.2020.578064

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003.

Mascia, T., and Gallitelli, D. (2016). Synergies and antagonisms in virus interactions. *Plant Sci.* 252, 176–192. doi: 10.1016/j.plantsci.2016.07.015

Massart, S., Olmos, A., Jijakli, H., and Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Res.* 188, 90–96. doi: 10.1016/j.virusres.2014.03.029

Moreno, P., Ambrós, S., Albiach-Martí, M. R., Guerri, J., and Pena, L. (2008). *Citrus tristeza* virus: a pathogen that changed the course of the citrus industry. *Mol. Plant Pathol.* 9, 251–268. doi: 10.1111/j.1364-3703.2007.00455.x

Muhire, B. M., Varsani, A., and Martin, D. P. (2014). SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PloS One* 9:e108277. doi: 10.1371/journal.pone.0108277

Niblett, C., Genc, H., Cevik, B., Halbert, S., Brown, L., Nolasco, G., et al. (2000). Progress on strain differentiation of *Citrus tristeza* virus and its application to the epidemiology of citrus tristeza disease. *Virus Res.* 71, 97–106. doi: 10.1016/s0168-1702(00)00191-x

Noda, H., Yamagishi, N., Yaegashi, H., Xing, F., Xie, J., Li, S., et al. (2017). Apple necrotic mosaic virus, a novel ilarvirus from mosaic-diseased apple trees in Japan and China. *J. Gen. Plant Pathol.* 83, 83–90. doi: 10.1007/s10327-017-0695-x

Palukaitis, P. (2017). "Satellite viruses and satellite nucleic acids," in *Viroids and Satellites*, eds A. Hadidi, R. Flores, J. W. Randles, and P. Palukaitis (London: Elsevier Academic Press), 545–552. doi: 10.1016/b978-0-12-801498-1.00050-4

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

Rybicki, E. P. (2015). A Top Ten list for economically important plant viruses. *Arch. Virol.* 160, 17–20. doi: 10.1007/s00705-014-2295-9

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sugawara, M., Epstein, B., Badgley, B. D., Unno, T., Xu, L., Reese, J., et al. (2013). Comparative genomics of the core and accessory genomes of 48 Sinorhizobium strains comprising five genospecies. *Genome Biol.* 14, 1–20. doi: 10.1089/omi.1.1999.4.1

Warnes, M. G. R., Bolker, B., Bonebakker, L., Gentleman, R., and Huber, W. (2016). *Package 'gplots'. Various R Programming Tools For Plotting Data*.

Wickham, H. (2011). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185. doi: 10.1002/wics.147

Xiao, Y., Rouzine, I. M., Bianco, S., Acevedo, A., Goldstein, E. F., Farkov, M., et al. (2016). RNA recombination enhances adaptability and is required for virus spread and virulence. *Cell Host Microbe* 19, 493–503. doi: 10.1016/j.chom.2016.03.009

Xu, D., Zhuo, Z., Wang, R., Ye, M., and Pu, B. (2019). Modeling the distribution of *Zanthoxylum armatum* in China with MaxEnt modeling. *Glob. Ecol. Conserv.* 19:e00691. doi: 10.1016/j.gecco.2019.e00691

Xu, Q., Chen, L.-L., Ruan, X., Chen, D., Zhu, A., Chen, C., et al. (2013). The draft genome of sweet orange (*Citrus sinensis*). *Nat. Genet.* 45, 59–66.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11, 367–372. doi: 10.1016/0169-5347(96)10041-0

Zhang, M., Wang, J., Zhu, L., Li, T., Jiang, W., Zhou, J., et al. (2017). *Zanthoxylum bungeanum* Maxim. (Rutaceae): a systematic review of its traditional uses, botany, phytochemistry, pharmacology, pharmacokinetics, and toxicology. *Int. J. Mol. Sci.* 18:2172. doi: 10.3390/ijms18102172

Zhang, X., Tang, N., Liu, X., Ye, J., Zhang, J., Chen, Z., et al. (2020). Comparative transcriptome analysis identified differentially expressed genes between male and female flowers of *Zanthoxylum armatum* var. *Novemfolius*. *Agronomy* 10:283. doi: 10.3390/agronomy10020283

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 July 2021
doi: 10.3389/fmicb.2021.699078

Full Genome Evolutionary Studies of Wheat Streak Mosaic-Associated Viruses Using High-Throughput Sequencing

*Carla Dizon Redila, Savannah Phipps† and Shahideh Nouri**

Department of Plant Pathology, College of Agriculture, Kansas State University, Manhattan, KS, United States

Wheat streak mosaic (WSM), a viral disease affecting cereals and grasses, causes substantial losses in crop yields. Wheat streak mosaic virus (WSMV) is the main causal agent of the complex, but mixed infections with Triticum mosaic virus (TriMV) and High plains wheat mosaic emaravirus (HPWMoV) were reported as well. Although resistant varieties are effective for the disease control, a WSMV resistance-breaking isolate and several potential resistance-breaking isolates have been reported, suggesting that viral populations are genetically diverse. Previous phylogenetic studies of WSMV were conducted by focusing only on the virus coat protein (CP) sequence, while there is no such study for either TriMV or HPWMoV. Here, we studied the genetic variation and evolutionary mechanisms of natural populations of WSM-associated viruses mainly in Kansas fields and fields in some other parts of the Great Plains using high-throughput RNA sequencing. In total, 28 historic and field samples were used for total RNA sequencing to obtain full genome sequences of WSM-associated viruses. Field survey results showed WSMV as the predominant virus followed by mixed infections of WSMV + TriMV. Phylogenetic analyses of the full genome sequences demonstrated that WSMV Kansas isolates are widely distributed in sub-clades. In contrast, phylogenetic analyses for TriMV isolates showed no significant diversity. Recombination was identified as the major evolutionary force of WSMV and TriMV variation in KS fields, and positive selection was detected in some encoding genomic regions in the genome of both viruses. Furthermore, the full genome sequence of a second Kansas HPWMoV isolate was reported. Here, we also identified previously unknown WSMV isolates in the Great Plains sharing clades and high nucleotide sequence similarities with Central Europe isolates. 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.

Keywords: wheat streak mosaic virus, Triticum mosaic virus, evolutionary studies, high-throughput RNA sequencing, high plains wheat mosaic emaravirus

OPEN ACCESS

Edited by:
Akhtar Ali,
University of Tulsa, United States

Reviewed by:
Xifeng Wang,
State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection (CAAS), China
Artemis Rumbou,
Humboldt University of Berlin, Germany
Satyanarayana Tatineni,
Agricultural Research Service, United States Department of Agriculture, United States

***Correspondence:**
Shahideh Nouri
shahidehnr@ksu.edu

†Present address:
Savannah Phipps,
Department of Crop and Soil Sciences, College of Agriculture, Washington State University, Pullman, WA, United States

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *22 April 2021*
Accepted: *24 June 2021*
Published: *30 July 2021*

Citation:
Redila CD, Phipps S and Nouri S (2021) Full Genome Evolutionary Studies of Wheat Streak Mosaic-Associated Viruses Using High-Throughput Sequencing. Front. Microbiol. 12:699078. doi: 10.3389/fmicb.2021.699078

INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the leading staple crops in the world. In 2019, the wheat production in Kansas estimated by United States Department of Agriculture's National Agricultural Statistics Service (USDA NASS) was $1.37 billion (National Agricultural Statistics Service (NASS), 2020). Kansas is the second leading producer of wheat behind North Dakota (National Agricultural Statistics Service (NASS), 2020) in the United States. Unfortunately, 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 total of $76 million in yield loss to Kansas farmers (Hollandbeck et al., 2017).

Wheat streak mosaic is a disease complex, which 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 (Slykhuis, 1955; Seifers et al., 1997; Seifers et al., 2009). WSMV and TriMV are type species classified under the *Potyviridae* family and are both filamentous viruses with positive-sense, single-stranded RNA genomes (Stenger et al., 1998; Fellers et al., 2009; Tatineni et al., 2009). In contrast, HPWMoV belongs to the *Fimoviridae* family, which is a multipartite, negative-sense virus consisting of eight RNA segments (Tatineni et al., 2014; Stewart, 2016). The typical symptoms of WSM caused by any of the three viruses in single infections are similar: yellow, mosaic-like streaks on the leaves (**Figure 1**), which lead to chlorosis and reduction in photosynthetic capabilities. Severe infection may also lead to stunted growth (**Figure 1**; Singh et al., 2018). For this reason, it is difficult to differentiate the causal virus phenotypically from symptoms, and serological or molecular biology techniques such as ELISA and RT-PCR are required to determine which virus or mixed-infection of viruses is 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 in the family *Potyviridae* and the genus *Tritimovirus* (Stenger et al., 1998). The genome of WSMV is ∼9.3kb in size and encodes one 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; Tatineni et al., 2011; Tatineni and French, 2014, 2016; Singh et al., 2018). The 5′ and the 3′ termini contain a VPg and a Poly (A) tail, respectively (Singh et al., 2018). Previous phylogenetic study based on the coat protein sequence of WSMV divided isolates into four different clades (clades A-D) based on their geographic regions (Stenger et al., 2002; Stenger and French, 2009). The U.S. isolates were placed in clade D and divided into four sub-clades, in which Kansas isolates were distributed throughout the clade (Stenger et al., 2002).

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*: (Fellers et al., 2009; Tatineni et al., 2009). The genome size of TriMV is ∼10.2 kb and, similar to WSMV, encodes a large polyprotein that is cleaved into 10 mature proteins (Fellers et al., 2009). In contrast to WSMV, TriMV has an unusual long 5′ untranslated region (UTR) spanning to 739 nt (Fellers et al., 2009; Tatineni et al., 2009). There are currently no studies demonstrating the phylogeny and genetic variation of TriMV.

HPWMoV, the other documented virus associated with WSM, was first described in 1993 as the causal agent of the high plains disease infecting maize and wheat in the Great Plains (Jensen et al., 1996). However, the genome sequence and organization of HPWMoV were not determined until 2014 (Tatineni et al., 2014). HPWMoV consists of eight 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 is the nucleocapsid protein, RNA 4 is 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).

To date, three resistant genes against WSM have been identified: *Wsm1* and *Wsm3* against both WSMV and TriMV isolates and *Wsm2* only against WSMV isolates (Triebe et al., 1991; Liu et al., 2011; Lu et al., 2011). In 2019, a WSMV resistant-breaking isolate has been reported and confirmed to overcome *Wsm2* in Kansas (Fellers et al., 2019). Potential resistant-breaking isolates for WSMV and TriMV have also been found to infect resistant varieties in the field (Kumssa et al., 2019). These events place a greater importance in understanding the current genetic structure of natural viral populations associated with WSM in order to determine the distribution of the associated viruses in fields and determine the major evolutionary forces acting upon WSM viruses.

In this study, we determined the current distribution and prevalence of WSM-associated viruses mainly in Kansas fields and some fields in other parts of the Great Plains and assessed the source of genetic variation of the WSM viruses. Additionally, and for the first time, the phylogenetic relationship among WSMV isolates was investigated based on the full genome sequence of historic and field isolates in this study. We also generated the first phylogenetic analysis of TriMV isolates using the whole genome sequence. The complete sequences of eight RNA segments of a new HPWMoV isolate from Kansas were reported as well.

MATERIALS AND METHODS

Wheat Survey and Sample Collection

In 2019, symptomatic and asymptomatic wheat leaf samples were collected through field surveys and sample submissions to the Kansas State University (KSU) Plant Disease Diagnostics Lab (**Supplementary Table 1**). In addition to these samples, historic WSMV samples were provided by the Agricultural

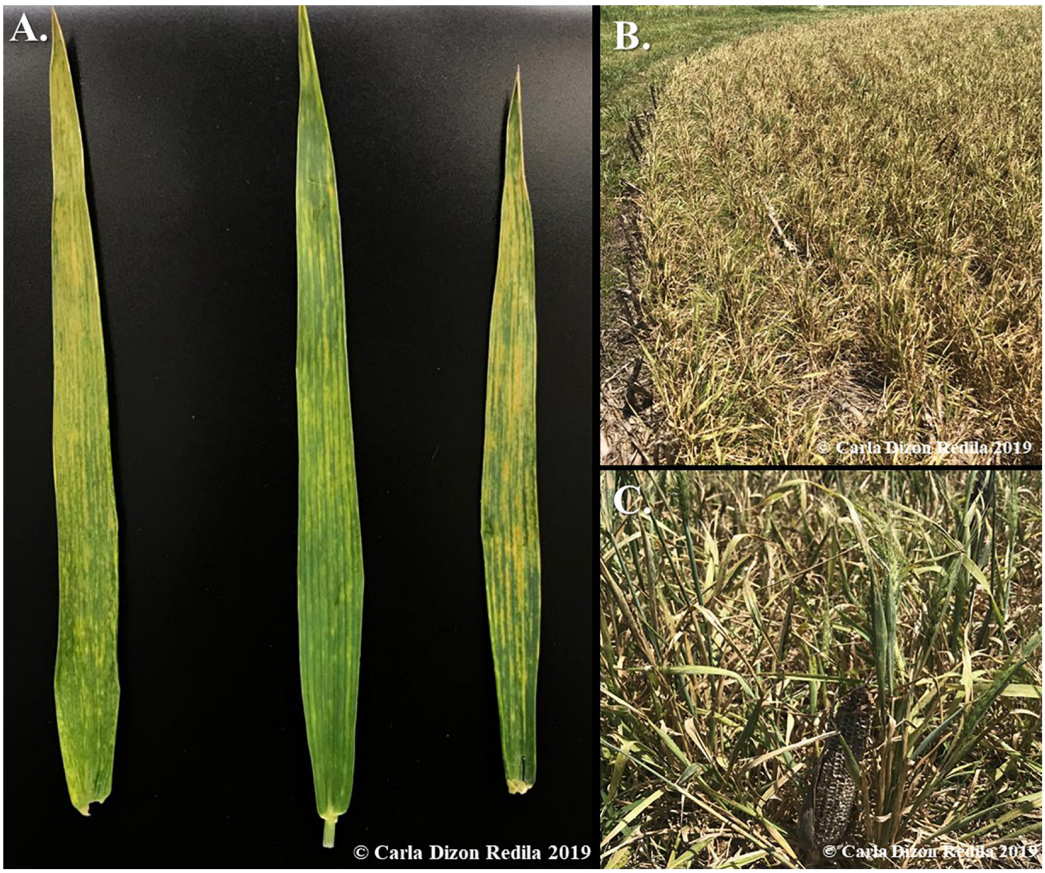


FIGURE 1 | Observed symptoms of WSM in the field. **(A)** Typical viral symptoms of WSM is yellow, mosaic-like streaks on leaves. **(B)** Severe symptoms of mixed infections of WSMV + TriMV include stunting, which leads to the underdevelopment and total loss of the crops. **(C)** A close-up image of 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 three times higher.

Research Center in Hays, KS (**Supplementary Table 1**). A few wheat samples from Nebraska and Colorado, other major wheat growing regions in the Great Plains, were also received and included in the study. In 2020, field samples were received from the KSU Plant Disease Diagnostics Lab and wheat samples from Montana were also obtained (**Supplementary Table 1**).

Screening Samples for WSM-Associated Viruses

Total RNAs were isolated from leaf tissues using TRIzol reagent (Invitrogen, CA, United States), according to the manufacturer's instructions. The extracted RNAs from the selected samples were treated with DNase I (Zymo Research, CA, United States). The first strand cDNAs were synthesized using the SuperScript II Reverse Transcriptase (Invitrogen, CA, United States). OligodT or gene specific primers (**Supplementary Table 2**) were used for the PCR step for each virus. PrimeSTAR GXL Premix (Takara Bio, CA, United States) was used to carry out a 25 μl reaction containing 1× PrimeSTAR GXL Buffer, nuclease-free water, and 0.5 μM each of the gene specific primers. The thermal cycler program used is as follows: 98°C for 2 min, 34 cycles of 98°C for 10 s, 55°C for 15 s, and 68°C for 2 min, and 72°C for 5 min to amplify ∼2 kb products.

For HPWMoV, one step RT-PCR was conducted to screen the samples. For the one step RT-PCR, the 25 μl reactions contained 1× GoTaq Flexi Buffer (Promega, WI, United States), 1 μM MgCl2, 0.1 μM dNTP, 0.4 μM of gene specific primers (**Supplementary Table 2**), 1.25 U GoTaq Flexi DNA Polymerase (Promega, WI, United States), 200 U of the SuperScript IV Reverse Transcriptase (Invitrogen, CA, United States), 40 U RNaseOUT, and DEPC treated water. The thermal cycler program used is as follows: 42°C for 10 min, 94°C for 2 min, 34 cycles of 94°C for 10 s, 55°C for 30 s, and 72°C for 1 min, and 72°C for 5 min to amplify 500 bp products. The RT-PCR products were visualized on agarose gels stained with SYBR Safe (Invitrogen, CA, United States).

RNA Library Construction and Sequencing

A total of 20 field sample from 2019 (15 samples) and 2020 (5 samples) were chosen for sequencing based on the results of the virus screenings and the geographic region with mixed infections being prioritized (**Supplementary Table 3**).

Additionally, five historic samples and 1 sample each from Colorado, Montana, and Nebraska were also selected for library preparation (**Supplementary Table 3**).

The RNA integrity of the DNase treated samples was measured using Qubit 4 (Invitrogen, CA, United States) with the RNA IQ assay kit, according to the manufacturer's instructions. The quantification of the RNA was carried out using the NanoDrop Spectrophotometer (Invitrogen, CA, United States). The TruSeq Stranded Total RNA with Ribo-Zero Plant Kit (Illumina Inc., CA, United States) was utilized to deplete the rRNA and prepare the libraries for sequencing, following the manufacturer's instructions. Agencourt RNAClean XP (Beckman Coulter, MA, United States) was used to purify the samples and ensure the removal of all traces of rRNA. TruSeq RNA Single Indexes Sets A and B (Illumina Inc., CA, United States) 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, United States).

The final libraries were subjected to quality control analysis using Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, United States) and were quantified using the Qubit 4 (Invitrogen, CA, United States) with the 1× dsDNA High Sensitivity Assay (Invitrogen, CA, United States). A total of 22 libraries from 2019 field and historic collections were pooled and sequenced in two lanes (11 pooled libraries in each lane) using the NextSeq 500 (Illumina Inc., CA, United States) high-output with a read length of 1 × 75 bp at the Kansas State Integrated Genomics Facility. From our 2020 collection, a total of six libraries were pooled and sequenced in one lane using the same platform as above.

Bioinformatics Analysis to Obtain Full Genome Sequences

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 (**Supplementary Table 4**) were retrieved from GenBank[1]. The trimmed reads were mapped against the reference genomes, and the consensus sequences were extracted using the CLC Genomics Workbench 20 (Qiagen, MD, United States).

Recombination Analysis

Multiple nucleotide alignments of the consensus sequences from this study (**Supplementary Table 6**) and the complete reference genome sequences obtained from the GenBank (**Supplementary Table 5**) were conducted using the MUSCLE alignment in the Geneious Prime 2020.2.4 (Edgar, 2004)[2]. The complete genome sequence alignments were then examined using seven different algorithms integrated in the RDP5 program (Martin et al., 2015). The seven algorithms used are as follows: RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), MaxChi (Maynard Smith, 1992), BootScan (Martin et al., 2005), Chimaera (Posada and Crandall, 2001), 3SEQ (Lam et al., 2018), and SiScan (Gibbs et al., 2000). The recombination events that were significant ($p < 0.01$) for at least four out of the seven detection methods were considered as putative 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 potential recombinants obtained from the RDP5 program were utilized as query sequences. To run this analysis, sequences of the major and minor parents detected by the RDP5 along with two selected reference sequences were used following the default settings for window width of 200 and step size of 20. The cutoff value for the percent of permuted trees to accept the sample as a potential recombinant was set at 70%. The basic principle of bootscanning is that high levels of phylogenetic relatedness between query and reference sequences found in different regions of the genome may be due to "mosaicism" (Salminen et al., 1995). In addition to Bootscan, SimPlot analysis (Lole et al., 1999) was also utilized using default parameters in order to determine the recombination breakpoints. For this analysis, the major and minor parents were used as references and the recombinant as the query sequence. The crossover points of each reference sequence were deemed as the recombination breakpoint sites.

Phylogenetic Analysis

The putative 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 (Guindon and Gascuel, 2003; Darriba et al., 2012). 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 for TriMV (Guindon and Gascuel, 2003; Darriba et al., 2012). For the phylogenetic analysis, mrBayes plugin within the Geneious Prime 2020.2.4 program was used to construct Bayesian consensus phylogenetic trees using the following default parameters: heated chains of 4, heated chain temp of 0.2, burn in length of 100,000, and sampling every 200 for every 1,100,00 generations (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003).

Population Genetics Analysis

To conduct the population genetics analyses, only the complete genome sequences of Kansas isolates were utilized. Twenty-six isolates for WSMV and 10 isolates for TriMV (**Supplementary Table 6**) were analyzed using the DnaSP version 6 (Rozas et al., 2003) to calculate the population genetics parameters and genetic diversity.

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

[1] www.ncbi.nlm.nih.gov/genbank/

[2] www.geneious.com

for each encoded protein (Kumar et al., 2004; Tamura et al., 2011). Using Hyphy 2.2.4 (Pond et al., 2005), the stop codons for the 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 that are implemented in the Hyphy package were used (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 (Kosakovsky Pond and 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 cutoff 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 and 14 field-collected leaf samples from 2019 and 2020, respectively, were screened for WSM-associated viruses by RT-PCR. **Figure 2** demonstrates the collection sites of the surveyed wheat samples[3]. The number of the collected samples was smaller in 2020 because of the COVID-19 pandemic and research restrictions. Sample screening for WSM viruses using RT-PCR revealed that single infections of WSMV dominated Kansas fields at 52% (44 positive out of 84 samples) and 29% (4 positive out of 14 samples) in 2019 and 2020, respectively (**Supplementary Table 1**). This was followed by 8% (7 positive out of 84 samples) and 14% (2 positive out of 14 samples) of mixed infections of WSMV + TriMV in 2019 and 2020, respectively (**Supplementary Table 1**). A mixed infection of WSMV + TriMV + HPWMoV was detected in only one sample of 2019 and none in 2020. **Figure 3** illustrates the distribution of WSM-associated viruses across the state of Kansas. No single TriMV or HPWMoV infections were detected in either years(see text footnote 3). Only WSMV was detected in historic samples as well as three samples from Nebraska, Colorado, and Montana (**Supplementary Table 1**).

RNASeq Analysis Obtained Full Genome Sequences of WSM Viruses

In total, the average of 33 million reads were produced for each library from total RNA sequencing (**Supplementary Table 3**). All raw sequences were deposited in the GenBank under the BioProject number PRJNA722004. After trimming, an average of 26 million clean reads were obtained which were used for mapping against reference viral sequences. Over 95% of the complete genome sequences of 21 WSMV, 9 TriMV, and 1 HPWMoV from field collected samples and 5 historical WSMV samples were obtained by mapping (**Supplementary Tables 5**, **9**). Complete genome sequences were used for the rest of analyses. Two samples infected with WSMV, NT19, and KE19, did not produce enough coverage to obtain full genome sequences and were excluded from the WSMV studies. Although full genome sequences of all eight RNA segments of a HPWMoV isolate were obtained in this study (**Supplementary Table 10**), we did not perform the phylogenetic or population genetics analyses for HPWMoV in this study for two reasons: First, the prevalence of HPWMoV was very low in our survey (only one infected sample), and second, there were only five complete genome sequences of HPWMoV available in the GenBank at the time of preparation of this report.

Full Genome Sequence Alignment of WSM Viruses

WSMV

The full genome sequences were aligned with sequences obtained from the GenBank. Most of the WSMV isolates from Kansas exhibited high nucleotide similarity (≥95%) with other isolates from the U.S. and lower similarity (≥ 88%) with Central Europe and Iran isolates. However, two isolates from Kansas (KM19 and RO20) and one isolate from Nebraska (NE01_19) showed higher similarity (∼94%) with Central Europe isolates and lower similarity (∼89%) with isolates from other U.S. states.

TriMV

The full genome sequence alignment of the TriMV isolates obtained in this study with available complete genome sequences retrieved from the GenBank including Nebraska, Colorado, and other Kansas isolates revealed high sequence similarities (≥98) between isolates.

HPWMoV

Complete genome sequences of eight segments of a HPWMoV isolate from Kansas obtained in this study (**Supplementary Table 10**) were aligned individually with reference genomes from the GenBank, resulting in high similarities (≤ 95%) of all segments with isolates from Nebraska, Kansas, Michigan, and one isolate from Ohio (GG1). In comparison, three other isolates from Ohio (NW1, NW2, and W1) showed lower similarities (∼60%–70%) to the isolate obtained in this study. These results suggest that HPWMoV U.S. field isolates are genetically diverse. However, a greater number of isolates should be studied in future to gain a better understanding of the genetic diversity of natural populations of HPWMoV.

Recombination Analysis

WSMV

The WSMV recombination analysis consisted of full genome sequences of 15 field samples collected in 2019, 6 field samples in 2020, 5 historical samples (**Supplementary Table 6**), and complete reference sequences (**Supplementary Table 5**) retrieved from the GenBank. In total, 11 potential recombinants (42%) were identified for WSMV. DC19, KM19, KSH294, EL17-1183, NE01-19, NS02-19, RO20, RH20,

[3]www.mapchart.net

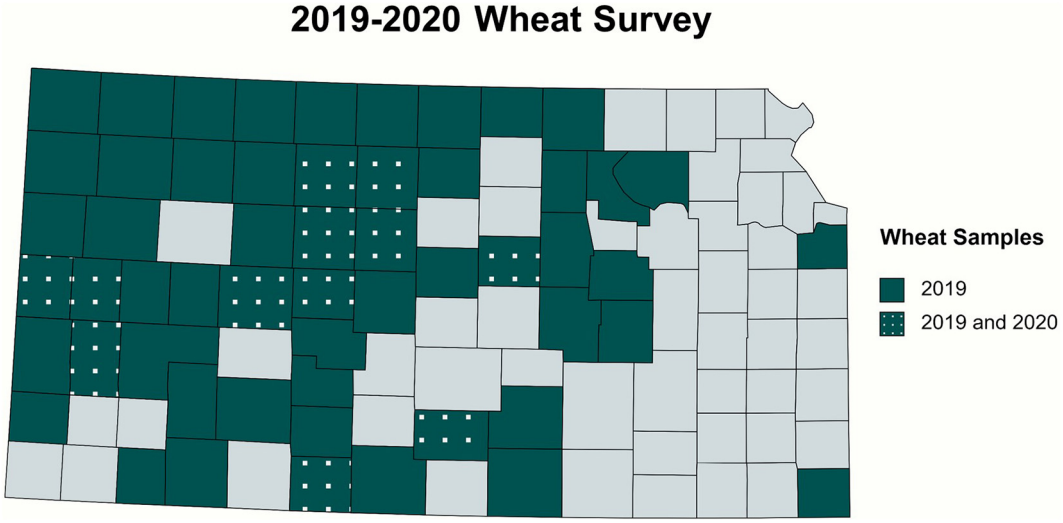


FIGURE 2 | Collection sites of the wheat samples surveyed in 2019 and 2020. Symptomatic and asymptomatic leaf samples were screened for WSM-associated viruses. The survey covered 54 counties from central and western Kansas and two eastern counties.

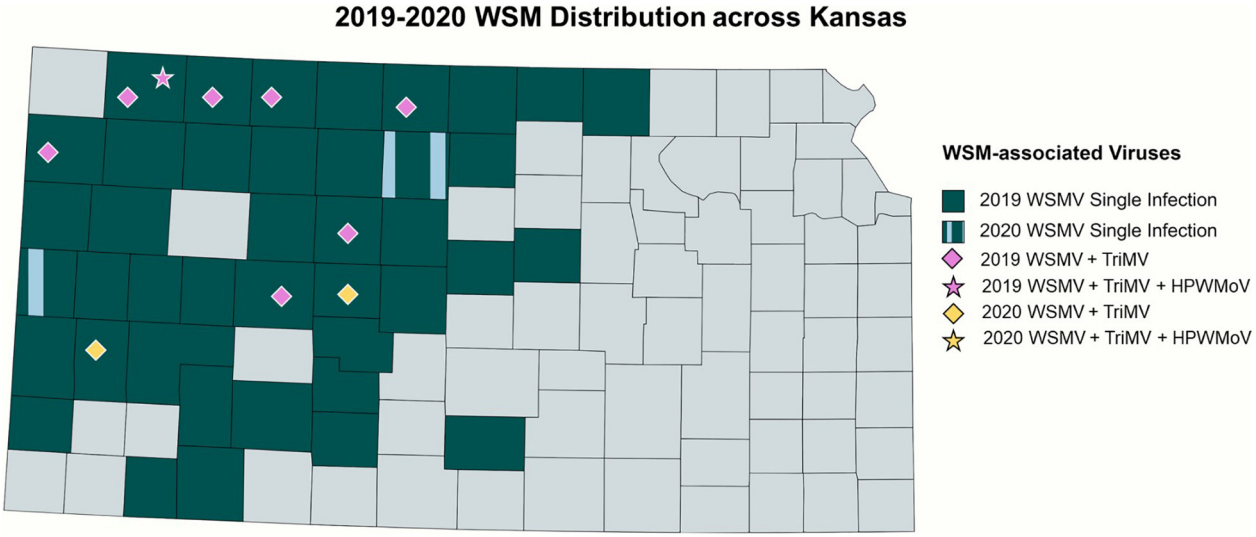


FIGURE 3 | Distribution of WSM-associated viruses in Kansas. 2019 single infections of WSMV is highlighted in dark blue, and single infections of WSMV in both 2019 and 2020 were highlighted with vertical patterns of light blue. The mixed infections of WSMV + TriMV are depicted by the diamonds, and the mixed infection of WSMV + TriMV + HPWMoV is shown in the star, with the purple color depicting the 2019 isolates and yellow for 2020.

COPhil, SM19, and KM19 isolates were detected as putative recombinants by at least five RDP5 detection methods with a significant support ($p < 0.05$; **Supplementary Table 7**). The detection methods also identified the potential parents, which included isolates from Nebraska, Washington, Czech Republic, Kansas, and Poland (**Supplementary Table 8**). The putative recombinants were confirmed using the BootScan method (**Supplementary Figure 1**) and SimPlot (**Supplementary Figure 2**). Recombination hotspots were detected in the regions of WSMV genome encoding HC-Pro, P3, NIb, CI, NIa-VPG, and P1 proteins (**Supplementary Figure 2**).

TriMV

For the recombination analysis of TriMV, genome sequences of seven field samples from 2019, two isolates from 2020 (**Supplementary Table 6**), and complete genomes obtained from the GenBank (**Supplementary Table 5**) were used. The RDP5 program found one significant putative recombinant (11%) in the 2020 isolate, RH20 (**Supplementary Table 9**), and it was confirmed using the BootScan method (**Supplementary Figure 3**). The recombination breakpoint sites were found in the NIa-Pro, CI, NIb, and 5′ UTR of the genome using the SimPlot method (**Supplementary Figure 4**).

Phylogenetic Analysis

WSMV

All putative recombinants were removed prior to phylogenetic analysis. The complete genome sequences of 10 WSMV isolates from 2019 field surveys, 3 field isolates from 2020, 4 WSMV historical samples, and reference isolates obtained from the GenBank (**Supplementary Table 5**) were used to build the phylogenetic tree. Oat necrotic mottle virus (ONMV) and Yellow oat-grass mosaic virus (YOgMV) were chosen as outgroups. The WSMV topology consists of four main clades: Clade A: an isolate from Mexico, Clade B: European isolates, Clade C: an isolate from Iran, and Clade D: United States, Argentina, and Turkey isolates (**Figure 4**). Clade D was further divided into four sub-clades (D1–D4), with Kansas isolates widely distributed in all sub-clades (**Figure 4**).

Sub-clade D1 contained isolates from the American Pacific Northwest (APNW) and the Kansas Type isolate, a previously reported isolate from Kansas (Robinson and Murray, 2013). Sub-cade D2 consisted of one isolate from Colorado, one potential resistant breaking isolate from Kansas, and both historical and 2019 isolates from Kansas. Sub-clade D3 included only Kansas isolates from this study. Sub-clade D4 included isolates from Nebraska, Colorado, Turkey, Idaho, and Kansas. There are polytomies and a small group found within sub-clade 4, which form clusters of isolates from Nebraska with Idaho and previously reported Kansas isolates and another cluster of Kansas isolates from this study.

TriMV

Phylogenetic trees were constructed with 12 TriMV isolates, including 9 isolates from this study and 6 complete genome sequences available in GenBank (**Supplementary Tables 5**, **6**). Sugarcane mosaic virus (ScSMV) and Caladenia virus A (CalVA) were used as outgroups for constructing the TriMV Bayesian tree (**Supplementary Table 5**). The topology of TriMV tree consisted of three clades (**Figure 5**). Clade A included a single isolate from 2019 KS isolate, RA02_19, which was an isolate found in the triple infection of WSM. 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 five Kansas isolates from this study, one reference isolate from 2016, one isolate from Colorado, and three previously reported potential resistant-breaking isolates from Kansas: KSGre2017, KSHm2015, and KSIct2017 (Fellers et al., 2019). Two of these potential resistant-breaking isolates are closely related and formed their own sister taxa. In addition to this, the Colorado isolate also forms a sister taxa group with a Kansas isolate from 2019 (**Figure 5**).

Population Genetics Parameters

The full genome sequences of a total of 26 and 10 Kansas isolates were used to evaluate the genetic diversity of the WSMV and TriMV, respectively (**Supplementary Table 6**). 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 DnaSP6 (**Tables 1**, **2**). The nucleotide diversity for both WSMV and TriMV isolates was relatively low with a mean of 0.035 and 0.0039, respectively.

WSMV

For WSMV, the CI gene showed the highest diversity ($\pi = 0.045$), while the CP possessed the lowest diversity ($\pi = 0.026$) (**Table 1**). The order of the average of nucleotide diversity for all encoded regions of WSMV was as follows: CI > HC-Pro > NIa-VPG > 6K2 > 6K1 > P1 > NIb > NIa-Pro > CP > P3. The degrees of tolerance for amino acid changes (dN/dS) were also calculated for each encoded region, with the 6K2 and the CP as the most and the least tolerant regions, respectively (6K2 > NIa-VPG > CI > 6K1 > NIa-Pro > P3 > NIb > HC-Pro > P1 > CP).

TriMV

For TriMV, the NIb and 6K1 genes possessed the greatest ($\pi = 0.0061$) and the lowest diversity ($\pi = 0.0012$), respectively (**Table 2**). In contrast to WSMV, the encoded regions with higher diversity for TriMV (NIb > P1 > NIa-PRO > HC-Pro > P3 > CI > NIa-VPG > CP > 6K2 > 6K1) contained greater tolerance for amino acid changes (6K1 > 6k2 > NIa-VPG > NIa-Pro > HC-Pro > CI > NIb > P3 > CP > P1).

Neutrality Tests

The ratio of dN/dS for all individual proteins was < 1 for both WSMV and TriMV isolates, suggesting purifying (negative) selection as the main selection pressure acting upon encoded proteins. To assess the selection imposed on each site (codon) of individual proteins, three different algorithms were used. Although most sites were detected under the negative selection (data not shown), sites 118 (D→N) and 2525 (G→E) located in the P1 and the NIb proteins of WSMV, respectively, were positively selected sites by two algorithms (**Table 3** and **Supplementary Figures 5**, **6**). Codon changes were detected in both historical and field isolates (**Supplementary Figures 5**, **6**). Moreover, site 2677 (L→I) located at the NIb protein of TriMV was detected as a positively selected site as well (**Table 3** and **Supplementary Figure 7**). This change was observed in only one TriMV isolate (RA02_19) (**Supplementary Figure 7**). All sites were supported significantly by two methods, FEL ($p < 0.1$) and FUBAR, with the Bayes posterior probability above 0.90 (**Table 3**).

DISCUSSION

Our study found WSMV as the most prevalent WSM-associated virus in Kansas fields in single infections followed by mixed infections of WSMV + TriMV. This result is consistent with previous surveys (Byamukama et al., 2013). While the findings of the previous studies demonstrated 91% of the TriMV infections in mixed with WSMV, our study did not find any single TriMV infections but all (100%) in mixed with WSMV. To interpret this result, we should know about the distribution of WSM in the field, which relies heavily on the successful transmission of the viruses by WCM. Transmission efficiency studies of WCM for both single and mixed infections found WCM to be efficient in transmitting WSMV alone, whereas TriMV had to

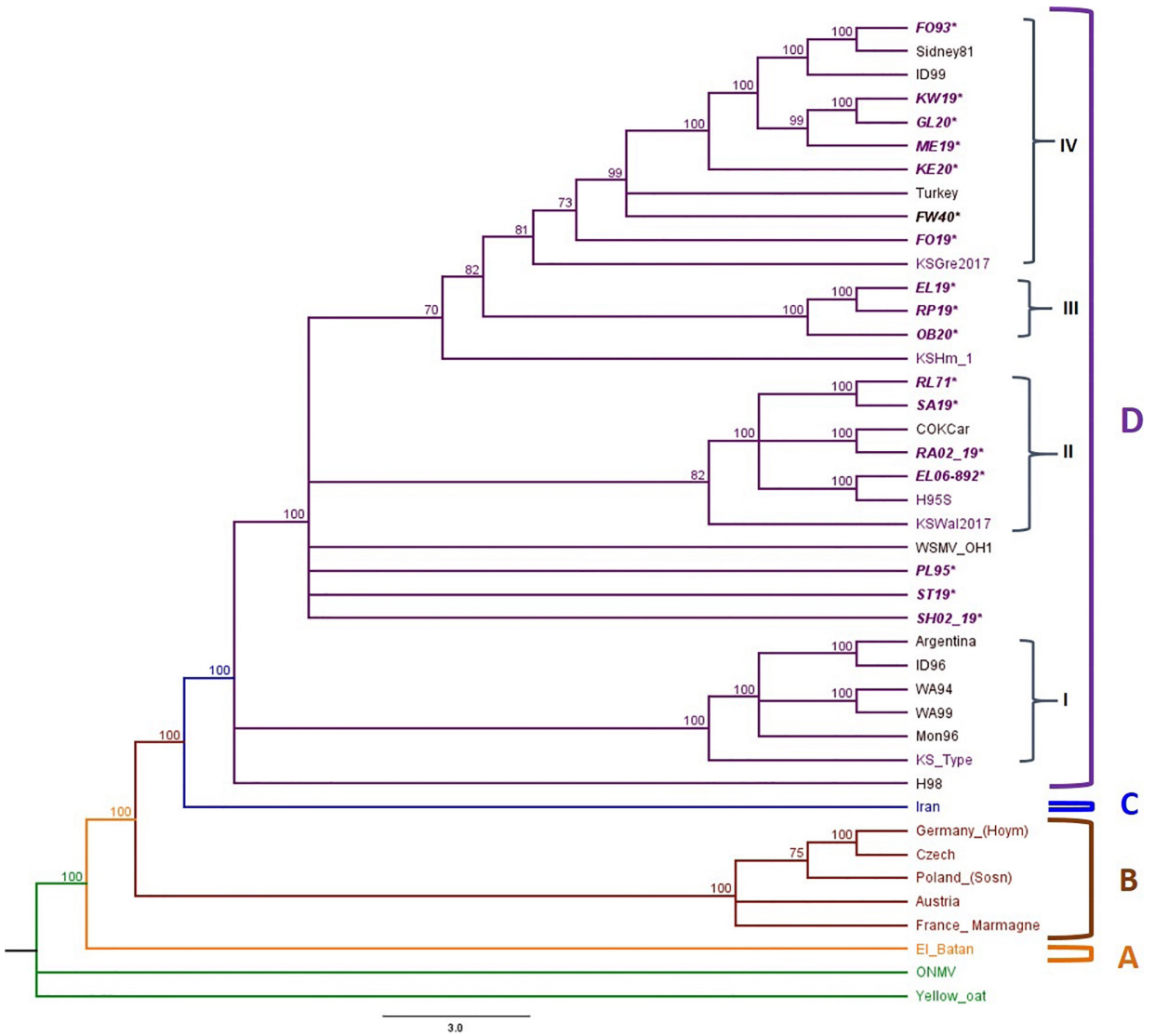


FIGURE 4 | Phylogenetic tree of WSMV isolates is divided into Clades A to C and Clade D is further divided into four sub-clades: I–IV. The sample IDs of Kansas isolates are in purple text, and the isolates sequenced in this study are written in bold and italicized text. The posterior probability of 70% was the cutoff value, and branches not supported were collapsed.

be in a mixed infection to increase the transmission efficiency (Seifers et al., 2009). In another 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 transmission efficiency of different biotypes of WCM is also different, in which Biotype 1 has a lower rate of virus transmission efficiency than biotype 2 (Oliveira-Hofman et al., 2015). Biotype 1 has been also shown to be a very poor vector of HPWMoV and could not transmit TriMV in single infections (Seifers et al., 2002; McMechan et al., 2014). Furthermore, it has been shown that the presence of WSMV in mixed infections led to increased efficiency of transmission of HPWMoV and TriMV by WCM (Seifers et al., 2002; Oliveira-Hofman et al., 2015). The presence of both WCM biotypes in Kansas wheat fields and the transmission of WSM viruses by different biotypes with the fitness advantage brought forth by mixed infections may explain the distribution of WSM associated viruses in this study and previous surveys. In addition to this, a mixed infection of WSMV and TriMV was observed to increase the titer of TriMV into the later stages of infection, which may explain the high occurrence of TriMV in mixed compared to single infections (Tatineni et al., 2019).

Our evolutionary analysis suggested recombination as the major evolutionary force operated upon field WSMV (42%) and TriMV (11%) populations. To the best of our knowledge, this is the first comprehensive analysis using full genome sequences to determine the major evolutionary mechanism of WSMV and

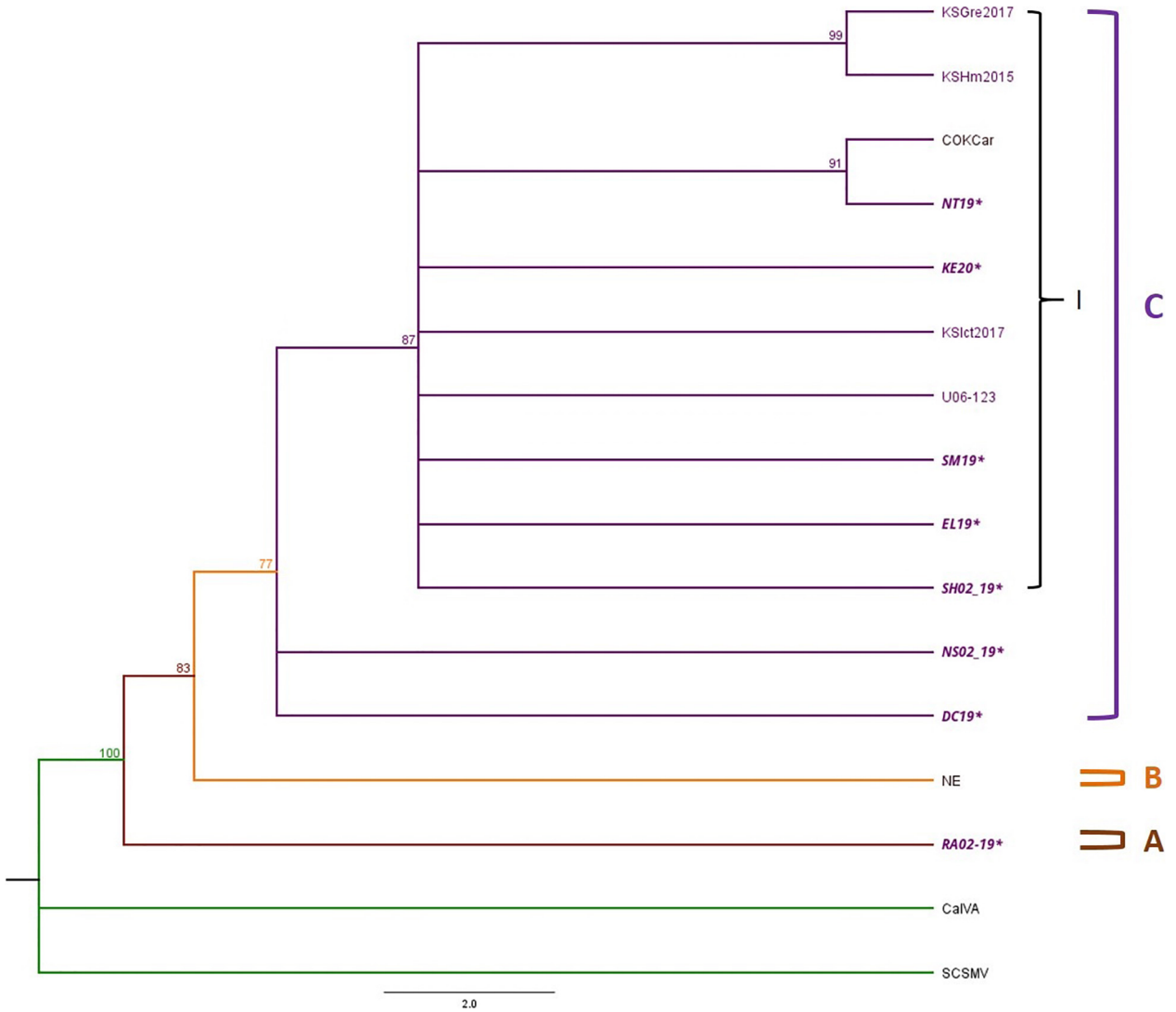


FIGURE 5 | Phylogenetic tree of TriMV isolates is divided into Clades A–C and Clade C contains sub-clade I. The Bayesian phylogenetic tree of the complete genome sequences of TriMV consisted of eight field isolates 2019 and 2020 and six isolates obtained from the GenBank. The sample IDs of Kansas isolates 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 cutoff value, and branches not supported were collapsed.

TriMV variations. The identified potential major and minor parents for WSMV putative recombinants suggested traces of isolates from other U.S. regions as well as countries from Central Europe in Kansas fields (**Supplementary Table 8**). Interestingly, two putative recombinants from Kansas and one from Nebraska were found in the same clade with Central Europe isolates with a higher nucleotide similarity compared to the rest of the U.S. isolates, suggesting close phylogenetic relationship between these isolates (**Supplementary Figure 8**). To increase our confidence about this result, we compared the sequence of the CP region of these three putative recombinants with the European isolates, and we observed that two of the three putative recombinants KM19 and NE01_19 from Kansas and Nebraska, respectively, contain three nucleotide deletions in the CP region corresponding to the Gly_{2761} codon (data not shown), which is a characteristic found in all European WSMV isolates (Gadiou et al., 2009). Taking this into account, this is the first report of European WSMV isolates found in the Great Plains fields, along with putative recombinants containing traces of these isolates in their genome. It still remains unclear about whether or not these European isolates were brought to the Great Plains directly from seed exchange with Central Europe or indirectly from the APNW, which first reported the presence of WSMV Central Europe isolates in the U.S. (Robinson and Murray, 2013).

While previous studies, which only focused on the sequence of the CP of WSMV, reported the 3′ terminus of the CP region as the recombination hotspot (Robinson and Murray, 2013), our analysis detected more hotspots in other regions of

TABLE 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 (ω)
P1	26	216	231	0.035 ± 0.007	0.053	0.12 ± 0.01	0.0068 ± 0.001	0.057
HC-Pro	26	268	297	0.044 ± 0.009	0.061	0.17 ± 0.01	0.0066 ± 0.001	0.039
P3	26	113	126	0.024 ± 0.005	0.036	0.082 ± 0.009	0.0022 ± 0.0007	0.027
6K1	26	29	32	0.036 ± 0.007	0.050	0.13 ± 0.031	0.0020 ± 0.002	0.015
CI	26	417	468	0.045 ± 0.007	0.057	0.18 ± 0.009	0.002 ± 0.0004	0.011
6K2	26	29	32	0.037 ± 0.007	0.050	0.20 ± 0.03	0.00 ± 0.00	0.000
Nla-VPg	26	131	136	0.040 ± 0.008	0.060	0.16 ± 0.01	0.0015 ± 0.0007	0.0093
Nla-Pro	26	1469	159	0.030 ± 0.00	0.057	0.10 ± 0.009	0.0023 ± 0.0008	0.023
Nlb	26	352	383	0.035 ± 0.008	0.062	0.13 ± 0.008	0.0046 ± 0.001	0.035
CP	26	187	201	0.026 ± 0.005	0.047	0.077 ± 0.006	0.0062 ± 0.001	0.081

[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 non-synonymous 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 | 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]θ_W	[5]dS	[6]dN	[7]dN/dS (ω)
P1	10	30	31	0.0055 ± 0.0009	0.0092	0.010 ± 0.002	0.0025 ± 0.001	0.25
HC-Pro	10	29	29	0.0046 ± 0.0008	0.0073	0.013 ± 0.003	0.00090 ± 0.0005	0.069
P3	10	17	18	0.0040 ± 0.0008	0.0067	0.0081 ± 0.002	0.0015 ± 0.0008	0.19
6K1	10	1	1	0.0012 ± 0.001	0.0021	0.0035 ± 0.004	0.00 ± 0.0	0.00
CI	10	38	39	0.0040 ± 0.0005	0.0069	0.011 ± 0.002	0.00084 ± 0.0004	0.076
6K2	10	1	1	0.0013 ± 0.001	0.0023	0.0040 ± 0.004	0.00 ± 0.0	0.00
Nla VPg	10	11	11	0.0037 ± 0.0006	0.0065	0.005 ± 0.001	0.0 ± 0.0	0.00
Nla Pro	10	16	16	0.0052 ± 0.001	0.0083	0.0079 ± 0.002	0.0 ± 0.0	0.00
Nlb	10	44	44	0.0061 ± 0.001	0.011	0.017 ± 0.003	0.0019 ± 0.0007	0.11
CP	10	12	12	0.0031 ± 0.0006	0.0048	0.0067 ± 0.002	0.0015 ± 0.001	0.22

[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 non-synonymous 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 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
WSMV	2525	3.5	0.077	8.80	0.19	9.2	0.97
TriMV	2677	16.66	0.076	84.64	0.19	29.62	0.94

[1]*These methods produced significant results.*

WSMV genome in addition to the CP. Unlike WSMV, only one recombination breakpoint site in TriMV was found in the 5′ UTR, which has been found to contain a translation enhancing element (Roberts et al., 2015). It is worth noting that the number of the studied TriMV isolates here was lower than WSMV because of the less incidence of TriMV in fields. Therefore, the analysis of a larger number of TriMV isolates would be needed for more accurate assessment of the recombination rate in the field. Overall, our full genome evolutionary analyses indicate that due to different functions of encoded proteins and the variability of evolutionary pressures placed upon them to increase fitness, it is crucial to conduct whole genome analyses of viruses in order to provide a thorough evolutionary study.

Furthermore, our phylogenetic analysis of WSMV using the full genome sequences placed the U.S. isolates in Clade D, which is further divided into four sub-clades. This is consistent with the

previous grouping of WSMV isolates based on the CP sequence (Rabenstein et al., 2002; Stenger et al., 2002; Stenger and French, 2009). The widespread distribution of WSMV KS isolates within sub-clades in clade D suggests that although KS isolates are closely related, there is enough diversity to group these isolates in separate clusters. The grouping of historical WSMV KS isolates with 2019 and 2020 isolates in sub-clades 2 and 4 demonstrates that the genetic structure of WSMV populations in the field has not significantly changed from past to present. However, a contrasting observation was found in sub-clade 3 and the cluster within sub-clade 4, which contained isolates only from 2019 and 2020. The opposing observations of the relationship between historical and current KS isolates show the diversity of the viral populations in KS fields and poses a problem with determining the most dominant isolate found in the field.

This study also presented the first recorded phylogenetic study of TriMV. Before the current study, the complete genome sequences of only six TriMV isolates were deposited in the GenBank, in which four of them were from Kansas (Fellers et al., 2009, 2019). Through this study, we were able to generate the full genome sequences of nine TriMV isolates from Kansas fields. Although the genetic variation of TriMV isolates was low, the phylogenetic relationship of these isolates provided an insight into the potential evolutionary pressures, which may be acting upon this virus. Interestingly, some of our 2019 field isolates were grouped together with recently reported potential resistant-breaking isolates with a high support (bootstrap value of 99%) along with a 2019 isolate (NT19) forming a sister taxa with a recently reported isolate from Colorado. This close relationship suggests that natural populations of TriMV 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 TriMV isolates would be needed to validate that claim.

Overall, we observed low genetic diversity for both WSMV and TriMV natural populations. This observation 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; Martín et al., 2006; Sun et al., 2013; Nouri et al., 2014). In fact, our findings are consistent with the concept that genetic stability is the rule in natural plant virus populations (García-Arenal et al., 2001). Genetic bottleneck during the cell to cell movement and vector transmission have been suggested to be possible reasons for the reduction in variation of WSMV (French and Stenger, 2003).

In addition, purifying selection may also be aiding in reducing diversity and maintaining a genetic stability in plant viruses. Purifying 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 1**, **2**), which is consistent with the previous studies of WSMV focusing only on the CP encoding genomic region (Choi et al., 2001; Stenger et al., 2002; Robinson and Murray, 2013). However, our comprehensive neutrality tests using the polyprotein sequences found a few positively selected sites (codons) for both WSMV and TriMV (**Table 3**). 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 may introduce changes to successfully expand host range or host and vector adaptation (Nigam et al., 2019). The two positive selection sites in WSMV were identified at the P1 and NIb regions, and for TriMV, the positively selected site was found in the NIb region. To date, no functional analysis has been performed for any of the codons associated with these specific sites. The P1 protein is the main silencing suppressor of WSMV, and its role as the pathogenicity enhancer has also been determined (Tatineni et al., 2012; Young et al., 2012). The NIb protein is the replicase, which aids in viral replication (Tatineni et al., 2009). The NIb is highly conserved among potyviruses and is often under strong purifying selection; however, mutations in the NIb were found in some potyviruses in order to adapt to different hosts (Stenger et al., 1998; Nigam et al., 2019; Shen et al., 2020). Hence, it is likely that sites 118 and 2525 have undergone evolutionary changes to successfully counter plant and/or vector defense, and also increase pathogenicity in order to adapt to infection of the different wheat varieties, WCM biotypes, and moving from the wheat to other grasses serving as the green bridge (Singh et al., 2018).

On the other hand, competitive replication may likely also pressure TriMV to introduce changes in the NIb region due to the high occurrence of mixed infection with WSMV. The NIb is also essential in interactions with viral and host proteins, leading to formation of viral replication complexes (VRCs) and also post translation, and targets plant defense pathway proteins to suppress immunity response, in addition to serving as the RdRp (Shen et al., 2020). The only TriMV isolate found in this study to have the amino acid changes in the NIb region was RA02_19, which was an isolate containing a triple infection. We hypothesize that the presence of WSMV and HPWMoV may lead into competition with TriMV in recruiting host proteins to form VRCs or trigger immune responses from the host, leading to the introduction of changes in the NIb protein. An improved understanding of the function of these regions and codons for both viruses remains as an interesting aspect that warrants further investigation.

The full genome sequences of eight segments of a HPWMoV isolate were also reported here, which based on our knowledge is the second Kansas isolate to be completely sequenced to date, but compared to the first Kansas isolate, it was isolated from wheat (**Supplementary Table 10**). In previous studies, the RNA 3 of HPWMoV was found to contain two variants: RNAs 3A and 3B (Tatineni et al., 2014; Stewart, 2016). The Kansas HPWMoV isolate from this study also contained both variants of RNA 3 (data not shown).

Taken together, the results obtained from this study demonstrated the importance of the whole genome sequence analyses to produce more informative data to study field 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 mechanisms utilized by the viruses to overcome current resistance to successfully infect the wheat.

DATA AVAILABILITY STATEMENT

The raw datasets presented in this study can be found in NCBI BioProject: PRJNA722004. The accession number(s) can be found in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

SN formulated and designed the experiments. CR and SP processed the samples and performed the experiments. CR conducted the bioinformatics and data analyses. SN and CR wrote and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by funding from Kansas State University and Kansas Wheat Commission award number KWC 2020-13.

ACKNOWLEDGMENTS

We thank Judy O'Mara, Erick DeWolf, Christian Webb, Nar Ranabhat, Dylan Mangel, Chandler Day, Jeffrey Ackerman, Marty Draper, Sarah Zukoff, Robert Whitworth, Stephen Wegulo (University of Nebraska-Lincoln), Punya Nachappa (Colorado State University), Mary Burrows (Montana State University), and Uta McKelvy (Montana State University) for their assistance in retrieving wheat samples for this study. We also want to extend our gratitude to Kansas State's Integrated Genomics Facility: Alina Akhunova, Jie Ren, and Sarah Bastian for processing and sequencing our samples and Wei Zhang for sharing her expertise in data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.699078/full#supplementary-material

REFERENCES

Andrews, S. (2010). *FastQc a Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed May 20, 2019).

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

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

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

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

Chung, B. Y.-W., Miller, W. A., Atkins, J. F., and Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5897–5902. doi: 10.1073/pnas.0800468105

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

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

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

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

French, R., and 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. doi: 10.1146/annurev.phyto.41.052002.095559

Gadiou, S., Kúdela, O., Ripl, J., Rabenstein, F., Kundu, J. K., and Glasa, M. (2009). An Amino Acid Deletion in *Wheat streak mosaic virus* Capsid Protein Distinguishes a Homogeneous Group of European Isolates and Facilitates Their Specific Detection. *Plant Dis.* 93, 1209–1213. doi: 10.1094/PDIS-93-11-1209

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

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

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

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

Hollandbeck, G., DeWolf, E., and Todd, T. (2017). *Kansas Cooperative Plant Disease Survey Report Preliminary 2017 Kansas Wheat Disease Loss Estimates*. Available online at: https://agriculture.ks.gov/docs/default-source/pp-disease-reports-2012/2017-ks-wheat-disease-loss-estimates.pdf?sfvrsn=d49587c1_0 (accessed June 10, 2020).

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

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. doi: 10.1094/PD-80-1387

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

Kumar, S., Tamura, K., Nei, M. (2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5, 150–163. doi: 10.1093/bib/5.2.150

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

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

Liu, W., Seifers, D. L., Qi, L. L., Friebe, B., and 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, 2382–2390. doi: 10.2135/cropsci2011.03.0118

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., et al. (1999). Full-Length Human Immunodeficiency Virus Type 1

Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *J. Virol.* 73, 152–160. doi: 10.1128/JVI.73.1.152-160.1999

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

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

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

Martin, D. P., Posada, D., Crandall, K. A., and Williamson, C. (2005). A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints. *AIDS Res. Hum. Retrov.* 21, 98–102. doi: 10.1089/aid.2005.21.98

Martín, S., García, M. L., Troisi, A., Rubio, L., Legarreta, G., Grau, O., et al. (2006). Genetic variation of populations of Citrus psorosis virus. *J. Gen. Virol.* 87, 3097–3102. doi: 10.1099/vir.0.81742-0

McKinney, H. (1937). *Mosaic Diseases Of Wheat And Related Cereals*. Available online at: https://archive.org/stream/mosaicdiseasesof442mcki (accessed May 20, 2019).

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

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

National Agricultural Statistics Service (NASS) (2020). *Kansas States 2019 Agricultural Overview*. Available online at: https://www.nass.usda.gov/Quick_Stats/Ag_Overview (accessed June 10, 2020).

Nigam, D., LaTourrette, K., Souza, P. F. N., and Garcia-Ruiz, H. (2019). Genome-Wide Variation in Potyviruses. *Front. Plant. Sci.* 10:1439. doi: 10.3389/fpls.2019.01439

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

Oliveira-Hofman, C., Wegulo, S. N., Tatineni, S., and 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, 1170–1174. doi: 10.1094/PDIS-08-14-0868-RE

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

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

Posada, D., and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13757–13762. doi: 10.1073/pnas.241370698

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

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

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

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

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

Salminen, M. O., Carr, J. K., Burke, D. S., and McCutchan, F. E. (1995). Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. *AIDS Res. Hum. Retrov.* 11, 1423–1425. doi: 10.1089/aid.1995.11.1423

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

Seifers, D. L., Harvey, T. L., Martin, T. J., and 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, 1161–1166. doi: 10.1094/PDIS.1997.81.10.1161

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

Seifers, D. L., Martin, T. J., Harvey, T. L., Fellers, J. P., Stack, J. P., Ryba-White, M., et al. (2008). Triticum mosaic virus: a New Virus Isolated from Wheat in Kansas. *Plant Dis.* 92, 808–817. doi: 10.1094/PDIS-92-5-0808

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

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

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

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

Stenger, D. C., and 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, 331–336. doi: 10.1007/s00705-008-0297-1

Stenger, D. C., Hall, J. S., Choi, I.-R., and 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, 782–787. doi: 10.1094/PHYTO.1998.88.8.782

Stenger, D. C., Seifers, D. L., and 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, 58–70. doi: 10.1006/viro.2001.1569

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

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

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

Tatineni, S., Alexander, J., Gupta, A. K., and French, R. (2019). Asymmetry in Synergistic Interaction Between *Wheat streak mosaic virus* and *Triticum mosaic virus* in Wheat. *MPMI* 32, 336–350. doi: 10.1094/MPMI-07-18-0189-R

Tatineni, S., and 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. *MPMI* 27, 150–162. doi: 10.1094/MPMI-09-13-0272-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, 10886–10905. doi: 10.1128/JVI.01697-16

Tatineni, S., McMechan, A. J., Wosula, E. N., Wegulo, S. N., Graybosch, R. A., French, R., et al. (2014). An Eriophyid Mite-Transmitted Plant Virus Contains Eight Genomic RNA Segments with Unusual Heterogeneity in the Nucleocapsid Protein. *J. Virol.* 88, 11834–11845. doi: 10.1128/JVI.01901-14

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

Tatineni, S., Van Winkle, D. H., and 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. doi: 10.1128/JVI.02044-10

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

Triebe, B., Mukai, Y., Dhaliwal, H. S., Martin, T. J., and 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. *Theoret. Appl. Genet.* 81, 381–389. doi: 10.1007/BF00228680

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fmicb.2021.715545

Identification of a Novel Quinvirus in the Family *Betaflexiviridae* That Infects Winter Wheat

Hideki Kondo[1], Naoto Yoshida[2†], Miki Fujita[1†], Kazuyuki Maruyama[1], Kiwamu Hyodo[1], Hiroshi Hisano[1], Tetsuo Tamada[1,2], Ida Bagus Andika[3] and Nobuhiro Suzuki[1]*

[1]Institute of Plant Science and Resources (IPSR), Okayama University, Kurashiki, Japan, [2]Agricultural Research Institute, HOKUREN Federation of Agricultural Cooperatives, Naganuma, Japan, [3]College of Plant Health and Medicine, Qingdao Agricultural University, Qingdao, China

OPEN ACCESS

Edited by:
Xifeng Wang,
State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection (CAAS), China

Reviewed by:
Mengji Cao,
Southwest University, China
Igor Koloniuk,
Academy of Sciences of the Czech Republic (ASCR), Czechia

***Correspondence:**
Hideki Kondo
hkondo@okayama-u.ac.jp

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbe and Virus Interactions With Plants, a section of the journal Frontiers in Microbiology

Received: *27 May 2021*
Accepted: *14 July 2021*
Published: *19 August 2021*

Citation:
Kondo H, Yoshida N, Fujita M, Maruyama K, Hyodo K, Hisano H, Tamada T, Andika IB and Suzuki N (2021) Identification of a Novel Quinvirus in the Family Betaflexiviridae That Infects Winter Wheat. Front. Microbiol. 12:715545. doi: 10.3389/fmicb.2021.715545

Yellow mosaic disease in winter wheat is usually attributed to the infection by bymoviruses or furoviruses; however, there is still limited information on whether other viral agents are also associated with this disease. To investigate the wheat viromes associated with yellow mosaic disease, we carried out *de novo* RNA sequencing (RNA-seq) analyses of symptomatic and asymptomatic wheat-leaf samples obtained from a field in Hokkaido, Japan, in 2018 and 2019. The analyses revealed the infection by a novel betaflexivirus, which tentatively named wheat virus Q (WVQ), together with wheat yellow mosaic virus (WYMV, a bymovirus) and northern cereal mosaic virus (a cytorhabdovirus). Basic local alignment search tool (BLAST) analyses showed that the WVQ strains (of which there are at least three) were related to the members of the genus *Foveavirus* in the subfamily *Quinvirinae* (family *Betaflexiviridae*). In the phylogenetic tree, they form a clade distant from that of the foveaviruses, suggesting that WVQ is a member of a novel genus in the *Quinvirinae*. Laboratory tests confirmed that WVQ, like WYMV, is potentially transmitted through the soil to wheat plants. WVQ was also found to infect rye plants grown in the same field. Moreover, WVQ-derived small interfering RNAs accumulated in the infected wheat plants, indicating that WVQ infection induces antiviral RNA silencing responses. Given its common coexistence with WYMV, the impact of WVQ infection on yellow mosaic disease in the field warrants detailed investigation.

Keywords: *Betaflexiviridae*, quinvirus, bymovirus, yellow mosaic disease, wheat, virome, soil borne, variants

INTRODUCTION

Wheat (*Triticum aestivum* L.) belongs to the grass family Poaceae and is one of the most important crops worldwide. Several diseases caused by vector-borne viruses have significant impacts on cereal crops, including wheat and barley (Lapierre and Hariri, 2008; Ordon et al., 2009; Jiang et al., 2020). Among these, soil-borne viruses, bymoviruses (family *Potyviridae*, phylum *Pisuviricota*), and furoviruses (family *Virgaviridae*, phylum *Kitrinoviricota*) are known as disease agents for yellow mosaic disease in wheat and barley and have been distributed widely in wheat/barley-growing regions of many countries (Kanyuka et al., 2003; Kuhne, 2009; Jiang et al., 2020). Yellow mosaic is one of the important diseases in winter barley and wheat

in Japan and China (Kashiwazaki et al., 1989; Chen et al., 1999; Han et al., 2000; Nishigawa et al., 2008; Ohki et al., 2014).

Wheat yellow mosaic virus (WYMV, a bymovirus) is the causal agent of the wheat yellow mosaic disease and is vectored by the obligate root-inhabiting *Polymyxa graminis* (a protist in the plasmodiophorid group; Inouye, 1969; Tamada and Kondo, 2013). WYMV has a bipartite plus-strand RNA genome (4.6 and 3.6 kb, respectively) that encodes large polyproteins (Namba et al., 1998; Wylie et al., 2017). WYMV and other bymoviruses require cool temperatures for multiplication, and their symptoms on infected leaves are masked when the average temperature exceeds 20°C (Cadle-Davidson and Bergstrom, 2004; Jiang et al., 2020). WYMV is widespread in the wheat production areas of Japan and China with the range of estimated yield losses due to the disease being 20–40% (Han et al., 2000; Zhang et al., 2011; Ohki et al., 2014). A closely related bymovirus – wheat spindle streak mosaic virus, which causes a similar yellow mosaic disease on winter wheat – is distributed mainly in North America and European countries (Clover and Henry, 1999; Jiang et al., 2020).

More than 57 viruses that infect wheat or barley have been identified using conventional diagnostic methods (Lapierre and Hariri, 2008). In recent years, the RNA sequencing (RNA-seq) analyses of field-grown crops, using next-generation sequencing (NGS) techniques, have enabled a more-detailed view on viral communities in the agroecosystem, often leading to new virus discoveries (Roossinck et al., 2015). Several studies have investigated the virome of wheat and barley plants using deep RNA-seq (Jo et al., 2018; Golyaev et al., 2019; Albrecht et al., 2020; Hodge et al., 2020; Singh et al., 2020), and some have identified the novel or uncharacterized wheat-infecting viruses, such as wheat stripe mosaic virus (a tentative benyvirus in the *Benyviridae*), wheat yellow striate virus (an alphanucleorhabdovirus in the family *Rhabdoviridae*), and European wheat striate mosaic virus (a putative tenuivirus in the family *Phenuiviridae*; Liu et al., 2018; Valente et al., 2019; Somera et al., 2020). Nevertheless, there is still very limited information concerning virome or viral agents in the wheat plants that are associated with yellow mosaic disease.

The members of the family *Betaflexiviridae* (order *Tymovirales*, phylum *Kitrinoviricota*) form two phylogenetically separate clades that are assigned to two subfamilies – *Trivirinae* and *Quinvirinae* (Adams et al., 2016). The subfamily *Quinvirinae* contains three genera – *Carlavirus*, *Foveavirus* and, *Robigovirus* – and three floating species, including *Banana mild mosaic virus* (recently proposed genus "Banmivirus"), *Banana virus X*, and *Sugarcane striate mosaic associated virus* (proposed genus "Sustrivirus"; Candresse et al., 2021). Quinviruses have a plus-sense RNA genome (5.8–9.0 kb; 5' capped and 3' polyadenylated ends) with six open reading frames (ORFs) for carlaviruses or five ORFs for the other two genera and floating species, with some exceptions (Caglayan et al., 2019; Zheng et al., 2020). Members of the *Quinvirinae* encode the so-called triple-gene block (TGB) for cell-to-cell movement in the leaf epidermis, and are clearly differentiated from the members of the subfamily *Trivirinae*, whose genomes encode a single movement protein (the 30K superfamily). Most quinviruses belonging to the genera *Foveavirus* (eight species) and *Robigovirus* (five species) naturally infect fruit trees, such as apple, cherry, and grapevine, while the members of the genus *Carlavirus* (53 species) infect various dicotyledonous plants (Ryu and Song, 2021; Yoshikawa and Yaegashi, 2021). Recently, several foveavirus candidates have also been reported from diseased fruit trees and garlic (*Allium sativum*) plants (Park et al., 2019; Gazel et al., 2020; Reynard et al., 2020; Yaegashi et al., 2020; Zheng et al., 2020; Luo et al., 2021). Some carlaviruses are transmitted by hemipteran insects, such as aphids and whiteflies, in non-persistent manners; while there have been no reports on the vectors of foveaviruses or robigoviruses, which raises speculation about their mechanical- or graft-transmission in the field (Ryu and Song, 2021). A quinvirus (sugarcane striate mosaic associated virus, SCSMaV) is apparently soil transmissible *via* an unknown mechanism (Choi et al., 1999; Thompson and Randles, 2001).

In the current study, a *de novo* meta-RNA-seq approach was used to investigate the virome of wheat plants sampled from a field in Hokkaido, Japan. We discovered three strains of a novel quinvirus tentatively named wheat virus Q (WVQ) that often co-infected wheat with a few known viruses. We also found that WVQ could be transmitted *via* soil and occurred at least 4 consecutive years in the field, raising the concerns about its potential impact on wheat production.

MATERIALS AND METHODS

Collection of Plant Materials

The plant materials used in this study were sampled at an experimental field in the HOKUREN Agriculture Research Institute, Naganuma, Hokkaido, Japan (43.3°N) in May and June of 2018 and 2019 (a total of 14 wheat samples or pools; **Table 1**). The leaves of wheat plants showing yellow mosaic and asymptomatic leaves, including from the cultivars "Kitahonami" (KTH) in 2018−2019, and "Kitanokaori" (KTN) and "Yumechikara" (YM, a WYMV resistant cultivar; Kojima et al., 2015) in 2019, were collected (**Supplementary Figure S1** and data not shown). The wheat plants collected in 2019 were grown on plots with or without the fungicide fluazinam (Flu; Ishihara Sangyo Kaisha, LTD.) soil treatment. Wheat (KTH) and rye (*Secale cereale* cv. Fuyumidori) samples (leaves and roots) were also collected from the same Naganuma field in 2020 and 2021 (**Table 1**). The collected plant materials were stored at −80°C until their analysis.

De novo RNA-Seq Analysis

Meta-RNA-seq analysis was basically conducted as has been described previously (Kondo et al., 2016; Lin et al., 2019). The total RNA from each wheat-leaf sample was extracted using NucleoSpin RNA Plant and Fungi Kit (Macherey and Nagel, Düren, Germany), following the manufacturer's instructions. The obtained RNA fractions from the 2018 (three samples: KTH-18-6, −7, and −8) and 2019 (three samples: KTH-19-2, −4, and −6) samples were pooled into two groups

TABLE 1 | List of field-collected samples used in this study and the viruses identified by RNA-seq.

No.	Sample name[1]	Wheat cultivar or plant name	Collected date	NGS group[2]	Virus[3]
2018					
1.	KTH-18-1	cv. Kitahonami	15 May	poo1-18si	W/Q
2	KTH-18-2	cv. Kitahonami	15 May	poo1-18si	W/Q
3.	KTH-18-3	cv. Kitahonami	15 May	NA	W/Q
4.	KTH-18-4	cv. Kitahonami	15 May	NA	W/Q
5.	KTH-18-5	cv. Kitahonami	15 May	NA	W/Q
6.	KTH-18-6 (1+2)	cv. Kitahonami	5 June	pool-18L	Q
7.	KTH-18-7 (3+4)	cv. Kitahonami	5 June	pool-18L	W/N/Q*
8.	KTH-18-8 (5+6)	cv. Kitahonami	5 June	pool-18L	W/Q*
2019					
9.	KTH-19-1	cv. Kitahonami	18 June	NA	Q
10.	KTH-19-2 (Flu)[4]	cv. Kitahonami	18 June	pool-19L	W/Q
11.	KTN-19-3	cv. Kitahonami	18 June	NA	Q
12.	KTN-19-4 (Flu)[4]	cv. Kitahonami	18 June	pool-19L	(Q)
13.	YM-19-5	cv. Yumechikara	18 June	NA	Q
14.	YM-19-6 (Flu)[4]	cv. Yumechikara	18 June	pool-19L	Q
2020					
15.	KTH-20-1	cv. Kitahonami	11 May	NA	W/Q
16.	Rye-20-2	Rye (*Secale cereale*)	11 May	NA	(W)/Q*
2021					
17.	KTH-21-1	cv. Kitahonami	19 April	NA	W/Q
18.	Rye-21-2	Rye (*Secale cereale*)	19 April	NA	(W)/Q*

[1]All plant materials were taken from an experimental field at HOKUREN Agricultural Research Institute, Naganuma, Japan (43.3°N).

[2]Total RNA fractions from the leaf samples were pooled according to the year of collection (pool-18L and pool-19L). Both pooled samples were used for transcriptomic sequencing, while poo1-18si was used for small RNA sequencing (RNA-seq). NA, not applicable.

*[3]The virus infection was assayed via RT-PCR (see the results in **Figures 1, 6**). W, wheat yellow mosaic virus (WYMV); N, northern cereal mosaic virus (NCMV); and Q, wheat virus Q (WVQ, a novel quinvirus). Parentheses indicate faint PCR products. Asterisks indicate strain c was undetected.*

[4]Sampled from an experimental plot treated with fluazinam, which is a broad spectrum fungicide.

based on the collection year – pool-18L (total 15.8 μg; RNA integrity number, RIN = 7.9) and pool-19L (total 7.8 μg; RIN = 7.7), respectively (**Table 1**). The two sample pools were subjected to ribosomal RNA (rRNA) depletion using a Ribo-Zero kit (Illumina, San Diego, CA, United States), and were subsequently used for the synthesis of a cDNA library using TruSeq Stranded Total RNA LT Sample Prep Kit (Plant; Illumina). The two cDNA libraries were then subjected to RNA-seq using the Illumina HiSeq 2000 platform (Illumina, 100-bp pair-end reads) performed by Macrogen Corp. Japan (Tokyo, Japan). After RNA-seq, the sequence reads were trimmed and *de novo* assembled using the CLC Genomics Workbench version 11 (CLC Bio-Qiagen, Aarhus, Denmark) using default parameters (minimum contig length = 200; mismatch cost = 2; length fraction = 0.5; similarity fraction = 0.8). Subsequently, assembled contigs were used as queries for basic local alignment search tool (BLAST) analyses (all contigs or contigs larger than 1.0 kb were subjected to BLAST-N or BLAST-X search, respectively) against the viral genome reference sequence (RefSeq) dataset of the National Center for Biotechnology Information (NCBI[1]; E-value cut-off set >0.05). Quinvirus-like sequences from public datasets were also used as queries for BLAST searches against assembled contigs generated in this study to detect further virus-related sequences. The sequence reads (or assembled contigs) were mapped to the assembled virus or virus-like contigs using the Read Mapping algorithm using default or more stringent mapping parameters (match score = 1; mismatch cost = 2; length fraction = 0.5 or 0.9; similarity fraction = 0.8 or 0.95) in the CLC Genomics Workbench.

RNA Extraction and RT-PCR

The total RNA was extracted from the plant materials using RNAiso Plus Reagent (TaKaRa Biotech. Co., Shiga, Japan) according to the manufacturer's instructions. The cDNA was synthesized using Moloney-murine leukemia virus (MMLV) reverse transcriptase (Thermo Fisher Scientific, Waltham, MA, United States) with random primers [nonadeoxyribonucleotide mixture; pd. (N)9; TaKaRa Bio], and was then used as a template for PCR amplification with QuickTaq (Toyobo, Osaka, Japan). The conditions of PCR were as follows: 94°C for 2 min; then 30 or 35 cycles of 94°C for 10 s, 53°C or 57°C for 30 s, and 72°C for 1 or 2 min; and 72°C for 10 min. Alternatively, the RNA samples were subjected to one-step reverse transcriptase PCR (RT-PCR) analysis by using PrimeScript One-Step RT-PCR Kit Ver. 2 (TaKaRa), following the manufacturer's instructions. The 5' and 3' rapid amplification of cDNA ends (5' and 3' RACE) analyses were conducted using the 5'-Full RACE Core Set and 3'-Full RACE Core Set (TaKaRa Bio), respectively, with virus-specific primers (Kondo et al., 2006; **Supplementary Table S1**). The wheat 18S rRNA gene was used as the reference target of an endogenous gene for the RT-PCR (Jarošová and Kundu, 2010). The primers used in the RT-PCR analyses are

[1]https://www.ncbi.nlm.nih.gov

listed in **Supplementary Table S1** and are available upon request. The selected PCR products were purified and subjected to Sanger sequencing to confirm their nucleotide sequences. For 5' and 3' RACE analyses, the obtained DNA products were ligated into the pGEM-T easy PCR cloning vector (Promega) and transformed into *Escherichia coli* strain DH5 alpha (TaKaRa Bio), and then plasmids were subjected to DNA sequencing.

Sequence Analysis and Database Search

Sequence data were analyzed using EnzymeX ver. 3.3.3[2] and 4peaks v1.8.[3] BLAST or reverse BLAST searches were conducted using the GenBank database through the NCBI web site running on the non-redundant (nr) DNA and protein databases from the NCBI (nucleotide collection – nr/nt; transcriptome shotgun assembly – TSA; and expression sequence tag – EST; see Footnote 1). Sequence identities were also calculated using BLAST program available from NCBI (BLAST-N suite-2 sequence program; see Footnote 1). The conserved domains were searched using the NCBI conserved domain database.[4] Pairwise sequence comparisons (PASCs) were made using the Sequence Demarcation Tool (SDT) ver. 1.2 (Muhire et al., 2014). A genome-based web tool for virus classification (pairwise sequence comparison – PASC) was used for the novel virus sequence analysis[5] (Bao et al., 2014).

Phylogenetic Analysis

For the phylogenetic analyses, maximum-likelihood (ML) tree was constructed, as described in Kondo et al. (2019, 2020), with minor modifications. Multiple amino-acid sequence alignments were generated by using MAFFT online ver. 7[6] (Katoh and Standley, 2013), with poorly aligned sites being removed using Gblocks online version 0.91b[7] (Talavera and Castresana, 2007). Phylogenetic trees were then constructed using the PhyML 3.0 online program[8] (Guindon et al., 2010; Lefort et al., 2017). Neighbor-joining (NJ) trees (Saitou and Nei, 1987) were constructed based on the MAFFT alignments masked with Gblocks. The reliability of the branches was obtained from 100 bootstrap replicates. The trees were refined using FigTree ver. 1.3.1 software.[9]

Small RNA-Seq

Small RNA deep sequencing was performed as described by Shahi et al. (2019). Two total RNA samples from 2018 (KTH-18-1 and −2) were mixed together (total 14.6 μg; RIN = 7.8) and subjected to small RNA seq (**Table 1**). After cDNA library preparation using a TruSeq Small RNA Library Prep Kit (Illumina), deep RNA-seq was conducted using an Illumina HiSeq 2500 (Illumina, 150-bp single-end reads) by Macrogen Corp., Japan. The raw sequence reads (total read number, 46,026,536; total read bases, 6,950-Mb) were trimmed of adapters and filtered for low-quality sequences and size range (15 to 32 nt, in length) using the CLC Genomic Workbench. The clean reads were subsequently mapped into the virus genomes. The virus-derived small RNA reads were used for further analysis using the program MISIS-2 (Seguin et al., 2016).

Virus Transmission Tests

For soil transmission, the soils that were collected from the field with WYMV-infected wheat plants (KTH-20-1) were used as a virus inoculation source. Commercial culture soil (Nihon Hiryo Co., Ltd., Tokyo, Japan) was also used for non-infested healthy soils. Wheat seeds (cv. "Kitahonami") were sown in plastic pots (7.5 cm in diameter) containing a mixture of the infested soil and the culture soil, and were grown in a growth cabinet (around 16°C, 12 h light/12 h dark) or in a greenhouse (non-controlled temperature conditions, around 15–25°C or occasionally slightly higher). At different periods after sowing, total RNA extracted from the roots of the plants was subjected to RT-PCR.

RESULTS

Identification of Virus-Like Sequences From the Wheat-Leaf RNA-Seq Analysis

RNA sequencing analysis of the two pooled, rRNA-depleted RNA preparations of wheat-leaf samples (**Table 1**; **Supplementary Figure S1**) resulted in totals of 55,585,610 (5,614 Mb; pool-18L) and 64,002,754 (6,464 Mb; pool-19L) raw reads, respectively. *De novo* assembly using the CLC Genomics Workbench yielded 124,777 (pool-18L) and 131,004 (pool-19L) sequence contigs. These assembled contigs were then subjected to local BLAST-N (all contigs as queries) and BLAST-X (contigs larger than 1.0 kb as queries) analysis against the viral RefSeq collection. At least, 86 contigs (44 from pool-18L and 42 from pool-19L, ranging from 230 to 8,590 nt) were indicated as candidates for virus-derived sequences. The 41 contigs had a sequence similarity to WYMV (a bymovirus; 11 contigs), northern cereal mosaic virus (NCMV, a plant rhabdovirus; 2 contigs), and betaflexiviruses (at least 28 contigs; **Supplementary Table S2**). The remaining 43 contigs were related to fungal viruses of the families *Narnaviridae*, *Botyourmiaviridae*, and *Mymonaviridae* and some others (data not shown). Approximately, 1.1% of the reads (636,069 reads) in pool-18L and 0.9% of the reads (550,442 reads) in pool-19 were assigned to virus-related reads in each library. Almost all the virus-associated reads were derived from plant viruses – 43.5% (515,962 reads) for WYMV, 0.4% (4,216 reads) for NCMV, and 53.3% (632,472 reads; several unassembled reads related to a betaflexivirus strain might also be presented in the two pools, see below) for betaflexiviruses, with 2.9% (33,861 reads) representing others, likely derived from fungal-associated viruses (**Supplementary Table S2** and data not shown).

Characterization of Two Known Wheat RNA Viruses

We obtained at least 11 contigs of WYMV RNA segments from the pool-18L and pool-19L libraries. Among these, two

[2]https://nucleobytes.com/enzymex/index.html
[3]http://nucleobytes.com/index.php/4peaks
[4]https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
[5]https://www.ncbi.nlm.nih.gov/sutils/pasc/viridty.cgi
[6]http://mafft.cbrc.jp/alignment/server/
[7]http://molevol.cmima.csic.es/castresana/Gblocks_server.html
[8]http://www.atgc-montpellier.fr/phyml-sms/
[9]http://tree.bio.ed.ac.uk/software/

contigs (Wh18L_c88 and Wh19L_c362) were nearly complete sequences of the WYMV RNA1 segment, while the RNA2 coding-complete sequence was generated using partially overlapping contigs in each dataset (Wh18L_c38/1970/1462 and Wh19L_c740/2865/2864, respectively; **Supplementary Figure S2A** and **Supplementary Table S2**). The WYMV sequences obtained from both libraries were almost identical, and the sequences were confirmed by mapping the reads (the representative contig sequences of RNA1 and RNA2 have been deposited in DDBJ under Accession Nos., LC63269 and LC63270, respectively; **Table 2**). The WYMV segments showed their highest sequence identities with a Japanese isolate Morioka (Accession No., AB627810) for RNA1 (99.6% nucleotide sequence identity) and a Chineses isolate (Shandong, KY354407) for RNA2 (99.0% nucleotide sequence identity).

To verify the presence of WYMV in the RNA samples used for the meta-RNA-seq analyses (**Table 1**), we performed RT-PCR, using the specific primer sets for RNA1 sequences (**Supplementary Table S1**). The typical yellow mosaic symptoms on the leaves of the cultivar "Kitahonami" samples are shown in **Supplementary Figure S1**. WYMV was detected in most of the 2018 "Kitahonami" samples, while one of two 2019 "Kitahonami" samples was positive (**Figure 1**). A phylogenetic analysis of the WYMV isolates, based on their RNA1 sequences, revealed that the

TABLE 2 | Annotated virus contigs from the wheat-leaf RNA-seq analyses.

Virus or tentative virus name	Contig or concatenated name	Size (nt)	Mapped reads no. pool-18/ pool-19[1]	Accession no.
WYMV RNA1	Wh19L_c362	7,635[2]	199,947/111,012	LC632069
WYMV RNA2	Wh19L_c740/2865/2864	3,643[2]	142,064/65,799	LC632070
NCMV	Wh18L_c253/20097	13,222[2]	4,186/0	LC632071
WVQ strain a		8,412[3]	166,410/329,116	LC632066
WVQ strain b		8,411[3]	178,802/70,997	LC632067
WVQ strain c		(~725)[4]	(12,531)/(14,320)	LC632068

[1]The data were obtained by the Read Mapping algorithm with following parameters: match score = 1; mismatch cost = 2; length fraction = 0.9; and similarity fraction = 0.95.
*[2]See **Supplementary Figures S2A,B** and **Supplementary Table S2** for the details of the contigs in each virus. Both WYMV and NCMV sequences were not included their terminal regions.*
*[3]The entire genomic regions of WVQ strains a and b were determined using RT-PCR and RACE (see **Figure 2**). The GG (strain a) or G (strain b) nucleotides in the majority of the 5' RACE clones are likely derived from a 5'-end cap structure.*
*[4]A partial genomic sequence of the third strain c of WVQ (see **Supplementary Figure S2C**) was also used for map-read analysis.*

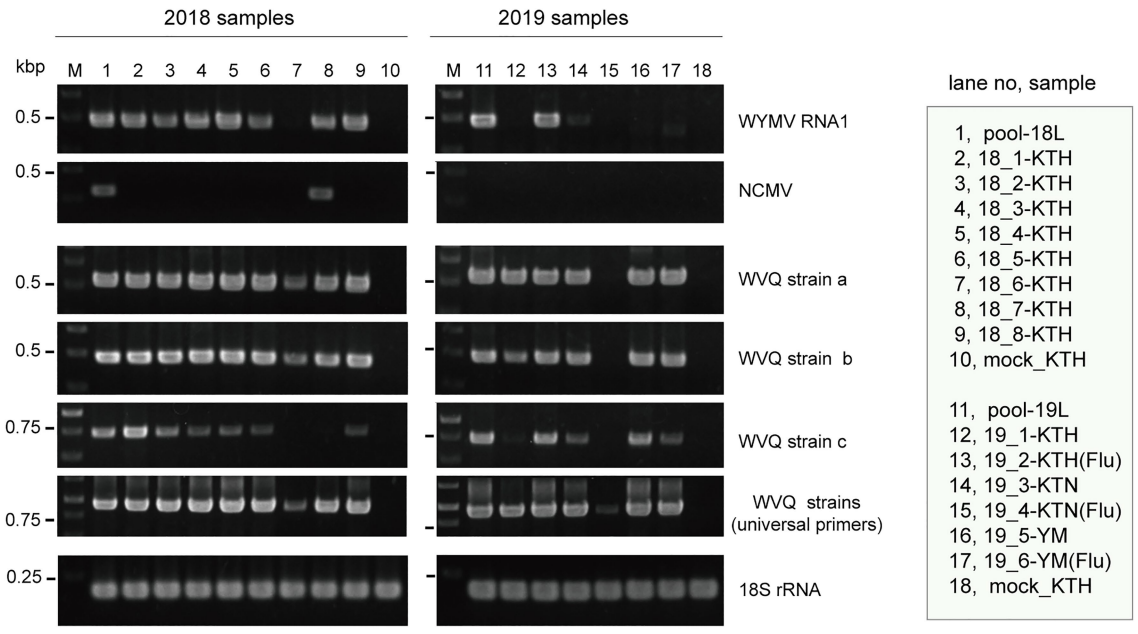


FIGURE 1 | Reverse transcriptase PCR (RT-PCR) detection of viruses in wheat leaf RNA samples. Sample names are shown in the right-hand box. Lane 1 (pool-18L) and lane 11 (pool-19L) are two pooled RNA samples subjected to NGS analysis. Lanes 10 and 18 are RNA samples of mock-inoculated plants. KTH, cv. Kitahonami; KTN, cv. Kitanokaori; YM, cv. Yumechikara; and Flu, fluazinam treatment to the soil. The wheat 18S ribosomal RNA (rRNA) gene was used as an endogenous RNA reference. The primers used in the RT-PCR analyses are listed in **Supplementary Table S1**. Selected PCR products were sequenced to confirm the amplification of the virus targets. The PCR products were analyzed in 1.5% agarose gel electrophoresis. The gel was stained with ethidium bromide and viewed under UV light.

obtained sequences belonged to genotype B, whose members are widely distributed in wheat-growing regions in northern Japan, including Hokkaido island (Ohki et al., 2014; **Supplementary Figure S3A**). In Japan, the WYMV isolates have been divided into three distinct pathotypes (I–III) based on their pathogenicity to wheat varieties, with pathotype II (genotype B) presenting in northern Japan, including Hokkaido island (Ohto et al., 2006). The WYMV isolate from the Naganuma field likely belongs to pathotype II, and symptoms could not be confirmed in the resistant wheat variety "Yumechikara," in which a major quantitative trait loci, designated as *Q. Ymym*, against WYMV on chromosome 2D is known (Kobayashi et al., 2020; **Figure 1**; **Supplementary Figure S3A** and data not shown).

Two negative-sense RNA virus-like contigs (Wh18L_c253 and Wh19L_c20097) were identified from the pool-18L library (**Table 2** and **Supplementary Table S2**). These two contig sequences are related to the NCMV (a planthopper-vectored rhabdovirus, genus *Cytorhabdovirus*, family *Rhabdoviridae*) and cover nearly the entire region of its genomic RNA (**Supplementary Figure S2B**). The concatenated contig sequence (Wh18L_c253/20097; deposited in DDBJ under Accession No., LC632071; **Table 2**) showed a high level of sequence identities with those of the Japanese (Accession No. AB030277) and Chinese (Hebei, GU985153) isolates of NCMV (98.5 and 93.8% nucleotide sequence identities, respectively). The distribution pattern of the mapped reads on the NCMV reference genome (4,186 mapped reads; **Table 2**) likely reflects the transcription gradient (a 3'–5' polar gradient of mRNA production), in which the transcripts of the N protein mRNA are the most abundant, with those of the downstream genes being at gradually lower levels (Dietzgen et al., 2017; Horie et al., 2021; **Supplementary Figure S2B**). The NJ tree, based on the L protein (RdRP), showed that the NCMV isolates are placed within a large sub-clade of cytorhabdoviruses (tentatively named subclade I), and formed a tight cluster together with other cereal rhabdoviruses, including barley yellow striate mosaic virus and maize yellow striate mosaic virus (**Supplementary Figure S3B**). Among the tested RNA samples in the pool-18L, NCMV was detected in one "Kitanokaori" sample (18_4-KTH; **Figure 1**). NCMV is known to be a viral agent of the stunted rosette symptom in wheat in East Asian countries (Tanno et al., 2000); however, the "Kitanokaori" plant (18_4-KTH) that was infected with the virus showed no stunting symptoms typical of the NCMV infection (data not shown). This may be due to the lower titer of NCMV in the host plants or other unknown reasons.

Genome Analysis of a Novel Quinvirus and Detection of Its Strains

We identified 28 contigs in the two datasets (14 contigs each from the two sample pools, ranging from 203 to 3,828 nt) related to quinviruses (genera *Foveavirus* and *Carlavirus*, subfamily *Quinvirinae*) using the local BLAST analysis on the two datasets (**Supplementary Table S2**). Two concatenated sequences related to quinviruses were obtained by using the overlapped contigs, which share high nucleotide sequence identities at their overlapping regions with neighboring ones (**Figure 2** and **Supplementary Table S2**). Subsequently, the entire genomic RNA region was identified by RT-PCR and Sanger sequencing using KTH-19-1 samples, following the 5' and 3' RACE analyses (**Figure 2**; **Supplementary Table S1** and data not shown). The complete genomic sequences of the putative quinvirus were 8,412 and 8,411 nt in length, excluding the poly(A) tail; these were designated "strain a" and "strain b," respectively (**Table 2**, Accession No., LC632066 and LC632067, respectively). In the 5' RACE clones, the GG (strain a, 9/16 clones) and G (strain b, 6/7 clones) nucleotides were mainly identified at the 5'-terminal ends of two genomic RNAs (data not shown). Both 5'-GG and 5'-G nucleotides at the 5'-terminal ends were most likely to be derived from 5'-end cap structure, as previously reported (Hirzmann et al., 1993; Kondo et al., 2014).

The genome organization of strain a was similar to that of foveaviruses and some others, with five ORFs, encoding replicase, TGB proteins, and CP, while strain b has an additional small ORF (ORF6) overlapping the CP gene (**Figure 2**). The presence of strain b with sequence variations within the 3'-terminal region were identified by the sequencing of the cloned 3' RACE products. The additional "A" residues presented at the potion 8,351 in some clones (UUAAAAA$_{8351}$AnCCC…, additional A residues different from deposited one was underlined; of 12, 2, 1, and 1 3'-RACE clones had A_1, A_4, and A_5; data not shown). These sequence variants lacked the extra ORF (ORF6) and it seems to be a minor population in the tested wheat sample (KTH-19-1), as the majority of the RACE clones (8/12), which would retain ORF6, had no additional A. The replicase of WVQ does not contain an alkylation B (AlkB)-like domain, which is commonly present in betaflexiviruses that infect perennial plants, such as fruit crops (Kondo et al., 2013; Moore and Meng, 2019; **Figure 2**). They showed moderate levels of nucleotide (79.5% for the entire genome) and amino acid sequence identities (68.6∼91.7%) to each other (**Table 3**). Since different species of quinviruses should have less than about 72% nt identity or 80% aa identity for quinviruses (King et al., 2011), these two sequences likely represent the strains of a single quinvirus species. It should be noted that there were other overlapping fragments (contigs 133 and 431 in pool-18L, marked with a white circle in **Figure 2**) with high nucleotide sequence identities (>97.0%) to the strain a reference sequence (**Supplementary Table S2** and data not shown), suggesting that this strain consists of multiple variants.

Along with these quinvirus sequences, some related contigs showed moderate levels of sequence similarity with the two strains (share 72.1–85.0% nucleotide sequence identity with the reference sequences), indicating the presence of a third strain (designated "strain c") in the sample pools (**Table 2**; **Supplementary Figure S2C**; **Supplementary Table S2** and data not shown). To confirm the presence of these three strains in the RNA samples, RT-PCR and subsequent direct Sanger sequencing were performed using strain-specific primer sets (**Supplementary Table S1**). A partial nucleotide sequence has been deposited as a representative of strain c sequence in DDBJ under Accession No., LC632068 (**Table 2**). All three strains were amplified in most of the tested samples in both pool-18L and pool-19L, although the amount of amplicons

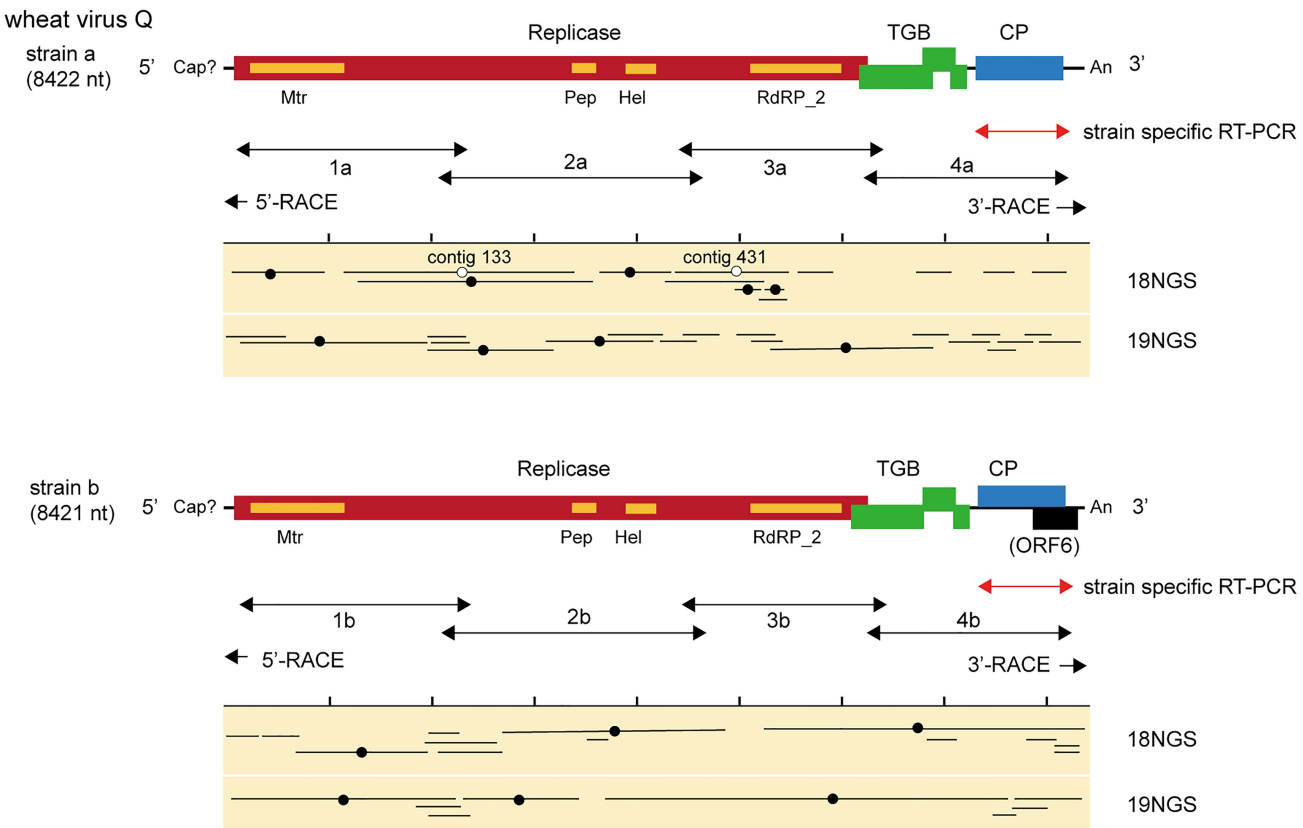


FIGURE 2 | Virus sequences of novel quinvirus (wheat virus Q, WVQ) strains identified from the RNA sequencing (RNA-seq) of wheat leaves. The black lines in the genomic RNA show ORFs. The conserved domains in the WVQ replicase are shown (Mtr, methyltransferase; Pep, Peptidase_C23 superfamily; HEL, RNA helicase; and RdRp_2, RNA-dependent RNA polymerase). The entire genomic sequence of the WVQ strains (a and b) was obtained from the overlapping RT-PCR products and the 5' and 3' RACE products (see the double-arrowed or arrowed lines below each schematic genome of the WVQ strains for RT-PCR target regions). A genomic region of WVQ, as indicated by the double-arrowed red lines, was used for strain-specific detection by RT-PCR in **Figures 1, 6**. The primers used in the RT-PCR and RACE analyses are listed in **Supplementary Table S1**. Virus-sequence contigs from two separate total RNA pools of wheat leaf samples (18NGS, pool-18L; 19NGS: pool-19L) are shown. The lines marked with a black filled-circle are the contigs listed in **Supplementary Table S2**, and they were used for constructions of the WVQ genomic sequences. Some unlisted contigs in **Supplementary Table S2** were also used for the construction of strain a sequence. The lines marked with an open circle were potentially derived from the variant(s) of WVQ strain a that have sequences slightly different from the reference strain a. The read depth coverage throughout the WVQ genomes (strains a and b) is presented in **Supplementary Figure S2C**.

derived from strain c were relatively lower compared to those of the other two strains in some of the tested wheat plants (**Figure 1**). Unfortunately, the entire genomic sequence of strain c could not be determined *via* combinational RT-PCR and direct sequencing. Some overlapping contigs (contigs 936 and 1832 or 935 in pool-19L), likely derived from strain c, show high nucleotide sequence identities >95.9% to each other (**Supplementary Table S2** and data not shown). Therefore, the mixed variant population of strain c likely existed in these pooled samples. This mixed population may have affected the direct sequencing analyses for the strain c. Further analyses are required to determine its entire genome.

Taxonomical and Phylogenetic Analyses of WVQ

BLAST-N analyses using the novel quinvirus sequences (strains a and b) as queries revealed that no significant genome-wide hits were detected (only for less than 12.0% for query coverage using other quinviruses; data not shown). PASC analysis of the genomic sequence of strain a showed 41.1% identity with that of a foveavirus – grapevine rupestris stem pitting-associated virus – which was the top hit sequence (**Supplementary Figure S4A**). Viruses among different genera in the *Quinvirinae* usually have less than about 45% nt identity in their genes (King et al., 2011), suggesting that WVQ represents a novel genus in the family. In the BLAST-P analyses, each predicted protein of the novel quinvirus (except for ORF6 in strain b) showed moderate levels of deduced amino-acid sequence identities (~38.1–47.5%) with the corresponding proteins encoded by other quinviruses (mainly foveaviruses infecting fruit trees; **Table 4**). PASCs using STD also showed similar results (**Figure 3** and **Supplementary Figures S4B**,**C**). Taken together, we propose that this newly discovered virus represents a virus species belonging to a novel genus in the family *Quinvirinae*.

To understand the phylogenetic relationships between WVQ and the known quinviruses, we constructed an ML tree using

the amino-acid alignment of their entire replicase sequences. The constructed tree showed that the WVQ strains form a separate clade within the subfamily *Quinvirinae*. WVQ is distantly related to members of the three genera *Foveavirus*, *Carlavirus*, and *Robigovirus*, and other floating or unassigned quinviruses, such as SCSMaV (proposed genus "Sustrivirus"), banana mild mosaic virus (proposed genus "Banmivirus"), garlic yellow mosaic-associated virus, and yam virus Y (Gambley and Thomas, 2001; Thompson and Randles, 2001; Da Silva et al., 2019; Silva et al., 2019; **Figure 4**). Similar trends were also observed in the NJ trees based on CP and TGB, in which WVQ was placed separately from the three established genera (see also Ryu and Song, 2021; **Supplementary Figures S5A**,**B**).

TABLE 3 | Comparison of predicted coding proteins between two WVQ strains.

Protein		Strain a	Strain b	Sequence identity[1]
Genome size		8,412 nt	8,411 nt	79.5%
ORF1	replicase	2069 aa (233.9 kDa)	2070 aa (233.9 kDa)	87.6%
ORF2	TGB1	223 aa (24.8 kDa)	223 aa (24.7 kDa)	86.6%
ORF3	TGB2	120 aa (13.4 kDa)	120 aa (13.4 kDa)	84.3%
ORF4	TGB3	85 aa (9.2 kDa)	85 aa (9.2 kDa)	68.6%
ORF5	CP	263 aa (28.6 kDa)	263 aa (28.6 kDa)	91.7%
ORF6	ORF6	–[2]	152 aa (17.3 kDa)[3]	–[2]

[1]*Using the NCBI BLAST-P-suite-2.*
[2]*–: Not presented or no hits.*
[3]*Some variants lack the ORF6 due to the insertion of A residue(s) at 8,351 nt (see text).*

Small RNA Profiles of WYMV and WVQ

Small RNA fractions of the wheat leaves (a pooled sample, KTH-18-1 and −2; **Table 1**) were deep sequenced to investigate the antiviral RNA silencing response. The small RNA reads (ranging 15–32 nt in length) mapped to the WYMV and WVQ were accounted for 0.5% (207,007 reads) and 0.2% (110,492 reads) of the total small RNAs, respectively. The small RNAs derived from both viruses were nearly equal positive and negative strands (45.9–59.6% for positive small RNA strands; **Supplementary Figure S6A**), predominantly 21 and 22 nt lengths for both strands (**Figure 5A** and **Supplementary Figure S6B**) and their 5'-terminal nucleotides were biased toward adenine (A) or uracil (U), which were particularly more prominent for 21 and 22 nt small RNAs in WVQ than those of WYMV (**Figure 5B**). Moreover, viral 21 and 22 nt small RNAs were distributed throughout the WYMV and WVQ genomes, with several distinct hotspots (**Supplementary Figure S6C**). Taken together, these analyses showed that WYMV- and WVQ-derived small RNAs possess the typical characteristics of viral small interfering RNAs (siRNAs; Li et al., 2017).

TABLE 4 | BLAST-P analyses using viral proteins of WVQ (strain a) as the queries.

Top hit virus	QC[1]	*E*-value	Identity	Accession
Query sequence: replicase				
Peach chlorotic mottle virus	85%	0.0	44.3%	AVD50414
Grapevine rupestris stem pitting-associated virus	84%	0.0	44.5%	QPB70005
Grapevine virus T	77%	0.0	47.9%	AYQ96168
Apple stem pitting virus	83%	0.0	42.7%	QKV49427
Panax ginseng flexivirus 1	75%		47.5%	YP_009552757
Query sequence: TGB1				
Rubus virus 1	99%	1e-53	44.0%	QLI58026
Apple stem pitting virus	97%	1e-49	43.8%	AGR66384
Peach chlorotic mottle virus	97%	6e-49	45.5%	YP_001497154
Apricot latent virus	97%	1e-48	44.2%	ADT91611
Apple green crinkle associated virus	97%	1e-48	43.3%	YP_006860590
Query sequence: TGB2				
Elderberry carlavirus A	87%	3e-26	47.2%	YP_009224930
Apple stem pitting virus	89%	1e-22	48.6%	AGR66384
Elderberry carlavirus B	89%	1e-22	44.4%	YP_009224936
Garlic common latent virus	89%	3e-21	46.9%	QED43133
Ilex cornuta carlavirus	89%	8e-21	45.0%	QJZ28443
Query sequence: TGB3				
No hits				
Query sequence: CP				
Peach chlorotic mottle virus	80%	3e-39	38.1%	AVD50418
Apple stem pitting virus	85%	2e-38	38.3%	AGR66254
Peach asteroid spot virus	71%	3e-38	40.1%	AAG48309
Garlic common latent virus	97%	1e-48	44.2%	QED43133
Ilex cornuta carlavirus	97%	1e-48	43.3%	QJZ28445

[1]*Query coverage (%).*

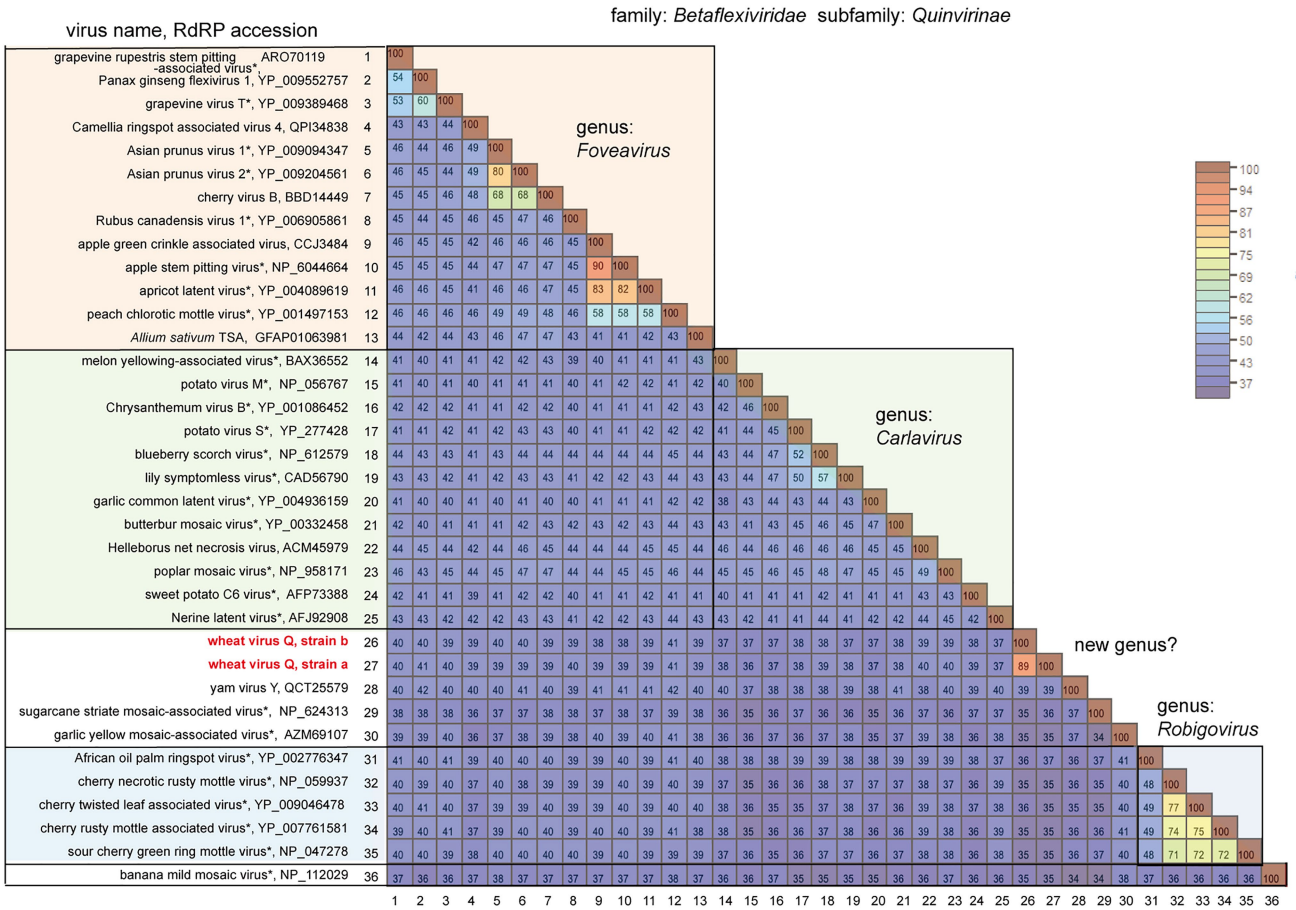


FIGURE 3 | Pairwise comparison of the replicase (RdRP) encoded by quinviruses or their candidates. The results of pairwise comparisons are shown as a heatmap with each pairwise amino acid identity calculated using SDT ver. 1.2. The virus names with asterisks are representative members of the viral species.

Transmission of WVQ

Our preliminary finding that WVQ was detected in WYMV-infected roots in the early spring (data not shown) when the temperature was relatively cool and thus the insects usually do not spread widely in the fields, suggests the possibility that WVQ may be transmitted to the roots through soil. Therefore, we aimed to examine the transmission of WVQ *via* soil under different conditions. Wheat seeds (cv. "Kitahonami") were sown in pots containing soils from WYMV-infested fields or commercial soils as a control. One group of pots (three pots for each treatment) being kept under growth-cabinet control at approximately 16°C, which is generally allowed for WYMV infection, while the other group of pots was kept in the greenhouse (non-controlled temperature), where WYMV infection is usually more difficult (Zhang et al., 2021). Under the growth cabinet condition, both WVQ and WYMV were detected by RT-PCR in the roots of wheat after 3 months of growth, but not after 1 month (**Figure 6A**). After 3 months of growth in the greenhouse, WVQ, but not WYMV, was detected in the roots of the plants (**Figure 6A**). No systemic infection or virus symptoms were observed in the aerial parts of the plants in either trial (data not shown). Thus, it was revealed that WVQ is potentially soil-transmissible, and this virus may require lower temperature and/or other unknown conditions for systemic infection but not for root infection unlike WYMV.

WVQ Infection in Rye Plants

WVQ strains were constantly detected by RT-PCR in the roots and leaves of the wheat plants (cv. "Kitahonami") that were grown in WYMV-infested fields at Naganuma over 4 consecutive years (2018–2021; **Table 1**; **Figures 1, 6B** and data not shown). To investigate the occurrence of WVQ in other crops, the rye samples were also collected from the same field in 2020 and 2021 and were subjected to RT-PCR. The results showed that WVQ strains (except for the strain c) were also detected in both leaves and roots of rye plants along with co-infecting WYMV (**Figure 6B**), indicating that WVQ is able to infect rye plants.

DISCUSSION

The occurrence of wheat yellow mosaic disease in East Asian countries has been predominantly associated with bymovirus

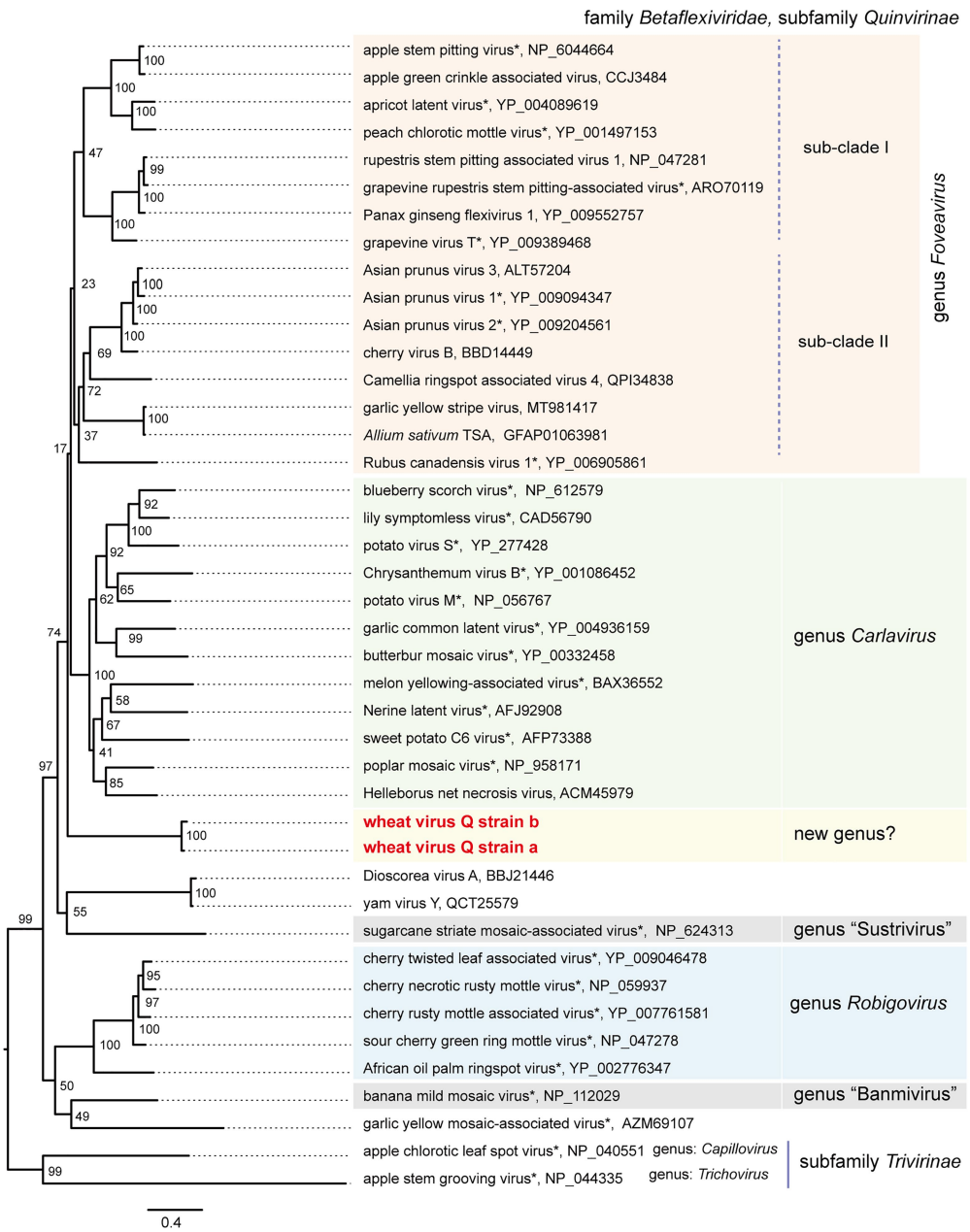


FIGURE 4 | Phylogenetic relationships of the quinviruses and other unassigned related viruses. A maximum likelihood phylogenetic tree was constructed using a MAFFT alignment of the full-length replicase amino acid sequences. Ambiguously aligned sequences were removed using Gblocks with the stringency levels lowered for all parameters. A model LG + I + G + F was selected as a best-fit model for the alignment (the 1,235 positions remaining in the input dataset). The tree with the highest log likelihood (–43,188,80) is shown. The tree was drawn using the midpoint rooting method. The virus names referring to plant viruses [genera *Foveavirus*, *Carlavirus* (selected), *Robigovirus*, and floating (proposed genera "Banmivirus" and "Sustrivirus") or unassigned members of subfamily *Quinvirinae*, family *Betaflexiviridae*] are followed by the GenBank accession or Ref-seq numbers of their sequences. Virus names with asterisks show representative members of the viral species. The two betaflexiviruses – apple chlorotic leaf spot virus (genus *Trichovirus*) and apple stem grooving virus (genus *Chapillovirus*) – of subfamily *Trivirinae* (family *Betaflexiviridae*) were used as the outgroups. The scale bar represents amino acid distances. The numbers at the nodes in the tree are bootstrap support values following 100 iterations.

infection, but it is possible that the magnitude of disease severity is also influenced by the viral communities in the wheat fields that sustain this disease. In this study, we investigated the wheat leaf virome associated with yellow mosaic disease in Hokkaido, Japan (**Table 1**). We identified three viruses – WYMV (a bymovirus), NCMV (a cytorhabdovirus), and a novel quinvirus (WVQ, at least three strains) – from two pooled samples obtained in 2 consecutive years (**Table 2**). In addition to the plant virus sequences, mycovirus-like contigs (at least 43) were also found in the two datasets (data not shown). Although the leaf meta-RNA-seq may

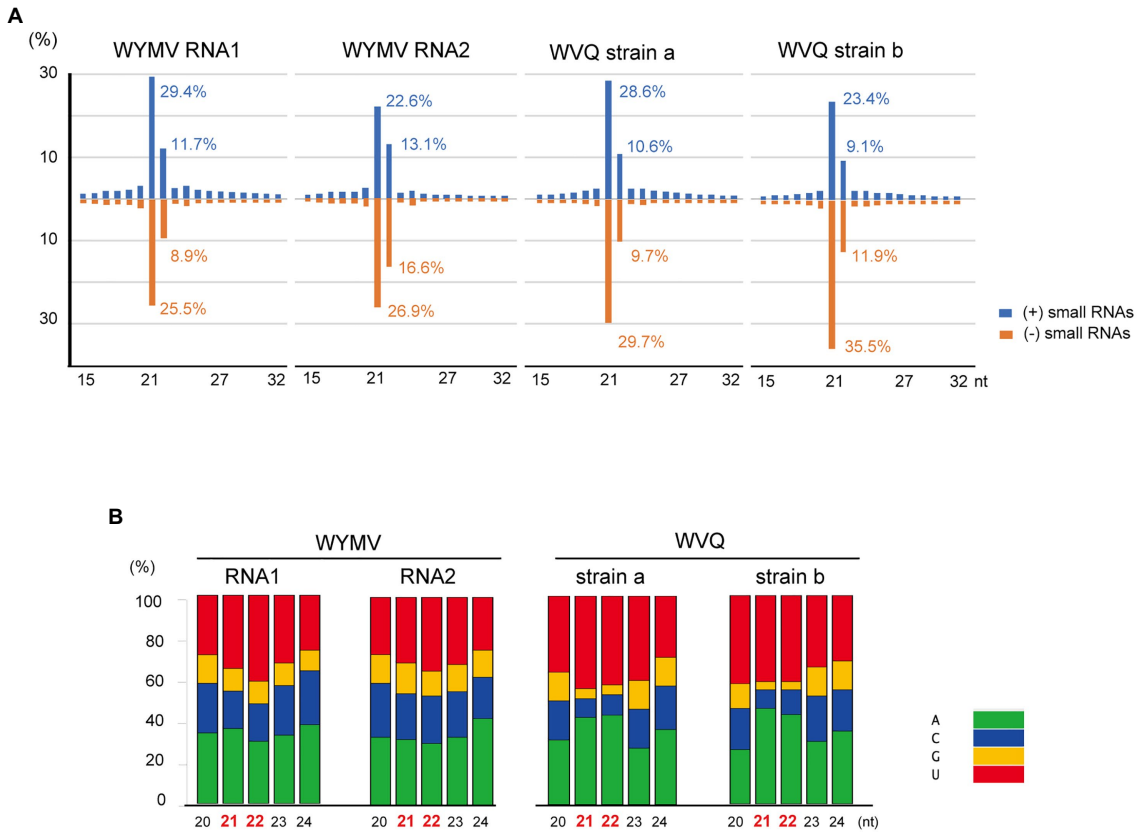


FIGURE 5 | Viral-derived small RNA profiles of WYMV and WVQ in the wheat leaves (the pooled KTH-18-1 and –2 samples). **(A)** Proportion of plus (+)- and minus (–)-strand small RNAs sizes, 15–32 nt, are shown. **(B)** Proportions of 5'-terminal nucleotide sequences of viral-derived small RNAs in the wheat leaves. The composite bar graphs represent the percentage of the 5'-terminal nucleotide sequences for each size-class (20–24 nt). Other profiles of the viral-derived small RNAs are presented in **Supplementary Figure S6**.

be able to provide information on the mycoviral communities in crop-associated fungal populations (Al Rwahnih et al., 2011; Marzano and Domier, 2016; Chiapello et al., 2020; Ma et al., 2021), many of the obtained contigs were only partial genomic sequences. Thus, further analyses are required to determine their entire genomes.

In this study, we have identified a novel quinvirus (family *Betaflexiviridae*), tentatively named "wheat virus Q." Among the three WVQ strains identified in the Naganuma field, two strains were determined on their complete genomic sequences (**Tables 2, 3**; **Figure 2** and **Supplementary Figure S2**). The sequence and phylogenetic analyses support the taxonomic status of WVQ as a distant evolutionary lineage (a member of the new genus) in the *Quinvirinae*, but the phylogenetic trees also show low statistical support for a deep node (for replicases) or formed two or three distinct clades (for CP and TGB1) of foveaviruses (**Table 4**; **Figures 3, 4**; **Supplementary Figures S4**, **S5**). Thus, more as-yet-unreported members of the foveaviruses and other quinviruses, including WVQ-related viruses, are needed in order to obtain more accurate phylogenetic relationships for these taxonomic studies. The three WVQ strains (a, b, and c) were simultaneously detected in most of the tested samples (**Figure 1** and **Table 1**). In addition, the RNA-seq data suggest the presence of WVQ variants belonging to each reference strains (at least strains a and c; **Figure 2**; **Supplementary Figure S2C** and **Supplementary Table S2**). Together, these indicate that WVQ infection in nature commonly forms a mixed population derived from multiple strains and variants. It is still unclear how these WVQ variants form mutant clouds (viral quasispecies) in the wheat plants (Domingo et al., 2012; Domingo and Perales, 2019) The biological significant of a mixed population for the WVQ infection is worth investigating in the future. The presence of WYMV and WVQ siRNAs in wheat leaves suggests that infection of both viruses induces antiviral RNA silencing responses in the field conditions (**Figure 5** and **Supplementary Figure S6**). Previous report showed that WYMV siRNAs have different characteristics between leaves and roots (Li et al., 2017). Because like WYMV, WVQ also potentially infects plant through roots, it is interesting to examine whether WVQ siRNAs also have differential characteristics between leaves and roots.

WVQ is the second quinvirus found to infect cereal crops following SCSMaV, which is distributed across a limited area in Queensland, Australia (Choi et al., 1999; Thompson and Randles, 2001). WVQ co-existed with WYMV in at least four

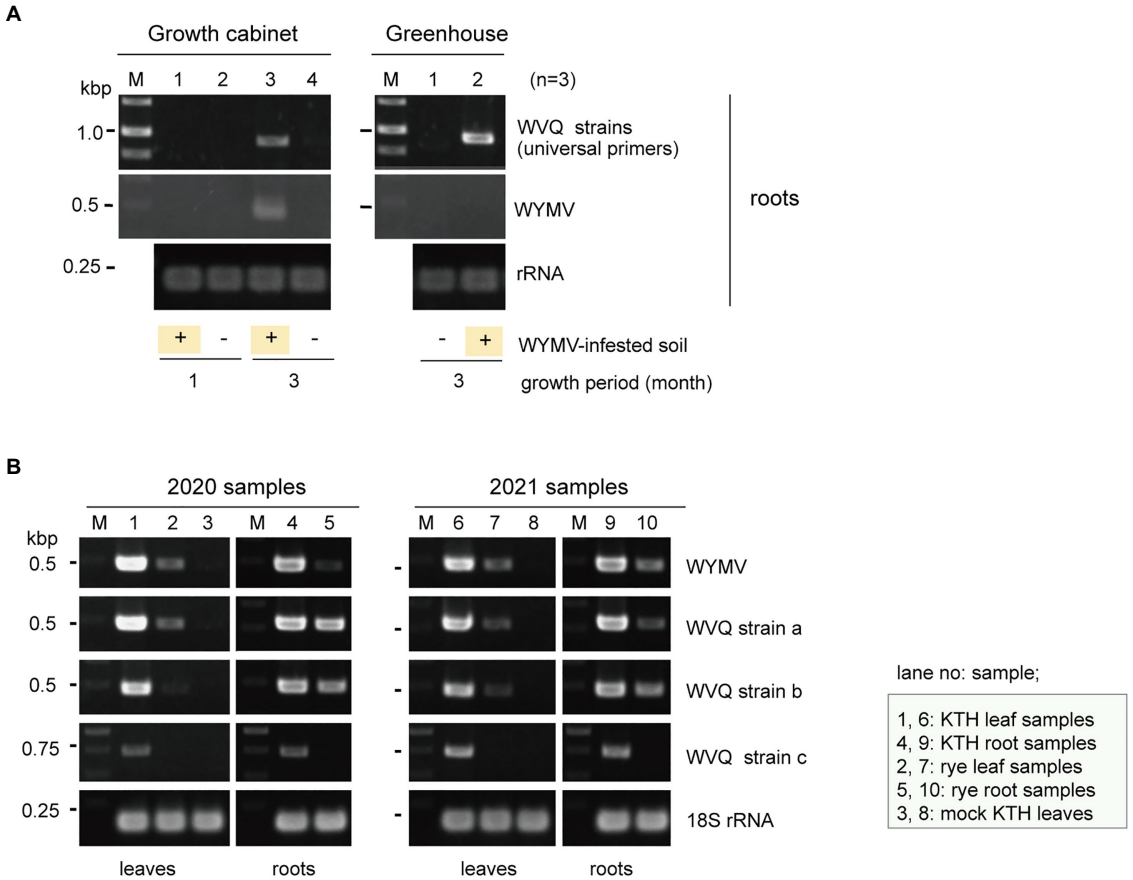


FIGURE 6 | (A) Transmission assay of WVQ. RT-PCR detection of viruses in the roots of wheat plants planted in WVQ and wheat yellow mosaic virus (WYMV)-infested soil. Wheat grains were sown in pots containing the field soil and grown under the growth-cabinet condition (below 16°C) or the greenhouse condition (non-controlled temperature). Three pots for each treatment were used. Total RNA fractions from the root samples were pooled according to the treatment and subjected to RT-PCR analysis with a common primer set for the WVQ strains. **(B)** RT-PCR detection of WVQ and WYMV in leaves or roots of wheat or rye plants. Sample names and their respective tissue origin are shown below the panel or in the bottom right. Lanes 3 and 8 are RNA samples of mock-inoculated plants. KTH, cv. Kitahonami; rye, cv. Fuyumidori. The wheat 18S rRNA gene was used as an endogenous RNA reference. The primers used in the RT-PCR analyses are listed in **Supplementary Table S1**. The PCR products were analyzed in 1.5% agarose gel electrophoresis. The gel was stained with ethidium bromide and viewed under UV light. Selected PCR products were sequenced to confirm the amplification of the virus targets.

different wheat-producing areas in the Hokkaido (NY, unpublished results), suggesting its widespread occurrence in this island. In addition, some unannotated short sequence fragments related to WVQ strains (mostly its 3'-terminal parts) were identified in the wheat EST libraries (cultivar "Chinese Spring"; **Supplementary Figure S7**), whose source plant materials were grown in Japan but outside of Hokkaido island (Kawaura et al., 2005; Mochida et al., 2006), suggesting the possibility that WVQ may also occurs in other islands in Japan. The natural infection by WVQ has also been found in the rye plants in the same Naganuma field (**Figure 6B**). WYMV is transmitted to wheat by the soil-borne vector *P. graminis* (Tamada and Kondo, 2013) and the field soil used for transmission experiment was infested with *P. graminis* (HK, unpublished results), while the underground mechanism, or possible biological vector, of WVQ transmission is still unknown (**Figure 6A**). Interestingly, as well as WYMV can be detected in nucleic acid fractions extracted from soil taken from an area in the Naganuma field where no wheat planting has taken place for the last 1 year (NY, unpublished results). This may suggests that WVQ is associated with unknown vector as for bymoviruses associated with *P. graminis* in the soil or present in a stable and/or transmissible form in the soil similarly to tobamoviruses (Ikegashira et al., 2004). It seems that SCSMaV is also soil transmissible *via* an unknown mechanism (Hughes, 1961; Thompson and Randles, 2001). Only natural vectors of some carlaviruses have been reported (Ryu and Song, 2021), while the vector of other members of subfamily *Quinvirinae* is still unknown. There is no particular sequence feature specifically present in WVQ and SCSMaV and absent in other members of the *Quinvirinae* (except for the cysteine-rich protein of carlaviruses).

WVQ was present in at least four consecutive growing seasons in the wheat plants coinfected with WYMV (**Table 1**; **Figures 1, 6B**), and were also detected in different wheat-producing areas in Hokkaido as mentioned above, revealing the common coexistence

of WVQ and WYMV in wheat fields in Hokkaido. This high incidence of co-infection by the two unrelated viruses is reminiscent of a soil-borne pathosystem, the lettuce big-vein disease (BVD). BVD has been associated with two unrelated negative-sense RNA viruses, Mirafiori lettuce big-vein virus (MLBVV, an ophiovirus in the family *Aspiviridae*) and lettuce big-vein associated virus (LBVaV, a varicosavirus in the family *Rhabdoviridae*), transmitted by a root-infecting fungus *Olpidium brassicae* (Kormelink et al., 2011; Maccarone, 2013). MLBVV is responsible to the induction of BVD symptoms, but LBVaV alone never shows symptoms (Lot et al., 2002; Roggero et al., 2003). The co-infection by these two viruses was frequently observed in the lettuce fields; thus, yet unknown interactions between two soil-borne viruses may play pivotal roles in this pathosystem. Therefore, the impact of WVQ infection on WYMV-induced yellow mosaic disease, particularly on symptom development and grain production, is worthy of detailed investigations. It will also be a future challenge to determine whether there is a trans enhancement in accumulation between the two wheat viruses.

DATA AVAILABILITY STATEMENT

The virus and virus-like sequences derived from this study have been submitted to the GenBank/DDBJ/ENA with the accession numbers LC632066–LC632071.

AUTHOR CONTRIBUTIONS

HK designed the experiments. NY and TT collected the samples. HK, NY, MF, KM, KH, HH, and NS performed the experimental work and discussion. HK analyzed the data and wrote the draft manuscript. lBA, NY, TT, and NS were involved in manuscript revision. All authors have given their approval to the final version of the manuscript.

FUNDING

This study was supported by grants from the Grants-in-Aids for Scientific Research (B; KAKENHI 20H02987) to HK, NS, KH, and HH, from the Scientific Research on Innovative Areas (16H06436, 16H06429, and 16H21723) to NS and HK from the Ministry of Education, Culture Sports, Science, and Technology (MEXT) of Japan and the Ohara Foundation for Agriculture Research, Kurashiki, Japan.

ACKNOWLEDGMENTS

We would like to thank Sakae Hisano, Hideki Nishimura, and Takakazu Matsuura for their helpful technical assistance. We also thank the handling editor and two reviewers for their assistance and valuable suggestions to improve the article. We are grateful to all the researchers who have shared their sequence data in publicly available databases prior to publication. During the proofing step of this paper, Fu et al. reported an isolate of WVQ strain b in winter wheat in China (Fu et al., 2021).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.715545/full#supplementary-material

REFERENCES

Adams, M. J., Lefkowitz, E. J., King, A. M. Q., Harrach, B., Harrison, R. L., Knowles, N. J., et al. (2016). Ratification vote on taxonomic proposals to the international committee on taxonomy of viruses (2016). *Arch. Virol.* 161, 2921–2949. doi: 10.1007/s00705-016-2977-6

Albrecht, T., White, S., Layton, M., Stenglein, M. D., Haley, S., and Nachappa, P. (2020). Ecology and epidemiology of wheat curl mite and mite-transmissible viruses in Colorado and insights into the wheat virome. bioRxiv [Preprint]. doi: 10.1101/2020.08.10.244806

Al Rwahnih, M., Daubert, S., Urbez-Torres, J. R., Cordero, F., and Rowhani, A. (2011). Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. *Arch. Virol.* 156, 397–403. doi: 10.1007/s00705-010-0869-8

Bao, Y. M., Chetvernin, V., and Tatusova, T. (2014). Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. *Arch. Virol.* 159, 3293–3304. doi: 10.1007/s00705-014-2197-x

Cadle-Davidson, L., and Bergstrom, G. C. (2004). The effects of postplanting environment on the incidence of soilborne viral diseases in winter cereals. *Phytopathology* 94, 527–534. doi: 10.1094/PHYTO.2004.94.5.527

Caglayan, K., Roumi, V., Gazel, M., Elci, E., Acioglu, M., Plesko, I. M., et al. (2019). Identification and characterization of a novel robigovirus species from sweet cherry in Turkey. *Pathogens* 8:57. doi: 10.3390/pathogens8020057

Candresse, T., Cao, M., Constable, F., Blouin, A., Cho, W. K., Nagata, T., et al. (2021). Create two new genera and 23 new species (*Tymovirales*: *Betaflexiviridae*). The taxonomy proposal submitted to the ICTV. Available at: https://talk.ictvonline.org/files/proposals/taxonomy_proposals_plant1/m/plant04/13138 (Accessed August 6, 2021).

Chen, J., Shi, N. N., Cheng, Y., Diao, A., Chen, J. P., Wilson, T. M. A., et al. (1999). Molecular analysis of barley yellow mosaic virus isolates from China. *Virus Res.* 64, 13–21. doi: 10.1016/S0168-1702(99)00076-3

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

Choi, Y. G., Croft, B. J., and Randles, J. W. (1999). Identification of sugarcane striate mosaic-associated virus by partial characterization of its double-stranded RNA. *Phytopathology* 89, 877–883. doi: 10.1094/PHYTO.1999.89.10.877

Clover, G., and Henry, C. (1999). Detection and discrimination of wheat spindle streak mosaic virus and wheat yellow mosaic virus using multiplex RT-PCR. *Eur. J. Plant Pathol.* 105, 891–896. doi: 10.1023/A:1008707331487

Da Silva, L. A., Oliveira, A. S., Melo, F. L., Ardisson-Araujo, D. M. P., Resende, F. V., Resende, R. O., et al. (2019). A new virus found in garlic virus complex is a member of possible novel genus of the family *Betaflexiviridae* (order *Tymovirales*). *PeerJ* 7:e6285. doi: 10.7717/peerj.6285

Dietzgen, R. G., Kondo, H., Goodin, M. M., Kurath, G., and Vasilakis, N. (2017). The family *Rhabdoviridae*: mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Res.* 227, 158–170. doi: 10.1016/j.virusres.2016.10.010

Domingo, E., and Perales, C. (2019). Viral quasispecies. *PLoS Genet.* 15:e1008271. doi: 10.1371/journal.pgen.1008271

Domingo, E., Sheldon, J., and Perales, C. (2012). Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev.* 76, 159–216. doi: 10.1128/MMBR.05023-11

Fu, S., Zhang, T., He, M., Sun, B., Zhou, X., and Wu, J. (2021). Molecular characterization of a novel wheat-infecting virus in family *Betaflexiviridae*. *Arch. Virol.* doi: 10.1007/s00705-021-05175-y (in press).

Gambley, C. F., and Thomas, J. E. (2001). Molecular characterisation of Banana mild mosaic virus, a new filamentous virus in *Musa* spp. *Arch. Virol.* 146, 1369–1379. doi: 10.1007/s007050170097

Gazel, M., Roumi, V., Ordek, K., Maclot, F., Massart, S., and Caglayan, K. (2020). Identification and molecular characterization of a novel foveavirus from *Rubus* spp. in Turkey. *Virus Res.* 286:198078. doi: 10.1016/j.virusres.2020.198078

Golyaev, V., Candresse, T., Rabenstein, F., and Pooggin, M. M. (2019). Plant virome reconstruction and antiviral RNAi characterization by deep sequencing of small RNAs from dried leaves. *Sci. Rep.* 9:19268. doi: 10.1038/s41598-019-55547-3

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Han, C., Li, D., Xing, Y., Zhu, K., Tian, Z., Cai, Z., et al. (2000). Wheat yellow mosaic virus widely occurring in wheat (*Triticum aestivum*) in China. *Plant Dis.* 84, 627–630. doi: 10.1094/PDIS.2000.84.6.627

Hirzmann, J., Luo, D., Hahnen, J., and Hobom, G. (1993). Determination of messenger RNA 5'-ends by reverse transcription of the cap structure. *Nucleic Acids Res.* 21, 3597–3598. doi: 10.1093/nar/21.15.3597

Hodge, B. A., Paul, P. A., and Stewart, L. R. (2020). Occurrence and high-throughput sequencing of viruses in Ohio wheat. *Plant Dis.* 104, 1789–1800. doi: 10.1094/PDIS-08-19-1724-RE

Horie, M., Akashi, H., Kawata, M., and Tomonaga, K. (2021). Identification of a reptile lyssavirus in Anolis allogus provided novel insights into lyssavirus evolution. *Virus Genes* 57, 40–49. doi: 10.1007/s11262-020-01803-y

Hughes, C. G. (1961). Striate mosaic, a new disease of sugarcane. *Nature* 190, 366–367. doi: 10.1038/190366b0

Ikegashira, Y., Ohki, T., Ichiki, U. T., Higashi, T., Hagiwara, K., Omura, T., et al. (2004). An immunological system for the detection of *pepper mild mottle virus* in soil from green pepper fields. *Plant Dis.* 88, 650–656. doi: 10.1094/PDIS.2004.88.6.650

Inouye, T. (1969). Viral pathogen of the wheat yellow mosaic disease. *Nogaku Kenkyu* 53, 61–68.

Jarošová, J., and Kundu, J. K. (2010). Validation of reference genes as internal control for studying viral infections in cereals by quantitative real-time RT-PCR. *BMC Plant Biol.* 10:146. doi: 10.1186/1471-2229-10-146

Jiang, C. C., Kan, J. H., Ordon, F., Perovic, D., and Yang, P. (2020). Bymovirus-induced yellow mosaic diseases in barley and wheat: viruses, genetic resistances and functional aspects. *Theor. Appl. Genet.* 133, 1623–1640. doi: 10.1007/s00122-020-03555-7

Jo, Y., Bae, J. Y., Kim, S. M., Choi, H., Lee, B. C., and Cho, W. K. (2018). Barley RNA viromes in six different geographical regions in Korea. *Sci. Rep.* 8:13237. doi: 10.1038/s41598-018-31671-4

Kanyuka, K., Ward, E., and Adams, M. J. (2003). *Polymyxa graminis* and the cereal viruses it transmits: a research challenge. *Mol. Plant Pathol.* 4, 393–406. doi: 10.1046/j.1364-3703.2003.00177.x

Kashiwazaki, S., Ogawa, K., Usugi, T., Omura, T., and Tsuchizaki, T. (1989). Characterization of several strains of barley yellow mosaic virus. *Ann. Phytopath. Soc. Japan* 55, 16–25.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kawaura, K., Mochida, K., and Ogihara, Y. (2005). Expression profile of two storage-protein gene families in hexaploid wheat revealed by large-scale analysis of expressed sequence tags. *Plant Physiol.* 139, 1870–1880. doi: 10.1104/pp.105.070722

King, A. M. Q., Adams, M. J., Carstens, E. B., and Lefkowitz, E. J. (2011). *Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses*. London: San Diego Elsevier Academic Press.

Kobayashi, F., Kojima, H., Tanaka, T., Saito, M., Kiribuchi-Otobe, C., and Nakamura, T. (2020). Characterization of the *Q. Ymym* region on wheat chromosome 2D associated with wheat yellow mosaic virus resistance. *Plant Breed.* 139, 93–106. doi: 10.1111/pbr.12759

Kojima, H., Nishio, Z., Kobayashi, F., Saito, M., Sasaya, T., Kiribuchi-Otobe, C., et al. (2015). Identification and validation of a quantitative trait locus associated with wheat yellow mosaic virus pathotype I resistance in a Japanese wheat variety. *Plant Breed.* 134, 373–378. doi: 10.1111/pbr.12279

Kondo, H., Chiba, S., Maruyama, K., Andika, I. B., and Suzuki, N. (2019). A novel insect-infecting virga/nege-like virus group and its pervasive endogenization into insect genomes. *Virus Res.* 262, 37–47. doi: 10.1016/j.virusres.2017.11.020

Kondo, H., Fujita, M., Hisano, H., Hyodo, K., Andika, I. B., and Suzuki, N. (2020). Virome analysis of aphid populations That infest the barley field: The discovery of two novel groups of nege/kita-like viruses and other novel RNA viruses. *Front. Microbiol.* 11:509. doi: 10.3389/fmicb.2020.00509

Kondo, H., Hirano, S., Chiba, S., Andika, I. B., Hirai, M., Maeda, T., et al. (2013). Characterization of burdock mottle virus, a novel member of the genus *Benyvirus*, and the identification of benyvirus-related sequences in the plant and insect genomes. *Virus Res.* 177, 75–86. doi: 10.1016/j.virusres.2013.07.015

Kondo, H., Hisano, S., Chiba, S., Maruyama, K., Andika, I. B., Toyoda, K., et al. (2016). Sequence and phylogenetic analyses of novel totivirus-like double-stranded RNAs from field-collected powdery mildew fungi. *Virus Res.* 213, 353–364. doi: 10.1016/j.virusres.2015.11.015

Kondo, H., Maeda, T., Shirako, Y., and Tamada, T. (2006). Orchid fleck virus is a rhabdovirus with an unusual bipartite genome. *J. Gen. Virol.* 87, 2413–2421. doi: 10.1099/vir.0.81811-0

Kondo, H., Maruyama, K., Chiba, S., Andika, I. B., and Suzuki, N. (2014). Transcriptional mapping of the messenger and leader RNAs of orchid fleck virus, a bisegmented negative-strand RNA virus. *Virology* 452, 166–174. doi: 10.1016/j.virol.2014.01.007

Kormelink, R., Garcia, M. L., Goodin, M., Sasaya, T., and Haenni, A. L. (2011). Negative-strand RNA viruses: the plant-infecting counterparts. *Virus Res.* 162, 184–202. doi: 10.1016/j.virusres.2011.09.028

Kuhne, T. (2009). Soil-borne viruses affecting cereals-known for long but still a threat. *Virus Res.* 141, 174–183. doi: 10.1016/j.virusres.2008.05.019

Lapierre, H. D., and Hariri, D. (2008). "Cereal viruses: wheat and barley," in *Encyclopedia of Virology*. *3rd Edn.* eds. B. W. J. Mahy and M. H. V. Regenmortel (Amsterdam: Elsevier), 490–497.

Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Li, L. Y., Andika, I. B., Xu, Y., Zhang, Y., Xin, X. Q., Hu, L. F., et al. (2017). Differential characteristics of viral siRNAs between leaves and roots of wheat plants naturally infected with wheat yellow mosaic virus, a soil-borne virus. *Front. Microbiol.* 8:1802. doi: 10.3389/fmicb.2017.01802

Lin, Y. H., Fujita, M., Chiba, S., Hyodo, K., Andika, I. B., Suzuki, N., et al. (2019). Two novel fungal negative-strand RNA viruses related to mymonaviruses and phenuiviruses in the shiitake mushroom (*Lentinula edodes*). *Virology* 533, 125–136. doi: 10.1016/j.virol.2019.05.008

Liu, Y., Du, Z. Z., Wang, H., Zhang, S., Cao, M. J., and Wang, X. F. (2018). Identification and characterization of wheat yellow striate virus, a novel leafhopper-transmitted nucleorhabdovirus infecting wheat. *Front. Microbiol.* 9:468. doi: 10.3389/fmicb.2018.00468

Lot, H., Campbell, R. N., Souche, S., Milne, R. G., and Roggero, P. (2002). Transmission by *Olpidium brassicae* of Mirafiori lettuce virus and lettuce big-vein virus, and their roles in lettuce big-vein etiology. *Phytopathology* 92, 288–293. doi: 10.1094/PHYTO.2002.92.3.288

Luo, Q., Hu, S. Z., Lin, Q., Xu, F., Peng, J. J., Zheng, H. Y., et al. (2021). Complete genome sequence of a novel foveavirus isolated from *Allium sativum* L. in China. *Arch. Virol.* 166, 983–986. doi: 10.1007/s00705-021-04957-8

Ma, Y., Fort, T., Marais, A., Lefebvre, M., Theil, S., Vacher, C., et al. (2021). Leaf-associated fungal and viral communities of wild plant populations differ between cultivated and natural ecosystems. *Plant Env. Interact.* 2, 87–99. doi: 10.1002/pei3.10043

Maccarone, L. D. (2013). Relationships between the pathogen *Olpidium virulentus* and viruses associated with lettuce big-vein disease. *Plant Dis.* 97, 700–707. doi: 10.1094/PDIS-10-12-0979-FE

Marzano, S. Y. L., and Domier, L. L. (2016). Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Res.* 213, 332–342. doi: 10.1016/j.virusres.2015.11.002

Mochida, K., Kawaura, K., Shimosaka, E., Kawakami, N., Shin-I, T., Kohara, Y., et al. (2006). Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat. *Mol. Gen. Genomics.* 276, 304–312. doi: 10.1007/s00438-006-0120-1

Moore, C., and Meng, B. Z. (2019). Prediction of the molecular boundary and functionality of novel viral AlkB domains using homology modelling and principal component analysis. *J. Gen. Virol.* 100, 691–703. doi: 10.1099/jgv.0.001237

Muhire, B. M., Varsani, A., and Martin, D. P. (2014). SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9:e108277. doi: 10.1371/journal.pone.0108277

Namba, S., Kashiwazaki, S., Lu, X., Tamura, M., and Tsuchizaki, T. (1998). Complete nucleotide sequence of wheat yellow mosaic bymovirus genomic RNAs. *Arch. Virol.* 143, 631–643. doi: 10.1007/s007050050319

Nishigawa, H., Hagiwara, T., Yumoto, M., Sotome, T., Kato, T., and Natsuaki, T. (2008). Molecular phylogenetic analysis of barley yellow mosaic virus. *Arch. Virol.* 153, 1783–1786. doi: 10.1007/s00705-008-0163-1

Ohki, T., Netsu, O., Kojima, H., Sakai, J., Onuki, M., Maoka, T., et al. (2014). Biological and genetic diversity of wheat yellow mosaic virus (genus *Bymovirus*). *Phytopathology* 104, 313–319. doi: 10.1094/PHYTO-06-13-0150-R

Ohto, Y., Hatta, K., and Ishiguro, K. (2006). Differential wheat cultivars to discriminate pathogenicity of Japanese wheat yellow mosaic virus (WYMV) isolates. *Jpn. J. Phytopathol.* 72, 93–100. doi: 10.3186/jjphytopath.72.93

Ordon, F., Habekuss, A., Kastirr, U., Rabenstein, F., and Kuhne, T. (2009). Virus resistance in cereals: sources of resistance, genetics and breeding. *J. Phytopathol.* 157, 535–545. doi: 10.1111/j.1439-0434.2009.01540.x

Park, D., Zhang, M., and Hahn, Y. (2019). Novel *Foveavirus* (the family *Betaflexiviridae*) species identified in ginseng (*Panax ginseng*). *Acta Virol.* 63, 155–161. doi: 10.4149/av_2019_204

Reynard, J. S., Brodard, J., Remoliff, E., Lefebvre, M., Schumpp, O., and Candresse, T. (2020). A novel foveavirus identified in wild grapevine (*Vitis vinifera* subsp. *sylvestris*). *Arch. Virol.* 165, 2999–3002. doi: 10.1007/s00705-020-04817-x

Roggero, P., Lot, H., Souche, S., Lenzi, R., and Milne, R. G. (2003). Occurrence of Mirafiori lettuce virus and lettuce big-vein virus in relation to development of big-vein symptoms in lettuce crops. *Eur. J. Plant Pathol.* 109, 261–267. doi: 10.1023/A:1023060830841

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

Ryu, K. H., and Song, E. G. (2021). "Quinviruses (*Betaflexiviridae*)," in *Encyclopedia of Virology. Vol. 3.* eds. B. W. J. Mahy and M. H. V. V. Regenmortel (London: Elsevier Academic Press, San Diego), 642–652.

Saitou, N., and Nei, M. (1987). The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Seguin, J., Otten, P., Baerlocher, L., Farinelli, L., and Pooggin, M. M. (2016). MISIS-2: A bioinformatics tool for in-depth analysis of small RNAs and representation of consensus master genome in viral quasispecies. *J. Virol. Methods* 233, 37–40. doi: 10.1016/j.jviromet.2016.03.005

Shahi, S., Eusebio-Cope, A., Kondo, H., Hillman, B. I., and Suzuki, N. (2019). Investigation of host range of and host defense against a mitochondrially replicating mitovirus. *J. Virol.* 93, e01503–e01518. doi: 10.1128/JVI.01503-18

Silva, G., Bomer, M., Rathnayake, A. I., Sewe, S. O., Visendi, P., Oyekanmi, J. O., et al. (2019). Molecular characterization of a new virus species identified in yam (*Dioscorea* spp.) by high-throughput sequencing. *Plants* 8:167. doi: 10.3390/plants8060167

Singh, K., Jarosova, J., Fousek, J., Huan, C., and Kundu, J. K. (2020). Virome identification in wheat in the Czech Republic using small RNA deep sequencing. *J. Integr. Agric.* 19, 1825–1833. doi: 10.1016/S2095-3119(19)62805-4

Somera, M., Kvarnheden, A., Desbiez, C., Blystad, D. R., Soovali, P., Kundu, J. K., et al. (2020). Sixty years after the first description: genome sequence and biological characterization of european wheat striate mosaic virus infecting cereal crops. *Phytopathology* 110, 68–79. doi: 10.1094/PHYTO-07-19-0258-FI

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Tamada, T., and Kondo, H. (2013). Biological and genetic diversity of plasmodiophorid-transmitted viruses and their vectors. *J. Gen. Plant Pathol.* 79, 307–320. doi: 10.1007/s10327-013-0457-3

Tanno, F., Nakatsu, A., Toriyama, S., and Kojima, M. (2000). Complete nucleotide sequence of northern cereal mosaic virus and its genome organization. *Arch. Virol.* 145, 1373–1384. doi: 10.1007/s007050070096

Thompson, N., and Randles, J. W. (2001). The genome organisation and taxonomy of sugarcane striate mosaic associated virus. *Arch. Virol.* 146, 1441–1451. doi: 10.1007/s007050170070

Valente, J. B., Pereira, F. S., Stempkowski, L. A., Farias, M., Kuhnem, P., Lau, D., et al. (2019). A novel putative member of the family *Benyviridae* is associated with soilborne wheat mosaic disease in Brazil. *Plant Pathol.* 68, 588–600. doi: 10.1111/ppa.12970

Wylie, S. J., Adams, M., Chalam, C., Kreuze, J., Lopez-Moya, J. J., Ohshima, K., et al. (2017). ICTV Virus Taxonomy Profile: *Potyviridae*. *J. Gen. Virol.* 98, 352–354. doi: 10.1099/jgv.0.000740

Yaegashi, H., Oyamada, S., Goto, S., Yamagishi, N., Isogai, M., Ito, T., et al. (2020). Simultaneous infection of sweet cherry with eight virus species including a new foveavirus. *J. Gen. Plant Pathol.* 86, 134–142. doi: 10.1007/s10327-019-00896-0

Yoshikawa, N., and Yaegashi, H. (2021). "Betaflexiviruses (*Betaflexiviridae*)," in *Encyclopedia of Virology*. *Vol. 3*. eds. B. W. J. Mahy and M. H. V. V. Regenmortel (London: Elsevier Academic Press, San Diego), 229–238.

Zhang, Z. Y., Liu, X. J., Li, D. W., Yu, J. L., and Han, C. G. (2011). Rapid detection of wheat yellow mosaic virus by reverse transcription loop-mediated isothermal amplification. *Virol. J.* 8:550. doi: 10.1186/1743-422X-8-550

Zhang, F., Liu, S., Zhang, T., Ye, Z., Han, X., Zhong, K., et al. (2021). Construction and biological characterization of an infectious full-length cDNA clone of a Chinese isolate of wheat yellow mosaic virus. *Virology* 556, 101–109. doi: 10.1016/j.virol.2021.01.018

Zheng, L. P., Chen, M., and Li, R. H. (2020). Camellia ringspot-associated virus 4, a proposed new foveavirus from *Camellia japonica*. *Arch. Virol.* 165, 1707–1710. doi: 10.1007/s00705-020-04655-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fmicb.2021.633710

Viruses Without Borders: Global Analysis of the Population Structure, Haplotype Distribution, and Evolutionary Pattern of Iris Yellow Spot Orthotospovirus (Family *Tospoviridae*, Genus *Orthotospovirus*)

*Afsha Tabassum[1], S. V. Ramesh[2], Ying Zhai[1], Romana Iftikhar[1], Cristian Olaya[1] and Hanu R. Pappu[1]**

[1] Department of Plant Pathology, Washington State University, Pullman, WA, United States, [2] Indian Council of Agricultural Research-Central Plantation Crops Research Institute, Kasaragod, India

OPEN ACCESS

Edited by:
Akhtar Ali,
University of Tulsa, United States

Reviewed by:
Steve Wylie,
Murdoch University, Australia
Adane Abraham,
Botswana International University of Science and Technology, Botswana

***Correspondence:**
Hanu R. Pappu
hrp@wsu.edu

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *26 November 2020*
Accepted: *24 June 2021*
Published: *20 September 2021*

Citation:
Tabassum A, Ramesh SV, Zhai Y, Iftikhar R, Olaya C and Pappu HR (2021) Viruses Without Borders: Global Analysis of the Population Structure, Haplotype Distribution, and Evolutionary Pattern of Iris Yellow Spot Orthotospovirus (Family Tospoviridae, Genus Orthotospovirus). Front. Microbiol. 12:633710. doi: 10.3389/fmicb.2021.633710

Iris yellow spot, caused by Iris yellow spot orthotospovirus (IYSV) (Genus: *Orthotospovirus*, Family: *Tospoviridae*), is an important disease of *Allium* spp. The complete N gene sequences of 142 IYSV isolates of curated sequence data from GenBank were used to determine the genetic diversity and evolutionary pattern. *In silico* restriction fragment length polymorphism (RFLP) analysis, codon-based maximum likelihood studies, genetic differentiation and gene flow within the populations of IYSV genotypes were investigated. Bayesian phylogenetic analysis was carried out to estimate the evolutionary rate. *In silico* RFLP analysis of N gene sequences categorized IYSV isolates into two major genotypes *viz*., IYSV Netherlands ($IYSV_{NL}$; 55.63%), IYSV Brazil ($IYSV_{BR}$; 38.73%) and the rest fell in neither group [IYSV other ($IYSV_{other}$; 5.63%)]. Phylogenetic tree largely corroborated the results of RFLP analysis and the IYSV genotypes clustered into $IYSV_{NL}$ and $IYSV_{BR}$ genotypes. Genetic diversity test revealed $IYSV_{other}$ to be more diverse than $IYSV_{NL}$ and $IYSV_{BR}$. $IYSV_{NL}$ and $IYSV_{BR}$ genotypes are under purifying selection and population expansion, whereas $IYSV_{other}$ showed decreasing population size and hence appear to be under balancing selection. $IYSV_{BR}$ is least differentiated from $IYSV_{other}$ compared to $IYSV_{NL}$ genotype based on nucleotide diversity. Three putative recombinant events were found in the N gene of IYSV isolates based on RDP analysis, however, RAT substantiated two among them. The marginal likelihood mean substitution rate was 5.08×10^{-5} subs/site/year and 95% highest posterior density (HPD) substitution rate between 5.11×10^{-5} and 5.06×10^{-5}. Findings suggest that IYSV continues to evolve using population expansion strategies. The substitution rates identified are similar to other plant RNA viruses.

Keywords: BEAST, evolutionary genomics, gene flow, genetic differentiation, genetic recombination, iris yellow spot orthotospovirus, *in silico* RFLP, phylogenetics

Abbreviations: IYSV, Iris yellow spot orthotospovirus; N gene, Nucleocapsid gene; RFLP, Restriction Fragment Length Polymorphism; RDP, Recombination Detection Program; RAT, Recombination Analysis Tool; BEAST, Bayesian Evolutionary Analysis by Sampling Trees.

INTRODUCTION

Tospoviruses continue to be a major production constraint for a wide range of agronomic and horticultural crops worldwide (Gent et al., 2006; Pappu et al., 2009; Mandal et al., 2012; Mandal et al., 2012; Bag et al., 2015; Oliver and Whitfield, 2016; Turina et al., 2016; Resende et al., 2020). Iris yellow spot orthotospovirus (IYSV; genus: *Orthotospovirus*, family: *Tospoviridae*) (Resende et al., 2020) primarily infect *Allium* spp., which includes onion (*Allium cepa*), green onion (*Allium fistulosum)*, garlic (*Allium tuberosum*), leek (*Allium porrum*) (Gent et al., 2006; Cordoba-Selles et al., 2007; Bag et al., 2015; Karavina et al., 2016; Tabassum et al., 2016. The virus was first described in southern Brazil in 1981 on infected onion (*Allium cepa*; family: Amaryllidaceae) inflorescence stalks (scapes). The disease was referred to as "Sapeca." In the US, the disease was first described in the Treasure Valley of southwestern Idaho and southeastern Oregon in 1989 (Gent et al., 2006). In 2003, the disease epidemic in Colorado (United States) caused a crop loss estimated at US $ 2.5–5 million (Gent et al., 2006). The disease has spread to most of the onion-growing areas in Africa, Asia, Europe, the Americas, and the Oceania (Centre for Agriculture and Bioscience International (CABI) - Invasive Species Compendium, 2019).

The disease caused by IYSV is characterized by chlorotic or necrotic, straw-colored to white, dry, elongated or spindle shaped lesions along the scape (**Figure 1A**). Lesions are frequently at middle to lower portions of the scape. The diamond-shaped lesions tend to be less defined on leaves (Pappu et al., 2008; Bag et al., 2015). The photosynthetic activity is affected in the infected plants leading to reduced bulb size. As the disease progresses, the lesions girdle the scape causing the seed head to collapse leading to severe crop losses (**Figure 1B**; Gent et al., 2006).

IYSV, as other tospoviruses, consists of a tripartite genome: Small (S) and Medium (M) RNAs encode proteins in both sense and antisense orientations (ambisense) while the Large (L) RNA encodes protein from negative sense strand. The L RNA codes for RNA dependent RNA polymerase (RdRp), M RNA codes for glycoprotein precursors (G_N and G_C) and the non-structural movement protein (NSm), and the S RNA codes for nucleocapsid (N) and non-structural silencing suppressor protein (NSs) (reviewed in Bag et al., 2015; Pappu et al., 2020; Resende et al., 2020).

Genetic evolution of viruses directly impacts the host-virus interactions and as such is important to ascertain genetic diversity within a viral species (Sacristan and Garcia-Arenal, 2008; Gibbs and Ohsima, 2010). Genetic drift, migration, mutation, natural selection, segment reassortment, and recombination are the major sources of evolutionary changes in the genetic architecture of viral populations (Moya et al., 2004; Butkovic et al., 2021). Phylo-geographical analysis is a powerful tool to determine the geographical distribution pattern of virus, assessing their genetic variation, and historical events that are shaping the genetic architecture of the viral populations (Hewitt, 2004; Chen et al., 2012). Comprehensive genetic architecture and evolutionary genomic analysis of viral populations have become a subject of increasing attention in a number of viruses.

While IYSV is widely distributed in the world, the complete genome of only a few isolates are sequenced. Since the N gene is considered as one of the descriptors for tospovirus identification and classification, the N gene of a large number of isolates was sequenced and the genetic diversity was determined (de Avila et al., 1993; Pappu et al., 2006; Nischwitz et al., 2007; Iftikhar et al., 2014; Bag et al., 2015). The number of N gene sequences of IYSV isolates reported since the last study (Iftikhar et al., 2014) has been on the rise. Building on the earlier findings, we carried out a more detailed and a global analysis of the extent of genetic recombination, genetic diversity, genetic differentiation, and gene flow among different genotypes of IYSV isolates reported from different parts of the world. Further, Bayesian model-based coalescent approaches were used to gain insights into the molecular evolutionary pattern of IYSV population.

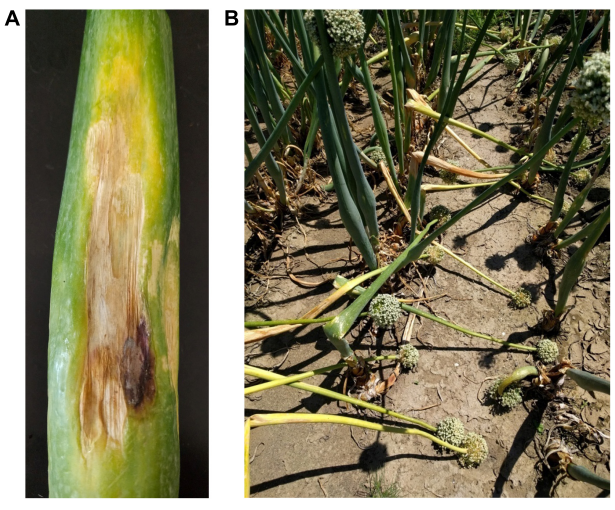


FIGURE 1 | Symptoms associated with Iris yellow spot orthotospovirus infection on onion **(A)** on scapes; **(B)** toppling of the seed head.

MATERIALS AND METHODS

Data Source of Nucleocapsid (N) Gene Sequences

Complete nucleocapsid (N) gene sequences of 142 IYSV isolates reported from across the globe were obtained the nucleotide sequence repository, GenBank. IYSV isolates analyzed were from 19 countries spread over six continents—Africa, Asia, Australia and New Zealand, Europe, North America (Canada, Mexico and the United States) and South America, infecting 10 different hosts including *Allium cepa* (the most commonly reported host), *Allium porrum, Eustoma russellianum, Allium tuberosum, Allium chinense,* Wild onion, *Alstroemeria* sp., *Allium sativum,* and *Allium fistulosum*. The N gene sequence (HQ267713) derived from tomato spotted wilt orthotospovirus (TSWV) infecting pepper crop in South Korea was used as an outgroup (**Supplementary Table 1**). Only complete IYSV N gene sequences

(822 nt-long open reading frame coding for a 273-amino acid protein) were considered for analysis.

In silico Restriction Fragment Length Polymorphism (RFLP) Analysis

N gene sequences were analyzed for sequence variations by performing *in silico* RFLP analysis using Restriction Mapper (Restriction Mapper, 2009). The complete N gene sequence was virtually digested, and sites were mapped as recognized by restriction enzyme *Hin*fI (Zen et al., 2005). Based on *Hin*fI digestion, IYSV isolates can be grouped into IYSV Netherlands ($IYSV_{NL}$) or IYSV Brazil ($IYSV_{BR}$) types. The size of the largest fragment generated by digestion is considered for differentiating the given isolate into two groups. The genotypes *viz.,* $IYSV_{NL}$ and $IYSV_{BR}$ are differentiated based on the resultant 308 and 468 bp fragments, respectively. Those isolates that yielded any different fragment size upon restriction digestion were grouped into $IYSV_{other}$.

Phylogeny Construction

Multiple sequence alignment (MSA) was performed using MUSCLE algorithm available in MEGA7 (Edgar, 2004). Best-fit model of nucleotide substitution was determined using MODEL TEST in MEGA7. Aligned sequence relatedness was evaluated using the Maximum Likelihood (default parameters with 1,000 bootstrap replicates) method based on Tamura parameter 3 model (T92) with Gamma distributed (G) available in MEGA7 (Kumar et al., 2016). The phylogenetic tree was rooted using TSWV N gene reported from South Korea as an outgroup.

Population Selection Studies and Neutrality Test

Mean rates of non-synonymous (dN) and synonymous substitutions (dS) were calculated using codon-based maximum likelihood methods, i.e., SLAC (single like ancestor counting), FEL (fixed effects likelihood), and REL (random effects likelihood). DATAMONKEY server (Weaver et al., 2018) was used to calculate dN/dS ratio. To test the theory of neutral evolution, the test statistics such as Tajimas's D (Tajima, 1989), Fu and Li's D, and Fu and Li's F (Fu and Li, 1993; Fu, 1997) were computed in DnaSP software.

Genetic Differentiation and Gene Flow Estimates

DnaSP was used to compute nucleotide test statistics such as Ks, Kst (Kst value close to zero indicate no differentiation), Snn (Snn value close to one indicates differentiation) (Hudson, 2000) and haplotype statistics Hs, Hst (Hudson et al., 1992a,b). These tests estimate genetic differentiation within the populations of IYSV genotypes. Fst statistics was used to estimate the extent of the gene flow (panmixia or free gene flow has values close to zero whereas infrequent gene flow attains values close to one) (Hudson et al., 1992b).

Recombination Detection Analysis (RDA)

Unaligned sequences were loaded in SDT v1.2 program, pairwise scan was performed with the MUSCLE, and the sequence data was saved with minimum identity of 70% and maximum of 100% to ensure sequences were properly aligned. The aligned IYSV N sequences were then used as an input query and analyzed for recombination events using Recombination Detection Program (RDP) v 4.0 (Martin and Rybicki, 2000), BOOTSCAN (Salminen et al., 1995), 3SEQ, GENECONV (Sawyer, 1999), MAXCHI (Maynard, 1992), CHIMAERA (Posada and Crandall, 2001) and SISCAN (Gibbs et al., 2000) available in RDP 4 Beta 4.88. Default settings for the different recombination detection methods and a Bonferroni corrected *P*-value cut-off of 0.05 were used for analysis.

Recombination Analysis Tool (RAT)

Recombination analysis tool (RAT) was used for the analysis of aligned nucleotide sequences (Etherington et al., 2005). RAT algorithm uses pairwise comparisons between sequences based on the distance method to identify recombinants in nucleotide sequence alignment. Percentage of nucleotide similarities were compared using a sliding window size of 10% of the sequence length and an increment size being half of the window size.

Bayesian Evolutionary Analysis by Sampling Trees (BEAST)

Bayesian phylogenetic analysis was performed in BEAST v2.4.6 (Bouckaert et al., 2014) to estimate evolutionary rate. Strict, relaxed (exponential, lognormal) and random local clocks were utilized for comparison (Bouckaert et al., 2014). Demographic models—coalescent constant population, coalescent exponential population, coalescent Bayesian skyline and coalescent extended Bayesian skyline were used to infer demographic history. "Temporal signal" (i.e., genetic changes between sampling times are sufficient and there is statistical relationship between genetic divergence and time) in the dataset was assessed using TempEst program (Rambaut et al., 2016). Using Markov Chain Monte Carlo (MCMC) method Bayesian phylogenies were constructed in BEAST v2.4.6. First 10% of the samples were discarded as burn-in. Convergence of the chain to stationary distribution and adequate sampling were analyzed using Tracer v1.6 (Tracer, 2018). Tracer was used to analyze the Effective Sample Size (ESS) and other prior parameter values. Tree Annotator was used for generating Maximum Clade Credibility (MCC) phylogenetic trees with common heights node. FigTree (2018) was used to generate the dendrograms.

RESULTS

In silico Restriction Fragment Length Polymorphism Analysis

Computational RFLP-based analysis of N gene sequences recognizing *Hin*fI restriction site divided the population into two major groups [79 NL (55.63%), 55 BR (38.73%)]. Thus, the genotype $IYSV_{NL}$ was found to be predominant

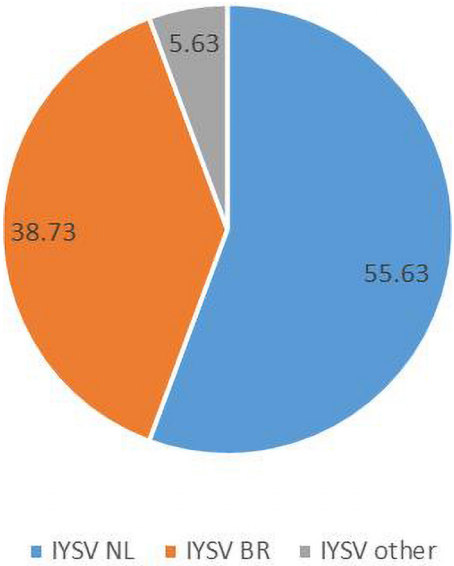


FIGURE 2 | Relative distribution of the global collection of Iris yellow spot orthotospovirus genotypes (indicated as percentages) based on Restriction Fragment Length Polymorphism.

over $IYSV_{BR}$ while the rest (5.63%) fell in neither category ($IYSV_{other}$) (**Figure 2**).

Molecular Phylogeny of IYSV N Gene

Phylogenetic tree of the N gene of IYSV constructed based on the aligned nucleotide sequences (**Figure 3**) using Maximum Likelihood method broadly clustered IYSV genotypes into two major clades (NL and BR types). Four $IYSV_{other}$ isolates (EU750697, KT27882, KT272883, and EU586203) and one $IYSV_{BR}$ isolate (FJ713700) clustered with the NL group. Similarly, four more $IYSV_{other}$ isolates (HQ148173, HQ148174, EU310290, and EU310291) and two $IYSV_{NL}$ isolates (FJ785835 and AF001387) clustered with the BR group. One $IYSV_{BR}$ isolate from Brazil (AF067070) formed a separate monophyletic clade along with a TSWV N gene isolate HQ267713 from South Korea (Outgroup). The clades also followed a geographical pattern as majority of $IYSV_{NL}$ genotypes are from North America and $IYSV_{BR}$ are from the Asian countries. Only one IYSV isolate has been reported from Brazil which formed a separate monophyletic clade even though *in silico* RFLP characterized it as an $IYSV_{BR}$ type.

Population Selection Studies, Neutrality Test, and Genetic Diversity Test

Gene codons that are in positive or negative selection pressure provide knowledge regarding the molecular evolution pattern of the N gene. The mean dN/dS (dN—rate of non-synonymous substitutions and dS—rate of synonymous substitutions) for N gene accessions belonging to $IYSV_{NL}$ group were found to be 0.192 with no positively selected codon site. SLAC methodology identified 20 negatively selected codons in the N gene of $IYSV_{NL}$ type. The data set when analyzed by FEL methodology revealed

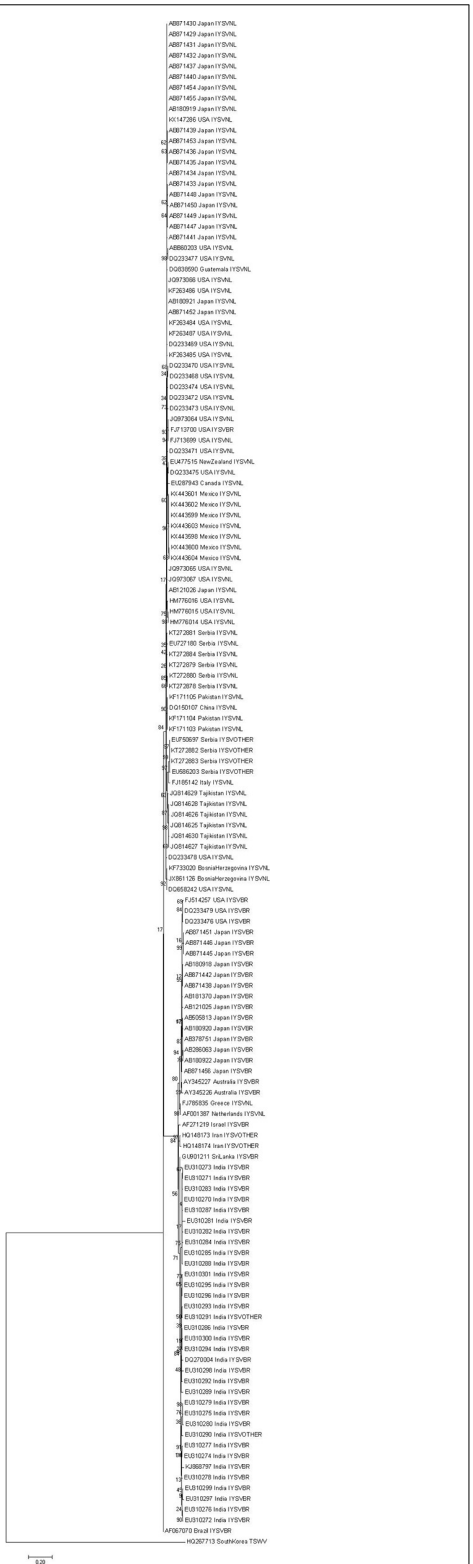


FIGURE 3 | Dendrogram based on the nucleocapsid (N) gene nucleotide sequences of Iris yellow spot orthotospovirus (IYSV). Tomato spotted wilt orthotospovirus N gene isolate HQ267713.1 from South Korea was used as an outgroup. GenBank accession number, place of origin, and IYSV genotype are the indicators of each IYSV isolate.

one positively selected codon site (codon no. 139) against 62 negatively selected codons. The dN-dS (mean difference between dN and dS) was –0.803 based on REL analysis denoting that the codon sites are under purifying selection acting against deleterious non-synonymous substitutions (**Table 1**).

For N gene accessions derived from $IYSV_{BR}$, the mean dN/dS was found to be 0.191 with no positively selected codon site. SLAC methodology identified 27 negatively selected codons. The data set when analyzed by FEL revealed one positively selected codon site (codon no. 139) against 63 negatively selected codons. The dN-dS was –0.813 based on REL analysis suggesting that the codon sites are under purifying selection acting against deleterious non-synonymous substitutions. Six positively selected codon (codon nos. 30, 40, 109, 139, 210, and 225) and seven negatively selected codons were identified by REL analysis, respectively (**Table 1**).

For N gene accessions of $IYSV_{other}$ group, the mean dN/dS was found to be 0.172 with no positively selected codon site. SLAC methodology identified three negatively selected codons. FEL methodology revealed no positively selected codon site against 38 negatively selected codons. The dN-dS was –0.805 based on REL analysis and it denotes codon sites are under purifying selection acting against deleterious non-synonymous substitutions. One positively selected codon (codon no. 270) and zero negatively selected codons were identified by REL analysis, respectively (**Table 1**).

Nucleotide diversity (π) of $IYSV_{BR}$ was about two folds higher than that for $IYSV_{NL}$ (0.04513 and 0.01990; respectively, **Table 2**).

TABLE 1 | Codon substitution in the nucleocapsid gene of Iris yellow spot orthotospovirus genotypes.

Genotype	Positively selected codon positions	No. of negatively selected codons	ω = dN/dS	dN-dS
$IYSV_{NL}$	139[b]	20[a]	0.192444	–0.803
		62[b]		
		73[c]		
$IYSV_{BR}$	139[b]	27[a]	0.19119	–0.813
	30[c]	63[b]		
	40[c]	07[c]		
	109[c]			
	139[c]			
	210[c]			
	225[c]			
$IYSV_{other}$	270[c]	3[a]	0.172079	–0.805
		38[b]		
$IYSV_{All}$	109[b]	54[a]	0.205279	–
	139[ab]	90[b]		

dN, the number of non-synonymous substitutions per non-synonymous site; dS, the number of synonymous substitutions per synonymous site ω—Ratio of dN/dS from SLAC (single like ancestor counting) methodology, dN-dS obtained from REL (random effects likelihood).
[a]Codons identified by SLAC at a cut-off p-value 0.1.
[b]Codons identified by FEL at a cut-off p-value 0.1.
[c]Codons identified by REL at a cut-off Bayes factor value 50.
$IYSV_{All}$ = $IYSV_{NL}$, $IYSV_{BR}$ and $IYSV_{other}$.

However, the highest nucleotide diversity among the IYSV isolates was found in $IYSV_{other}$ (0.08042) indicating $IYSV_{other}$ is more genetically diverse than the $IYSV_{NL}$ and $IYSV_{BR}$ (**Table 2**). Number of polymorphic sites (S) was 136 from the N gene sequences of eight isolates of $IYSV_{other}$, $IYSV_{NL}$ showed 189 polymorphic sites among the N gene sequences obtained from the 79 isolates, whereas $IYSV_{BR}$ showed 230 polymorphic sites obtained from the 55 isolates (**Table 2**). $IYSV_{other}$ is more diverse than IYSV NL and BR based on number of polymorphic sites (S).

Neutrality Test

The test of neutral evolution analyzed based on the total number of mutations and segregating sites, revealed statistically significant and non-significant negative values of test statistic Tajimas's D for $IYSV_{NL}$ and $IYSV_{BR}$, respectively (**Tables 3**, **4**). It indicates the operation of purifying selection and population expansion in major IYSV genotypes (NL and BR). Similarly, negative values of other test statistics such as Fu and Li's D and Fu and Li's F also corroborate the above findings with regard to $IYSV_{BR}$ and $IYSV_{NL}$ genotypes. However, positive values of all the test statistics such as Tajimas's D, Fu and Li's D, and Fu and Li's F with respect to the genotype $IYSV_{other}$ indicate the decrease in population size and act of balancing selection.

TABLE 2 | Genetic diversity test of Iris yellow spot orthotospovirus genotypes.

Genotype	N	S	π	Hd
$IYSV_{NL}$	79	189	0.01990	0.982
$IYSV_{BR}$	55	230	0.04513	0.999
$IYSV_{Other}$	08	136	0.08042	0.964
$IYSV_{All}$	142	317	0.07220	0.994

N, Number of isolates; S, Number of polymorphic (segregating) sites; Hd, haplotype diversity; π, nucleotide diversity within species; $IYSV_{All}$ = $IYSV_{NL}$, $IYSV_{BR}$ and $IYSV_{other}$.

TABLE 3 | Neutrality test of Iris yellow spot orthotospovirus genotypes based on total number of mutations.

Genotypes	Tajimas's D	Fu and Li's D	Fu and Li's F
$IYSV_{NL}$	−2.14425*	−1.43877	−2.07919
$IYSV_{BR}$	−1.24392	−1.92209	−1.98807
$IYSV_{Other}$	1.08486	0.71725	0.89745

*Calculated using total number of mutations. *statistically significant at $P < 0.01$.*

TABLE 4 | Neutrality tests of Iris yellow spot orthotospovirus genotypes based on total number of segregating sites.

Genotypes	Tajimas's D	Fu and Li's D	Fu and Li's F
IYSV NL	−1.96208*	−1.83392	−2.26501
IYSV BR	−0.93734	−2.14943	−2.01338
IYSV Other	1.42119	0.69566	0.96703

*Calculated using total number of segregating sites. *statistically significant at $P < 0.01$.*

TABLE 5 | Gene flow and genetic differentiation of Iris yellow spot orthotospovirus genotypes.

Genotypes	H_s	H_{st}	χ^2	*P*-value	K_t	K_s	K_{st}	S_{nn}	Z	F_{st}
$IYSV_{BR}$ vs. $IYSV_{NL}$	0.98869	0.00484	134	0.0768	58.49254	2.85914*	0.21377*	0.97761	*7.53452	0.72066
$IYSV_{BR}$ vs. $IYSV_{other}$	0.99516	0.00331	63	0.3368	44.90220	3.43202*	0.03724*	0.95238	6.42171*	0.26181
$IYSV_{NL}$ vs. $IYSV_{other}$	0.98056	0.00427	87	0.0427*	24.75033	2.56320*	0.06501*	0.98851	7.09306*	0.35394

H_s, H_{st}—measure genetic differentiation based on haplotype statistics.
Ks, Kst, Snn, Z—measure genetic differentiation based on nucleotide statistics.
*Fst—measures extent of gene flow. *statistically significant at $P < 0.05$.*

TABLE 6 | Recombination events in Iris yellow spot orthotospovirus nucleocapsid (N) gene detected by Recombination Detection Program (RDP).

Isolate	Parental isolate		Recombination Detection Program	Recombination event #	*P*-values
	Major	Minor			
AF067070_Brazil	JX861126_Bosnia	AY345226_Australia	GeneConv, 3Seq	1	3.193×10^{-5}–1.333×10^{-4}
HQ148174_Iran	EU310281_India	AB180922_Japan	MaxChi	2	–
AB180918_Japan	HQ148174_Iran	EU586203_Serbia	SiScan, 3Seq	22	2.879×10^{-08}–9.951×10^{-1}

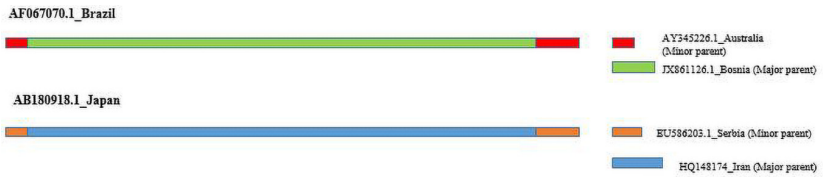


FIGURE 4 | Potential recombinants detected in Iris yellow spot orthotospovirus nucleocapsid (N) gene by Recombination Detection Program (RDP4). Major and Minor parent for each recombinants are shown.

Genetic Differentiation and Gene Flow

Haplotype-based statistics (Hs and Hst) and nucleotide-based statistics (Ks, Kst, Snn) were estimated to evaluate genetic differentiation between the IYSV genotypes (**Table 5**). The statistically significant test values of Ks, Kst and Z reveals strong genetic differentiation among the IYSV genotypes studied. Snn value close to one indicates genetic differentiation even though insignificant test statistical values were obtained. $IYSV_{BR}$ is more differentiated from $IYSV_{other}$ (Kst value of 0.03724*) compared to $IYSV_{NL}$ genotypes (0.21377*) based on the Kst values. Fst values show that the extent of gene flow between major genotypes, $IYSV_{BR}$ and $IYSV_{NL}$, is relatively high than the gene flow between individual BR and NL genotypes with $IYSV_{other}$. Among the major genotypes, $IYSV_{NL}$ shows greater gene flow with $IYSV_{other}$ than $IYSV_{BR}$.

Recombination Detection Analysis

Three potential recombination events were detected among the IYSV N genes analyzed (**Table 6** and **Figure 4**). AF067070 (BR type) IYSV isolate from Brazil is a potential recombinant of isolates: JX861126 Bosnia (major parent) and AY345226 Australia (minor parent). This recombinant was detected by GeneConv, 3Seq, algorithms in RDP. The recombination breakpoint begins at 789 in alignment (789 without gaps) with breakpoint clustering at 99% confidence interval ranging from 730 to 809 in alignment (730–809 without gaps) and breakpoint ends at 12 in alignment (12 without gaps) with breakpoint clustering at 99% confidence interval ranging from 822 to 44 in alignment (822–44 without gaps). The second recombination event involved isolate HQ148174 ($IYSV_{other}$) from Iran putatively arising from EU310281 India (major parent) and AB180922 Japan (minor parent). However, only MaxChi algorithm detected this recombinant. The third recombinant, AB180918 (BR type) IYSV isolate from Japan, is the result of a potential recombination event 22 arising from HQ148174 Iran (major parent) and EU586203 Serbia (minor parent). This recombinant was detected by SiScan and 3Seq algorithms of recombination detection program. The recombination breakpoint begins at 20 in alignment (20 without gaps) with breakpoint clustering at 99% confidence interval ranging from 731 to 3 in alignment (731–3 without gaps) and breakpoint ends at 820 in alignment (820 without gaps) with breakpoint clustering at 99% confidence interval ranging from 731 to 3 in alignment (731–3 without gaps). Among the three putative recombinants detected, two belonged to $IYSV_{BR}$ type and the remaining one belonged to $IYSV_{other}$ type. IYSV isolate from Brazil is a potential recombinant and hence this isolate formed a separate monophyletic clade in the phylogenetic tree (**Figure 3**).

Recombination Analysis Tool

Recombination analysis tool (RAT) was used to substantiate the findings of the RDP. An isolate was considered recombinant when the major and minor parent isolates intersect at two points in the graph (**Figure 5**). Based on this criterion, HQ148174_Iran and AB180918_Japan were considered potential recombinants

even though only MaxChi in the RDP4 program detected HQ148174_Iran as a recombinant.

Bayesian Evolutionary Analysis by Sampling Trees (BEAST)

The rates of nucleotide substitution in tospovirus genomes have not been reported. Therefore, the global repository of IYSV N gene sequences was used to estimate the rates of nucleotide substitution and discern the rapidity with which molecular evolution might occur in the tospoviruses. Genetic recombinants were removed for BEAST analysis since their inclusion violates the assumption of coalescent-based analyses and thus could result in incorrect estimates of the rate of evolution. For nucleotide models, Hasegawa-Kishino-Yano (HKY)-based analysis was performed and it converged satisfactorily. While comparing two models if the marginal posterior distributions of the log-likelihoods do not overlap then the model with the higher posterior distribution of log-likelihood was preferred. Estimate is a better approximation of the true posterior distribution when larger Effective Sample Size (ESS) is available (ESS > 200 are desirable). Based on the above criteria, General Time Reversible (GTR) relaxed exponential growth clock model with coalescent

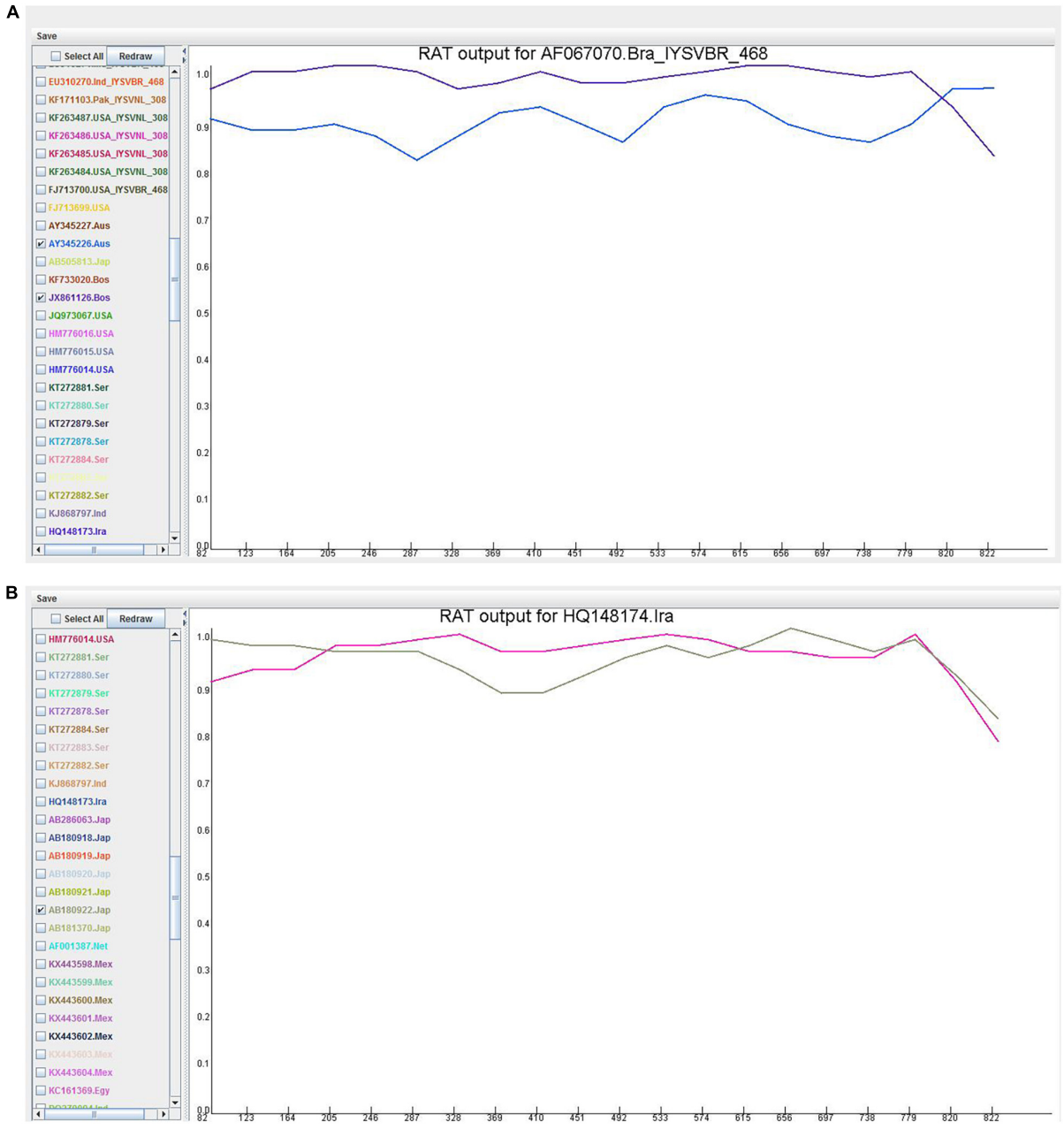


FIGURE 5 | (Continued)

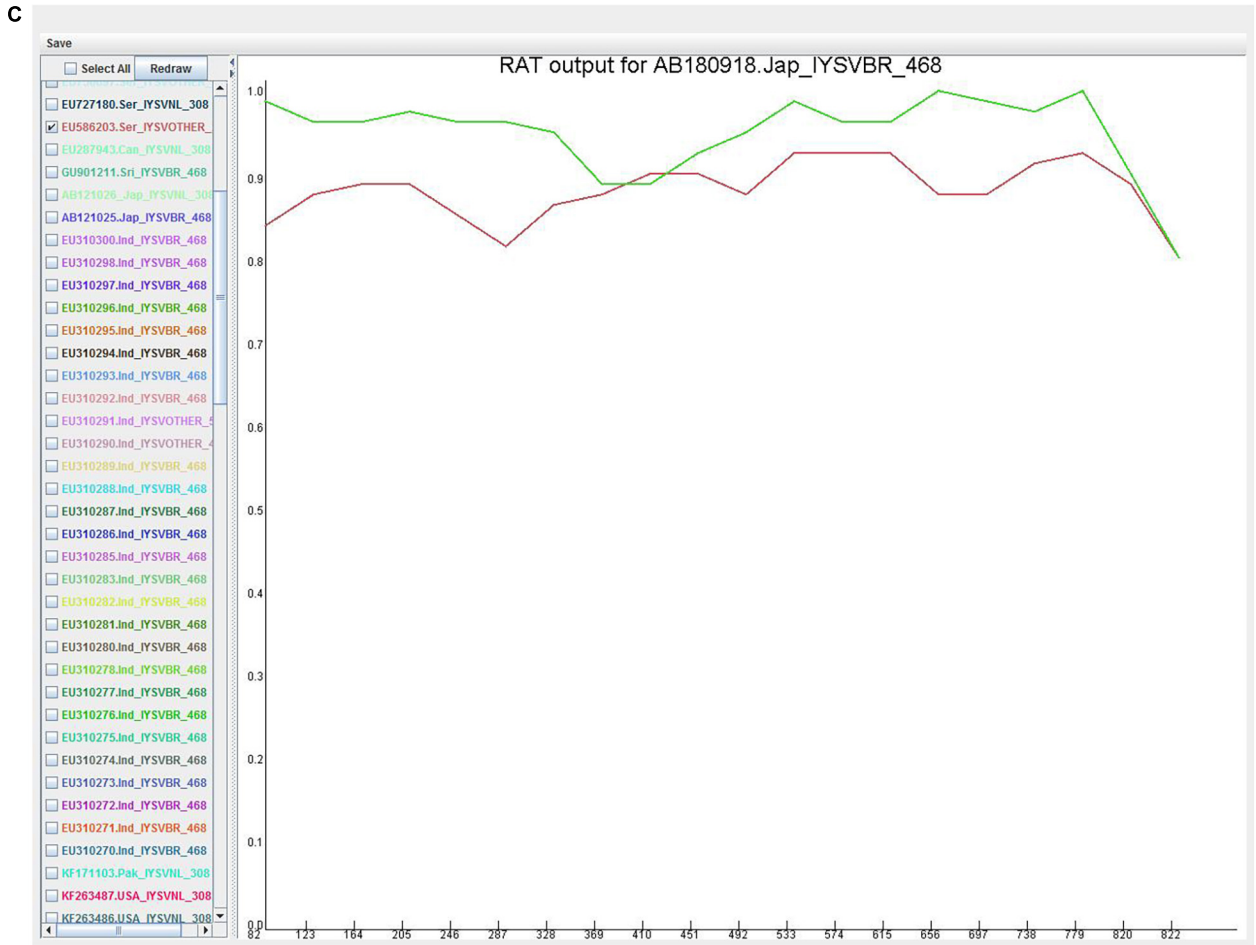


FIGURE 5 | Recombination cross over sites in the Iris yellow spot orthotospovirus nucleocapsid (N) gene generated by Recombination Analysis Tool. The *X*-axis denotes sequence location along the genome and *Y*-axis represents the genetic distance. **(A)** AF067070_Brazil, the crossover site is located at 3′ end portion of the genome; **(B)** HQ148174_Iran, the crossover sites were located toward the 5′ and 3′ ends of the genome; **(C)** AB180918_Japan, the crossover site is located in the middle portion of the genome.

constant population was found to be the best fit with a marginal likelihood mean substitution rate of 5.08×10^{-5} subs/site/year, 95% highest posterior density (HPD) substitution rate between $5.11 \times 10^{-5}$and 5.06×10^{-5} and ESS was 305 (**Supplementary Table 2**). Bayesian phylogenetic tree separated the IYSV isolates into two distinct clades, clade I comprising of $IYSV_{BR}$ isolates and clade II comprising $IYSV_{NL}$ isolates. The isolates that belonged to the same geographic region (or same country) clustered together (**Figure 6**).

DISCUSSION

The global population structure and temporal dynamics of the IYSV conducted previously (Iftikhar et al., 2014) based on N gene sequences delineated that the viral isolates could be categorized into two major genotypes ($IYSV_{BR}$ and $IYSV_{NL}$). Further, temporal dynamics of IYSV showed greater incidence of $IYSV_{BR}$ post-2005 compared to $IYSV_{NL}$. Since the last publication, the number of N gene sequences added to the public repository has increased significantly. To gain a better understanding of the evolutionary genomics and to further gain deeper insights into the evolution rate of IYSV, we analyzed 142 complete N gene sequences using a wide range of computational tools to infer molecular evolutionary genomics.

In silico RFLP analysis to categorize the genotype of IYSV isolates showed that the majority of the isolates belonged to $IYSV_{NL}$ category (55.63%), whereas 38.73% of $IYSV_{BR}$ isolates were observed. There was an increment in $IYSV_{NL}$ genotype incidence or characterization compared to $IYSV_{BR}$ since the last report (Iftikhar et al., 2014). Interestingly, gene flow estimates showed greater gene flow between NL and BR genotypes, rather than between the individual major genotypes and the "other" category. Even between the major genotypes, $IYSV_{NL}$ exhibited a greater gene flow with $IYSV_{other}$. Also, a greater genetic diversity was observed in $IYSV_{other}$, compared to NL and BR. However, codon substitution analysis of N gene showed little change since the last study (Iftikhar et al., 2014). In fact, the positively selected

FIGURE 6 | Maximum Clade Credibility (MCC) tree of Iris yellow spot orthotospovirus (IYSV) nucleocapsid (N) gene (nucleotide sequences) generated in BEAST v2.4.6. Nodes of IYSV isolates reported from different countries are colored.

codons (codon positions 139 in BR and NL and 270 in $IYSV_{other}$) remained intact despite the substantial increase in the number of isolates examined, suggesting the importance of these codon positions in improving the fitness of nucleocapsid protein. The negative selections in the other codon positions imply that the deleterious mutations in those positions are effectively removed in the IYSV population as a whole. Most of the codons of the N gene are neither under positive nor negative selection suggesting the neutral evolution of these codons.

Recombination is a common phenomenon in RNA viruses but the implications of recombination for evolution is not well studied (Sztuba-Solinska et al., 2011). There is a serious limitation in understanding the contribution of recombination to evolution of IYSV due to lack of full-length genome sequences. The potential recombinants identified in this study belonged to BR type which seems to be evolving using population expansion strategies. The recombination breakpoints were at 5' and 3' ends suggesting that these are potential hot spots for recombination (Gawande et al., 2015).

In the BEAST analysis, General Time Reversible (GTR) relaxed exponential growth clock model with coalescent constant population was found to be the best fit model explaining the genetic architecture of IYSV population. In a similar analysis of PVY genomic sequences, it was found that the relaxed uncorrelated log normal clock was the best fit with a population of constant size (Gibbs et al., 2017). Further, similar topology of the phylogenetic tree was obtained by both ML method and Bayesian MCC based phylogeny for IYSV.

PVY dating was reported by comparing the estimated phylogenetic dates with historical events in the worldwide adoption of potato and other PVY hosts (Gibbs et al., 2017). While the potato-PVY analysis was based on the sample collection dates over several decades, onion-IYSV interactions are relatively new and hence predicting the phylodynamic patterns and demographic history of IYSV require more such data on temporal scale. The PVY demographic history and population expansion was deduced and compared with that of geographic distribution of host (potato) suggesting direct influence of potato cultivation area on the population size of the virus (Mao et al., 2019). In this context, further studies how expansion of onion cultivation area influences the population expansion of IYSV will be interesting.

Bayesian coalescent estimates of evolutionary dynamics of citrus tristeza virus, based on the *p25* gene, showed that the rate of substitution was at 1.19×10^{-3} subs/site/year (Benítez-Galeano et al., 2017). Similarly, Bayesian phylogenetic reconstruction-based nucleotide substitution rates of CP gene derived from four species of viruses in *Secoviridae* family estimated it to range 9.29×10^{-3} to 2.74×10^{-3} (subs/site/year) (Thompson et al., 2014) while for tobamovirus the estimate ranged from 1×10^{-5} and 1.3×10^{-3} substitutions per site, per year (Pagan et al., 2010). Further, Bayesian analysis of VPg gene of PVY reveals that it has been evolving at a rate of 5.60×10^{-4} subs/site/year (Mao et al., 2019). Thus, the mean substitution rates identified for the IYSV N gene are comparable to those found in other plant-infecting RNA viruses. Substitution rates tend to be higher in RNA viruses as they are shown to mutate at faster rate. These mutations help in viral emergence on novel hosts but are not adaptive (Sacristan and Garcia-Arenal, 2008). Furthermore, the time of divergence of PVY clades, clade N and clade O, was found to be the year 1861

CE (95% credibility interval 1750–1948 CE) (Mao et al., 2019). Similar estimation of temporal divergence of $IYSV_{BR}$ and $IYSV_{NL}$ and the role of geographically driven adaptation of IYSV are worth exploring for a better understanding of the evolutionary dynamics of IYSV.

There are a very limited number of sequences of the other IYSV genes (NSm, NSs, G_N/G_C, RdRp) and even fewer complete genome sequences. Evolutionary analysis on such small sample size is not feasible. In the absence of complete genome sequences of a considerable number of the virus isolates (as is the case of IYSV) extrapolation of results of a single (or a few) gene(s) for the entire species is not uncommon. There are increasing number of studies on molecular evolutionary analysis, including phylodynamics and temporal evolutionary features of plant viruses based on a single or few viral gene sequences, such as VPg of PVY (Mao et al., 2019), NABP and CP genes of Potato virus M (PVM) (He et al., 2019) and P3, CI, Nib genes of (PVY) (Gao et al., 2020). However, to avoid any discrepancies in extrapolating the evolutionary analysis based on one or a few genes of a virus to the entire virus species, next generation sequencing–based sequencing followed by *de novo* assembly would provide a near complete genomic sequences that could be used to generate a more comprehensive picture of the genetic diversity of the virus populations (Zarghani et al., 2018).

CONCLUSION

$IYSV_{NL}$ was found to be the predominant genotype on a global scale. Interestingly, the $IYSV_{other}$ genotype is genetically more diverse than $IYSV_{BR}$ and $IYSV_{NL}$ genotypes. Population structure analysis revealed that it is under purifying selection and the phenomenon of population expansion is occurring. BEAST-based molecular clock analysis showed that the rates of molecular evolution of IYSV N gene are similar to other plant RNA viruses. This study is a step forward in identifying molecular factors that contribute to the evolution of IYSV, and serves as a foundation for further evolutionary genomic studies on one of the economically important plant virus groups.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AT and HP conceived and designed the experiments. AT and YZ performed the experiments. AT, SR, YZ, and HP analyzed the data. AT, CO, SR, YZ, RI, and HP contributed reagents, materials, and analysis tools. AT, SR, and HP wrote the manuscript, proof-read and finalized the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported in part by the Specialty Crop Block Grant Program from the Washington State Department of Agriculture (Grant # K2527), USDA National Institute of Food and Agriculture (NIFA), Specialty Crop Research Initiative (Grant No. 2018-5118128435), and USDA-NIFA Hatch project Accession #1016563 "Reducing the Impact of Pests and Diseases Affecting Washington Agriculture." The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

AT would like to thank the Indian Council of Agricultural Research and Washington State University for their fellowship and financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.633710/full#supplementary-material

REFERENCES

Bag, S., Schwartz, H. F., Cramer, C. S., Havey, M. J., and Pappu, H. R. (2015). Iris yellow spot virus (*Tospovirus*: *Bunyaviridae*): from obscurity to research priority. *Mol. Plant Pathol.* 16, 224–237. doi: 10.1111/mpp.12177

Benítez-Galeano, M. J., Castells, M., and Colina, R. (2017). The evolutionary history and spatiotemporal dynamics of the NC lineage of citrus Tristeza Virus. *Viruses* 9:272. doi: 10.3390/v9100272

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Butkovic, A., Gonzalez, R., and Elena, S. F. (2021). Revisiting *Orthotospovirus* phylogeny using full-genome data and testing the contribution of selection, recombination and segment reassortment in the origins of members of new species. *Arch. Virol.* 166, 491–499. doi: 10.1007/s00705-020-04902-1

Centre for Agriculture and Bioscience International (CABI) - Invasive Species Compendium (2019). Available online at: https://www.cabi.org/isc/datasheet/28848 (accessed March 29, 2021).

Chen, S., Xing, Y., Su, T., Zhou, Z., Dilcher, D. L., and Soltis, D. E. (2012). Phylogeographic analysis reveals significant spatial genetic structure of *Incarvillea sinesisas* a product of mountain building. *BMC Plant Biol.* 12:58. doi: 10.1186/1471-2229-12-58

Cordoba-Selles, C., Cebrián-Mico, C., Alfaro-Fernández, A., Muñoz-Yerbes, M. J., and Jordá-Gutiérrez, C. (2007). First report of Iris yellow spot virus in commercial leek (*Allium porrum*) in Spain. *Plant Dis.* 91:1365. doi: 10.1094/PDIS-91-10-1365B

de Avila, A. C., de Haan, P., Kormelink, R., Resende, R. O., Goldbach, R. W., and Peters, D. (1993). Classification of tospoviruses based on phylogeny of nucleoprotein gene sequences. *J. Gen. Virol.* 74, 153–159. doi: 10.1099/0022-1317-74-2-153

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

Etherington, G. J., Dicks, J., and Roberts, I. N. (2005). Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics* 21, 278–281. doi: 10.1093/bioinformatics/bth500

FigTree (2018). Available online at: http://tree.bio.ed.ac.uk/software/figtree/ (accessed March 29, 2021).

Fu, Y. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915–925.

Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133, 693–709.

Gao, F., Kawakubo, S., Ho, S. Y., and Ohshima, K. (2020). The evolutionary history and global spatio-temporal dynamics of potato virus Y. *Virus Evol.* 6:veaa056. doi: 10.1093/ve/veaa056

Gawande, S., Gurav, V. S., Ingle, A. A., Martin, D. P., Asokan, R., and Gopal, J. (2015). Sequence analysis of Indian iris yellow spot virus ambisense genome segments: evidence of interspecies RNA recombination. *Arch. Virol.* 160, 1285–1289. doi: 10.1007/s00705-015-2354-x

Gent, D. H., Mohan, S. K., du Toit, L. J., Pappu, H. R., Fichtner, S. F., and Schwartz, H. F. (2006). *Iris Yellow Spot Virus*: an emerging threat to onion bulb and seed production. *Plant Dis.* 90, 1468–1480. doi: 10.1094/PD-90-1468

Gibbs, A. J., and Ohsima, K. (2010). Potyviruses and the digital revolution. *Annu. Rev. Phytopathol.* 48, 205–223. doi: 10.1146/annurev-phyto-073009-114404

Gibbs, A., Ohshima, K., Yasaka, R., Mohammadi, M., Gibbs, M. J., and Jones, R. A. C. (2017). The phylogenetics of the global population of potato virus Y and its necrogenic recombinants. *Virus Evol.* 3:vex002. doi: 10.1093/ve/vex002

Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000). Sister – scanning: a monte carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573–582. doi: 10.1093/bioinformatics/16.7.573

He, Z., Chen, W., Yasaka, R., Chen, C., and Chen, X. (2019). Temporal analysis and adaptive evolution of the global population of potato virus M. *Infect. Genet. Evol.* 73, 167–174. doi: 10.1016/j.meegid.2019.04.034

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the quaternary. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 359, 183–195. doi: 10.1098/rstb.2003.1388

Hudson, R. R. (2000). A new statistic for detecting genetic differentiation. *Genetics* 155, 2011–2014.

Hudson, R. R., Boos, D. D., and Kaplan, N. L. (1992a). A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* 9, 138–151. doi: 10.1093/oxfordjournals.molbev.a040703

Hudson, R. R., Slatkin, M., and Maddison, W. P. (1992b). Estimations of levels of gene flow from DNA sequence data. *Genetics* 132, 583–589.

Iftikhar, R., Ramesh, S. V., Bag, S., Ashfaq, M., and Pappu, H. R. (2014). Global analysis of population structure, spatial and temporal dynamics of genetic diversity, and evolutionary lineages of *Iris yellow spot virus* (*Tospovirus:Bunyaviridae*). *Gene* 547, 111–118. doi: 10.1016/j.gene.2014.06.036

Karavina, C., Ibaba, J. D., Gubba, A., and Pappu, H. R. (2016). First report of *Iris yellow spot virus* infecting garlic and leek in Zimbabwe. *Plant Dis.* 100:657. doi: 10.1094/PDIS-09-15-1022-PDN

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Mandal, B., Jain, R. K., Krishnareddy, M., KrishnaKumar, N. K., Ravi, K. S., and Pappu, H. R. (2012). Emerging problems of Tospoviruses (*Bunyaviridae*) and their management in the Indian subcontinent. *Plant Dis.* 96, 468–479. doi: 10.1094/PDIS-06-11-0520

Mao, Y., Bai, Y., Shen, J., Gao, F., Sun, X., Qiu, G., et al. (2019). Molecular evolutionary analysis of potato virus Y infecting potato based on the VPg gene. *Front. Microbiol.* 10:1708. doi: 10.3389/fmicb.2019.01708

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

Maynard, S. J. (1992). Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34, 126–129. doi: 10.1007/BF00182389

Moya, A., Holmes, E. C., and Gonzalez-Candelas, F. (2004). The population genetics and evolutionary epidemiology of RNA viruses. *Nat. Rev. Microbiol.* 2, 279–288. doi: 10.1038/nrmicro863

Nischwitz, C., Pappu, H. R., Mullis, S. W., Sparks, A. N., Langston, D., Csinos, A. S., et al. (2007). Phylogenetic analysis of *Iris yellow spot virus* isolates from onion *(Allium cepa)* in Georgia (USA) and Peru. *J. Phytopathol.* 155, 531–535.

Oliver, J. E., and Whitfield, A. E. (2016). The genus *Tospovirus*: emerging Bunyaviruses that threaten food security. *Annu. Rev. Virol.* 3, 101–124. doi: 10.1146/annurev-virology-100114-055036

Pagan, I., Firth, C., and Holmes, E. C. (2010). Phylogenetic analysis reveals rapid evolutionary dynamics in the plant RNA virus genus *Tobamovirus*. *J. Mol. Evol.* 71, 298–307. doi: 10.1007/s00239-010-9385-4

Pappu, H. R., du Toit, L. J., Schwartz, H. F., and Mohan, K. (2006). Sequence diversity of the nucleoprotein gene of Iris yellow spot virus (genus *Tospovirus* family *Bunyaviridae*) isolates from the western region of the United States. *Arch. Virol.* 151, 1015–1023. doi: 10.1007/s00705-005-0681-z

Pappu, H. R., Jones, R. A. C., and Jain, R. K. (2009). Global status of tospovirus epidemics in diverse cropping systems: successes gained and challenges ahead. *Virus Res.* 141, 219–236. doi: 10.1016/j.virusres.2009.01.009

Pappu, H. R., Rosales, I. M., and Druffel, K. L. (2008). Serological and molecular assays for rapid and sensitive detection of *Iris yellow spot virus* infection of bulb and seed onion crops. *Plant Dis.* 92, 588–594. doi: 10.1094/pdis-92-4-0588

Pappu, H. R., Whitfield, A. E., and Oliveira, A. (2020). *Tomato Spotted Wilt Virus. Encylopedia of Virology, Reference Module in Life Sciences*, 4th Edn, Cambridge, MA: Elsevier Press. doi: 10.1016/B978-0-12-809633-8.21329-0

Posada, D., and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13757–13762. doi: 10.1073/pnas.241370698

Rambaut, A., Lam, T. T., Carvalho, L. M., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Resende, R., Whitfield, A. E., and Pappu, H. R. (2020). "Orthotospoviruses (*Tospoviridae*)," in *Encylopedia of Virology*, 4th Edn., eds D. H. Bamford and M. Zuckerman (Cambridge, MA: Academic Press), 507–515. doi: 10.1016/B978-0-12-809633-8.21337-X

Restriction Mapper (2009). Available online at: www.restrictionmapper.org (accessed March 29, 2021).

Sacristan, S., and Garcia-Arenal, F. (2008). The evolution of virulence and pathogenicity in plant. *Mol. Plant Pathol.* 9, 369–384. doi: 10.1111/j.1364-3703.2007.00460.x

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

Sawyer, S. A. (1999). *Geneconv: A Computer Package for the Statistical Detection of Gene Conversion*. St. Louis, MI: Department of Mathematics, Washington University.

Sztuba-Solinska, J., Urbanowicz, A., Figlerowicz, M., and Bujarski, J. J. (2011). RNA-RNA recombination in plant virus replication and evolution. *Annu. Rev. Phytopathol.* 49, 415–443. doi: 10.1146/annurev-phyto-072910-095351

Tabassum, A., Reitz, S., Rogers, P., and Pappu, H. R. (2016). First report of Iris yellow spot virus infecting green onion (*Allium fistulosum*) in the United States. *Plant Dis.* 100:2539. doi: 10.1094/PDIS-05-16-0599-PDN

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.

Thompson, J. R., Kamath, N., and Perry, K. L. (2014). An evolutionary analysis of the *Secoviridae* family of viruses. *PLoS One* 10:e0119267. doi: 10.1371/journal.pone.0119267

Tracer (2018). Available online at: http://tree.bio.ed.ac.uk/software/tracer/ (accessed March 29, 2021).

Turina, M., Kormelink, R., and Resende, R. O. (2016). Resistance to Tospoviruses in vegetable crops: epidemiological and molecular aspects. *Annu. Rev. Phytopathol.* 54, 347–371. doi: 10.1146/annurev-phyto-080615-095843

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., and Pond, S. L. K. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* 35, 773–777. doi: 10.1093/molbev/msx335

Zarghani, S. N., Hily, J. M., Glasa, M., Marais, A., Wetzel, T., Faure, C., et al. (2018). Grapevine virus T diversity as revealed by full-length genome sequences assembled through high-throughput sequence data. *PLoS One* 13:e0206010. doi: 10.1371/journal.pone.0206010

Zen, S., Okuda, M., Ebihara, K., Uematsu, S., Hanada, K., Iwanami, T., et al. (2005). Genetic differentiation of *Iris yellow spot virus* on onion (*Allium cepa*) and pathogenicity of two IYSV strains on onion and leaf onion (*A. schoenoprasum*). *Jpn. J. Phytopathol.* 71, 123–126. doi: 10.3186//jjphytopath.71.123

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership