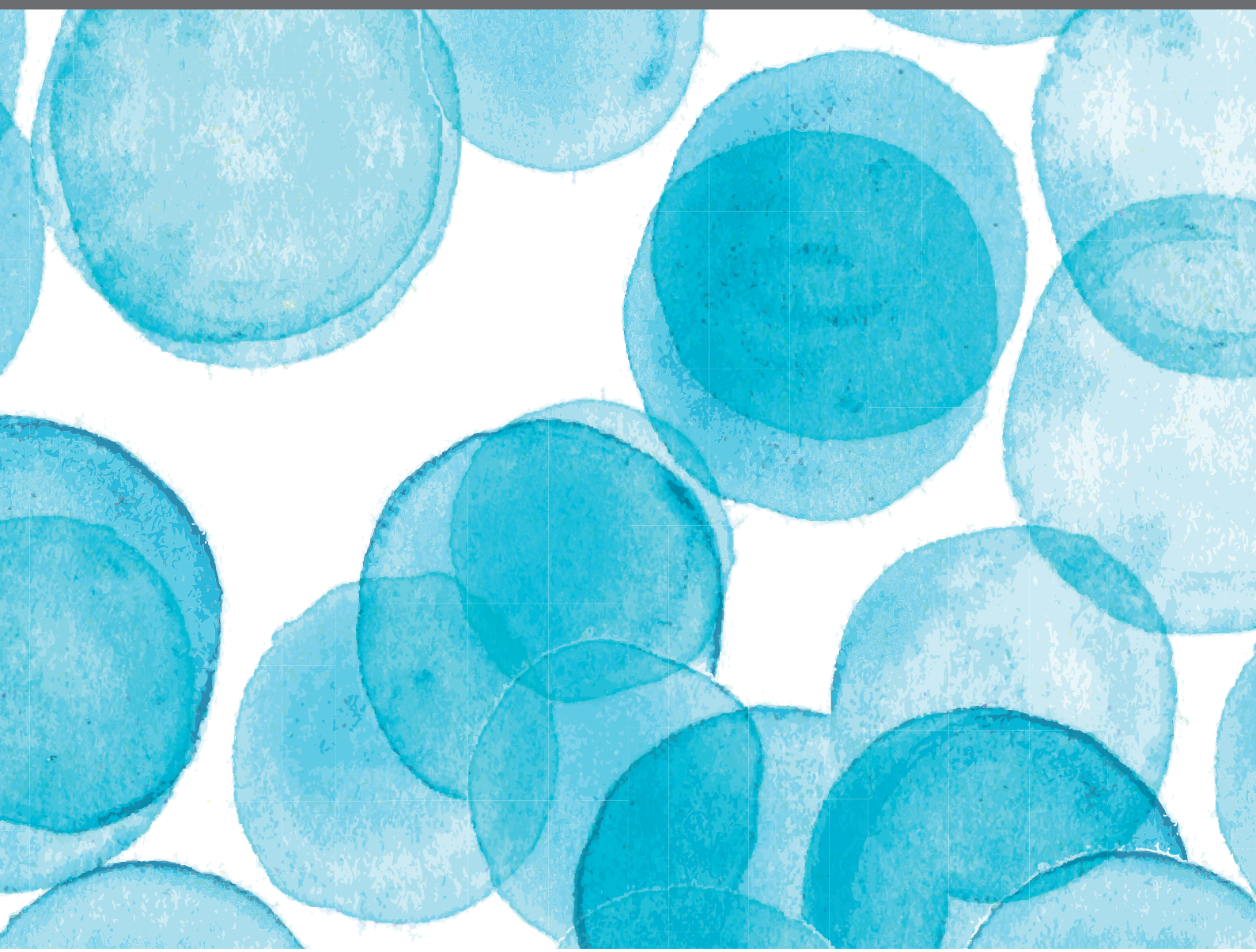
NEXT-GENERATION SEQUENCING AND CRISPR-CAS EDITING IN PLANT VIROLOGY

EDITED BY: Ahmed Hadidi, Henryk Hanokh Czosnek and John Wesley Randles

PUBLISHED IN: Frontiers in Microbiology and Frontiers in Plant Science



frontiers Research Topics

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88971-702-6
DOI 10.3389/978-2-88971-702-6

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

NEXT-GENERATION SEQUENCING AND CRISPR-CAS EDITING IN PLANT VIROLOGY

Topic Editors:
Ahmed Hadidi, United States Department of Agriculture, United States
Henryk Hanokh Czosnek, Hebrew University of Jerusalem, Israel
John Wesley Randles, University of Adelaide, Australia

Citation: Hadidi, A., Czosnek, H. H., Randles, J. W., eds. (2021). Next-Generation Sequencing and CRISPR-Cas Editing in Plant Virology.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-702-6

Table of Contents

05 ***Editorial: Next-Generation Sequencing and CRISPR-Cas Editing in Plant Virology***
Ahmed Hadidi, Henryk Czosnek and John W. Randles

09 ***Virome of* Camellia japonica*: Discovery of and Molecular Characterization of New Viruses of Different Taxa in Camellias***
Song Zhang, Liu Yang, Lisha Ma, Xin Tian, Ruhui Li, Changyong Zhou and Mengji Cao

23 ***Robust Virome Profiling and Whole Genome Reconstruction of Viruses and Viroids Enabled by Use of Available mRNA and sRNA-Seq Datasets in Grapevine (*Vitis vinifera *L.)***
V. Kavi Sidharthan, Amitha Mithra Sevanthi, Sarika Jaiswal and V. K. Baranwal

37 ***Insights Into Potato Spindle Tuber Viroid Quasi-Species From Infection to Disease***
Charith Raj Adkar-Purushothama, François Bolduc, Pierrick Bru and Jean-Pierre Perreault

53 ***Apple Russet Ring and Apple Green Crinkle Diseases: Fulfillment of Koch's Postulates by Virome Analysis, Amplification of Full-Length cDNA of Viral Genomes,* in vitro *Transcription of Infectious Viral RNAs, and Reproduction of Symptoms on Fruits of Apple Trees Inoculated With Viral RNAs***
Chunjiang Li, Hajime Yaegashi, Ryusuke Kishigami, Ayaka Kawakubo, Noriko Yamagishi, Tsutae Ito and Nobuyuki Yoshikawa

68 ***Genome Editing of* eIF4E1 *in Tomato Confers Resistance to Pepper Mottle Virus***
Yoo-Joung Yoon, Jelli Venkatesh, Joung-Ho Lee, Jinhee Kim, Hye-Eun Lee, Do-Sun Kim and Byoung-Cheorl Kang

79 ***Novel Fig-Associated Viroid-Like RNAs Containing Hammerhead Ribozymes in Both Polarity Strands Identified by High-Throughput Sequencing***
Alejandro Olmedo-Velarde, Beatriz Navarro, John S. Hu, Michael J. Melzer and Francesco Di Serio

90 ***High-Throughput Sequencing for Deciphering the Virome of Alfalfa (*Medicago sativa *L.)***
Nicolas Bejerman, Philippe Roumagnac and Lev G. Nemchinov

107 ***The Plant Negative-Sense RNA Virosphere: Virus Discovery Through New Eyes***
Nicolás Bejerman, Humberto Debat and Ralf G. Dietzgen

121 ***Artificially Edited Alleles of the Eukaryotic Translation Initiation Factor 4E1 Gene Differentially Reduce Susceptibility to Cucumber Mosaic Virus and Potato Virus Y in Tomato***
Hiroki Atarashi, Wikum Harshana Jayasinghe, Joon Kwon, Hangil Kim, Yosuke Taninaka, Manabu Igarashi, Kotaro Ito, Tetsuya Yamada, Chikara Masuta and Kenji S. Nakahara

131 ***Next-Generation Sequencing Combined With Conventional Sanger Sequencing Reveals High Molecular Diversity in* Actinidia *Virus 1 Populations From Kiwifruit Grown in China***
Shaohua Wen, Guoping Wang, Zuokun Yang, Yanxiang Wang, Min Rao, Qian Lu and Ni Hong

147 ***Efficient, Rapid, and Sensitive Detection of Plant RNA Viruses With One-Pot RT-RPA–CRISPR/Cas12a Assay***
Rashid Aman, Ahmed Mahas, Tin Marsic, Norhan Hassan and Magdy M. Mahfouz

157 ***A Mixed Infection of Helenium Virus S With Two Distinct Isolates of Butterbur Mosaic Virus, One of Which Has a Major Deletion in an Essential Gene***
John Hammond, Michael Reinsel, Samuel Grinstead, Ben Lockhart, Ramon Jordan and Dimitre Mollov

172 ***Next-Generation Sequencing Reveals a Novel Emaravirus in Diseased Maple Trees From a German Urban Forest***
Artemis Rumbou, Thierry Candresse, Susanne von Bargen and Carmen Büttner

184 ***Next-Generation Sequencing and the CRISPR-Cas Nexus: A Molecular Plant Virology Perspective***
Muhammad Shafiq Shahid, Muhammad Naeem Sattar, Zafar Iqbal, Amir Raza and Abdullah M. Al-Sadi

208 ***Application of CRISPR/Cas for Diagnosis and Management of Viral Diseases of Banana***
Leena Tripathi, Valentine Otang Ntui, Jaindra Nath Tripathi and P. Lava Kumar

221 ***Next-Generation Sequencing Identification and Characterization of MicroRNAs in Dwarfed Citrus Trees Infected With Citrus Dwarfing Viroid in High-Density Plantings***
Tyler Dang, Irene Lavagi-Craddock, Sohrab Bodaghi and Georgios Vidalakis

frontiers in Microbiology

EDITORIAL
published: 04 October 2021
doi: 10.3389/fmicb.2021.723278

Editorial: Next-Generation Sequencing and CRISPR-Cas Editing in Plant Virology

Ahmed Hadidi[1]*, *Henryk Czosnek*[2] and *John W. Randles*[3]

[1] US Department of Agriculture, Agricultural Research Service, Beltsville, MD, United States, [2] The Hebrew University of Jerusalem, Rehovot, Israel, [3] The University of Adelaide, Waite Campus, Glen Osmond, SA, Australia

Keywords: next-generation sequencing, CRISPR-Cas editing, viruses, viroids, plant virus diagnosis

Editorial on the Research Topic

Next-Generation Sequencing and CRISPR-Cas Editing in Plant Virology

The Research Topic was initiated in October 2019 to demonstrate how next-generation sequencing (NGS) and clustered regularly interspaced short palindromic repeats-associated Cas protein (CRISPR-Cas) editing are being applied in plant virology. NGS combined with bioinformatics has changed both basic and applied research in many biological disciplines by its ability to deliver fast, inexpensive and accurate genome and transcriptome information (Hadidi and Barba, 2012; Barba et al., 2014). NGS has been used since 2009 in a large number of studies in plant virology, including discovery of novel viruses and viroids or their variants, sequencing the genome of known pathogens or the 21–24 nucleotide pathogen-derived small RNAs (sRNAs) generated during the infection process which frequently cover the whole virus or viroid genome. Applications include the detection and identification of pathogens, extension of their known host range, investigating pathogen-host or -vector interactions, analysis of genome diversity and evolution, mRNA targeting, symptom expression, and the study of pathogen biology (Barba et al., 2014; Hadidi et al., 2016; Hadidi, 2019; Villamor et al., 2019). In addition, metagenomics coupled with NGS of insect vectors of plant viruses allowed the discovery of virus species and variants not known to be present in the region and infecting important crops (Ng et al., 2011; Rosario et al., 2015; Fontenele et al., 2018).

CRISPR-Cas and derived systems confer adaptive immunity against bacteriophages and plasmids in many bacteria and most archaea (Dounda and Charpentier, 2014). CRISPR-Cas systems act as RNA-guided programmable nucleases to degrade DNA and/or RNA derived from foreign nucleic acids by preserving molecular memory information of prior infections (Dounda and Charpentier, 2014; O'Connell et al., 2014; Abudayyeh et al., 2016). The CRISPR-Cas-based genome editing is emerging as a powerful tool for developing plant virus-resistant crops by directly targeting the virus genome or indirectly by editing host susceptibility factors (Sanfaçon, 2015). The technique is precise in editing the target genome with and without double-stranded breaks or donor templates.

The objectives of the Research Topic were to publish high-quality research papers and review articles focusing on the following. (1) Utilization of NGS in research and diagnosis of plant viruses and viroids; (2) Direct targeting of specific nucleotide sites of plant viruses and viroids by CRISPR-based genome editing to enable the fast introduction of resistance; (3) Targeting sites of specific plant genes such as eIF4E gene or eIF (iso) 4E gene by CRISPR-Cas editing to develop plants resistant to viruses; (4) Application of the CRISPR-Cas systems for rapid and accurate diagnosis of plant viruses and viroids,. Sixteen articles have been included in this Research Topic, comprising four reviews and 12 research papers. Nine of the research papers were on plant viruses and three on

OPEN ACCESS

Edited by:
Magdy Mahfouz, King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Peter Palukaitis, Seoul Women's University, South Korea
Lev G. Nemchinov, United States Department of Agriculture (USDA), United States

***Correspondence:**
Ahmed Hadidi
ahadidi@yahoo.com

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

Received: *10 June 2021*
Accepted: *30 August 2021*
Published: *04 October 2021*

Citation:
Hadidi A, Czosnek H and Randles JW (2021) Editorial: Next-Generation Sequencing and CRISPR-Cas Editing in Plant Virology. Front. Microbiol. 12:723278. doi: 10.3389/fmicb.2021.723278

viroids. Articles dealt with utilization of NGS in discovering and characterizing novel and known viruses and viroids, and applying CRISPR-Cas editing to plant translation factors for developing plants resistant to virus infection, and for rapid detection of RNA plant viruses.

NEW KNOWLEDGE GAINED BY NGS FOR NOVEL AND KNOWN PLANT VIRUSES AND VIROIDS

Apple russet ring and green crinkle diseases were of undetermined etiology for many years. Using NGS, other technologies, and applying Koch's postulates, it was demonstrated that one of the sequence variants of apple chlorotic leaf spot virus causes a characteristic ring-shaped rust on fruits of infected apple trees and that a sequence variant of apple stem pitting virus causes green crinkle symptoms on an infected apple fruit. Koch's postulates were fulfilled to demonstrate the viral etiology of both apple diseases (Li et al.). It was also revealed by NGS-based analyses of mRNA and sRNA of grapevine in India that 23 known viruses and viroids occur in this host (Sidharthan et al.). The mRNA -based approach identified more pathogens than that of the sRNA. The former approach was at the same level as that of the whole transcriptome in viral identification. Genomes of 19 viruses and viroids were characterized. Identification of three recombination events and phylogenetic analyses using characterized genomes suggested possible introduction of grapevine viruses and viroids into India from several continents through infected vegetatively propagated planting material. In kiwifruit, NGS analyses revealed high molecular diversity in Actinidia virus 1 (AcV-1) populations, with the highest sequence variation among its 12 open reading frames (ORFs) occurring at ORF1a, ORF2, and ORF3 (Wen et al.). Different domain compositions were shown for the first-time in the viral ORF1a. In addition, molecular recombination events among AcV-1 variants were revealed.

In diseased *Camellia japonica,* the common camellia, five novel viruses, and two known viruses (geminivirus and blunervirus) and a large number of betaflexiviruses were discovered by NGS (Zhang et al.). NGS and phylogenetic analyses of two of the novel viruses, tentatively named camellia chlorotic ringspot viruses, suggested that they may represent a novel genus in the family *Fimoviridae.* The other three novel viruses belong to the genera *Idaeovirus*, *Badnavirus,* and *Marafivirus*. These findings may serve as a basis for better management of the above viruses in common camellia and possibly other hosts. NGS was also used for deciphering the virome of alfalfa plants which led to the characterization of several previously known but not fully described viruses as well as the identification of many novel viral pathogens, including alfalfa dwarf virus, alfalfa enamovirus, alfalfa leaf curl virus, alfalfa virus F., alfalfa ringspot-associated virus, and others (Bejerman et al.). Exploring the alfalfa virome by NGS has also demonstrated the impotence of viral infections in a single plant. NGS from a mixed infection of an unknown cultivar of *Veronica sp.* revealed the genome sequences of two isolates of helenium virus S (HelVS) and two distinct isolates of butterbur mosaic virus (ButMV), ButMV-A and ButMV-B. A major deletion in an essential gene of ButMV-B was identified for the first time by NGS which is maintained through complementation by ButMV-A. The HelVS host range was extended to *Veronica* sp. and it was confirmed that the virus is a distinct species in the genus *Carlavirus* (Hammond et al.). The full-length genome of a novel emaravirus was identified and characterized by NGS from diseased symptomatic sycamore maple (*Acer pseudoplatanus*)—a tree species of significance in urban and forest areas. The virus was tentatively named maple mottle-associated virus (MaMaV). Phylogenetic and sequence analyses place MaMaV in the distinct "subgroup a" clade within the *Emaravirus* genus. This is the first time an emaravirus has been described from maple and fully genetically characterized (Rumbou et al.). NGS also allowed the discovery and genome sequences of at least 70 negative-sense and ambisense RNA (NSR) plant viruses. These viruses belong to several genera in seven families: *Ophiovirus* in the family *Aspiviridae*; *Coguvirus*, *Rubodvirus*, and *Tenuivirus* in the family *Phenuiviridae*; *Orthotospovirus* in the family *Tospoviridae*; *Emaravirus* in the family *Fimoviridae*; *Cytorhabdovirus*, *Betanucleorhabdovirus*, *Alphanucleorhabdovirus*, *Dichorhavirus*, and *Varicosavirus* in the family *Rhabdoviridae*. It is predicted that the increasing use of NGS, not only for plant samples but also in arthropod vectors, will allow the identification of many novel NSR viruses which will be crucial to unraveling the evolution of many NSR virus clades (Bejerman et al.).

The 10 most abundant sequence variants of potato spindle tuber viroid (PSTVd) RG1, expressed 1–4 weeks after infecting tomato plants, were identified by NGS in the regions favoring mutations. The findings of the effect of mutations on PSTVd secondary structure and its derived small RNAs increased our knowledge of the biological role of sequence variants, PSTVd interaction with host components, stability of structures generated by mutants during the course of infection, and stabilizing viroid population dynamic as influenced by variant sequences (Adkar-Purushothama et al.). Novel circular RNAs, 357–360 nt, containing hammerhead ribozymes in both polarity strands were discovered by NGS analyses of fig tree leaves (Olmedo-Velarde et al.). Bioassays, however, are needed to demonstrate whether the RNAs are viroids or viral satellites. MicroRNAs (miRNAs) were identified by NGS in dwarfed citrus trees in response to infection by citrus dwarfing viroid (CDVd). The 60 miRNAs identified were conserved in stem and root tissues. Three conserved miRNAs (csi-miR479, csi-miR171b, and csi-miR156) were significantly downregulated in the stems of CDVd-infected trees compared to the non-infected controls. These miRNAs are known to be involved in various physiological and developmental processes some of which may be related to the characteristic dwarfed phenotype displayed by CDVd-infected sweet orange on trifoliate orange rootstock field trees. Only one miRNA (csi-miR535) was significantly downregulated in CDVd-infected roots and it was predicted to target genes controlling a wide range of cellular functions. These findings indicate that CDVd-responsive plant miRNAs play a role in regulating important citrus tree growth and developmental processes which may participate in the cellular changes leading to the observed citrus tree dwarf phenotype (Dang et al.).

NEW KNOWLEDGE GAINED BY CRISPR-Cas EDITING

A-Developing Plants Resistant to Virus Infection

The eukaryotic translation initiation factor *eIF4E1* gene on chromosome 3 of a commercial cultivar of tomato was mutated by CRISPR/Cas9. Artificially edited alleles of *eIF4E1* gene differentially reduced susceptibility to cucumber mosaic virus and potato virus Y in tomato (Atarashi et al.). Similarly, site specific mutation of the tomato *eIF4E1* gene by CRISPR-Cas9 successfully conferred enhanced resistance to infection by pepper mottle virus (Yoon et al.). A banana streak virus (BSV) genome sequence that integrates at a single locus into the banana B genome is known as endogenous BSV (eBSV) whereas the virus genome in the replicative form in cells is known as the episomal form. CRISPR-Cas editing was applied to banana to control BSV by inactivating the eBSV integrated into the host genome (Tripathi et al.).

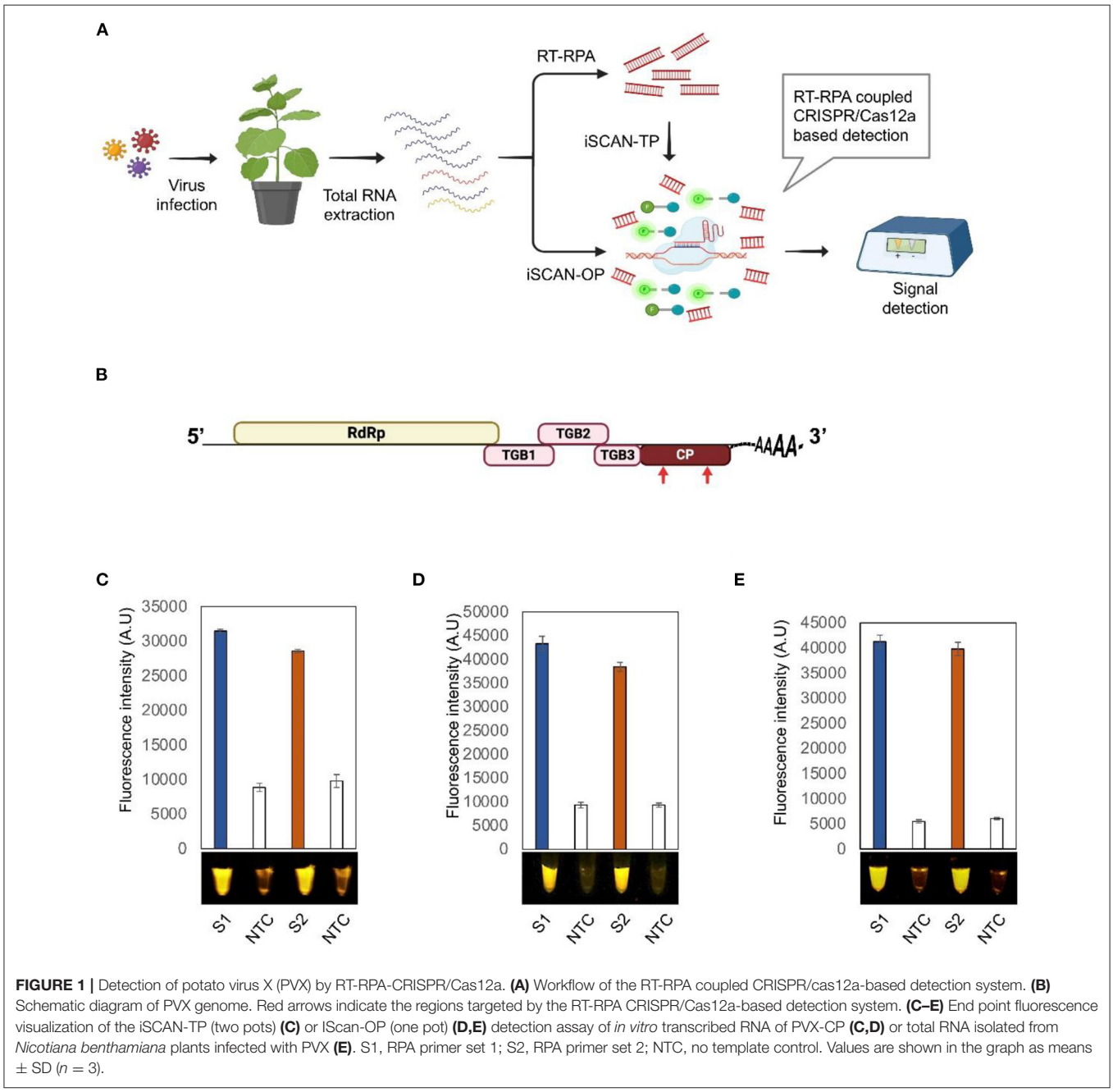


FIGURE 1 | Detection of potato virus X (PVX) by RT-RPA-CRISPR/Cas12a. **(A)** Workflow of the RT-RPA coupled CRISPR/cas12a-based detection system. **(B)** Schematic diagram of PVX genome. Red arrows indicate the regions targeted by the RT-RPA CRISPR/Cas12a-based detection system. **(C–E)** End point fluorescence visualization of the iSCAN-TP (two pots) **(C)** or IScan-OP (one pot) **(D,E)** detection assay of *in vitro* transcribed RNA of PVX-CP **(C,D)** or total RNA isolated from *Nicotiana benthamiana* plants infected with PVX **(E)**. S1, RPA primer set 1; S2, RPA primer set 2; NTC, no template control. Values are shown in the graph as means ± SD ($n = 3$).

NGS and CRISPR-Cas editing technologies were presented in a review article that discussed in detail the development, various applications, advantages and drawbacks, as well as potential future of both technologies in plant virology (Shahid et al.).

B-Detection of Plant RNA Viruses

The *in vitro* Specific CRISPR-based Assay for Nucleic acids detection (iSCAN) was combined with reverse transcription-recombinase polymerase amplification (RT- RPA), to develop a one-pot detection assay termed iSCAN-one-pot (iSCAN-OP), for specific, rapid, and sensitive detection of plant RNA viruses (Aman et al.). The RT-RPA pre-amplification step converts the RNA genome of the virus into dsDNA that serves as a substrate for the Cas12a cis activity. Cas12a targeting of the dsDNA triggers its collateral activity, which in turn cleaves the ssDNA reporter molecules and releases the signal. The reaction was incubated isothermally at 42°C for 20 min. The fluorescence signal as a result of trans-cleavage activity of Cas12a was measured with a commercially available P51 Molecular Fluorescence viewer using the Tecan plate reader (**Figure 1**). This detection method has the potential to be used for RNA plant viruses and viroids.

AUTHOR CONTRIBUTIONS

AH wrote the editorial and prepared the final version of the editorial. HC and JR reviewed the editorial and made useful suggestions. All authors approved the editorial.

ACKNOWLEDGMENTS

We are indebted to Frontiers and members of the editorial office of Frontiers in Microbiology for their complete and unwavering support: Ander Baranda, Giacomo Cecot, Margaux Dreyer, Helene Jensen, Pauline Maturana, Lydia Nurse, and Camilla Stanton. We are also grateful to colleagues from many countries for their valuable review of manuscripts submitted to this Research Topic: Nicolas Bejerman, Thierry Candresse, Fabrizio Cillo, Jose-Antonio Daros, Francesco Di Serio, Elvira Fiallo-Olive, Donato Gallitelli, Hernan Garcia-Ruiz, John Hammond, Wilhelm Jelkmann, C. M. Lawrence, Magdy Mahfouz, Beatriz Navarro, Peter Palukaitis, Michael N. Pearson, Pedro L. Ramos-Gonzales, Pasquale Saldarelli, Teruo Sano, Gerhard Steger, Satyanarayana Tatineni, Jeremy R. Thompson, Herve Vanderschuren, Anupam Varma, Aiming Wang, Jin Wang, Xiao-Wei Wang, Xuefeng Wang, Yi Xu, Zhixiang Zhang, and Kaijun Zhao.

REFERENCES

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA- targeting CRISPR effector. *Science* 353:aaf5573. doi: 10.1126/science.aaf5573

Barba, M., Czosnek, H., and Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106–136. doi: 10.3390/v6010106

Dounda, J. A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096. doi: 10.1126/science.1258096

Fontenele, R. S., Alves-Freitas, D. M. T., Silva, P. I. T., Foresti, J., Silva, P. R., Godinho, M. T., et al. (2018). Discovery of the first maize-infecting mastrevirus in the Americas using a victor-enabled metagenomics approach. *Arch. Virol.* 163, 263–267. doi: 10.1007/s00705-017-3571-2

Hadidi, A. (2019). Next-generation sequencing and CRISPR/Cas13 editing in viroid research and molecular diagnosis. *Viruses* 11:120. doi: 10.3390/v11020120

Hadidi, A., and Barba, M. (2012). Next-generation sequencing: historical perspective and current application in plant virology. *Petria* 22, 262–277.

Hadidi, A., Flores, R., Candresse, T., and Barba, M. (2016). Next-generation sequencing and genome editing in plant virology. *Front. Microbiol.* 7:13. doi: 10.3389/fmicb.2016.01325

Ng, T. F. F., Duffy, S., Polston, J. E., Bixby, E., Vallad, G. E., and Breitbart, M. (2011). Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLoS ONE* 6:e19050. doi: 10.1371/journal.pone.0019050

O'Connell, M. R., Oakes, B. I., Sternberg, S. H., East-Saletsky, A., Kaplan, M., and Doudna, J. A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* 516, 263–266. doi: 10.1038/nature13769

Rosario, K., Seah, Y. M., Marr, C., Varsani, A., Kraberger, S., Stainton, D., et al. (2015). Vector-enabled metagenomic (VEM) surveys using whiteflies (Aleyrodidae) reveal novel Begomovirus species in the New and Old Worlds. *Viruses* 7, 5553–5570. doi: 10.3390/v7102895

Sanfaçon, H. (2015). Plant translation factors and virus resistance. *Viruses* 7, 3392–3419. doi: 10.3390/v7072778

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109, 716–725. doi: 10.1094/PHYTO-07-18-0257-RVW

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.

Copyright © 2021 Hadidi, Czosnek and Randles. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 May 2020
doi: 10.3389/fmicb.2020.00945

Virome of *Camellia japonica*: Discovery of and Molecular Characterization of New Viruses of Different Taxa in Camellias

Song Zhang[1,2], *Liu Yang*[1,2], *Lisha Ma*[1,2], *Xin Tian*[1,2], *Ruhui Li*[3], *Changyong Zhou*[1,2] *and Mengji Cao*[1,2*†]

[1] National Citrus Engineering and Technology Research Center, Citrus Research Institute, Southwest University, Chongqing, China, [2] State Cultivation Base of Crop Stress Biology for Southern Mountainous Land, Academy of Agricultural Sciences, Southwest University, Chongqing, China, [3] USDA-ARS, National Germplasm Resources Laboratory, Beltsville, MD, United States

OPEN ACCESS

Edited by:
Ahmed Hadidi,
Agricultural Research Service (USDA), United States

Reviewed by:
Donato Gallitelli,
University of Bari Aldo Moro, Italy
Pasquale Saldarelli,
Institute for Sustainable Plant Protection (CNR), Italy

***Correspondence:**
Mengji Cao
caomengji@cric.cn

†ORCID:
Mengji Cao
orcid.org/0000-0003-0396-262X

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

Received: 15 February 2020
Accepted: 20 April 2020
Published: 15 May 2020

Citation:
Zhang S, Yang L, Ma L, Tian X, Li R, Zhou C and Cao M (2020) Virome of Camellia japonica: Discovery of and Molecular Characterization of New Viruses of Different Taxa in Camellias. Front. Microbiol. 11:945. doi: 10.3389/fmicb.2020.00945

Many species of the genus *Camellia* are native to China, and several species such as *C. japonica* have been cultivated as garden plants for over 1,000 years. Virus-like symptoms have been recorded for years. In this study, *C. japonica* plants with various leaf symptoms were observed in Jiangxi and Chongqing provinces. The species composition of potential viruses in the symptomatic plants was analyzed by next-generation sequencing of six libraries prepared from total RNAs of specimens from 10 trees. Five new viruses were discovered, and their genome sequences were determined. These viruses were tentatively named Camellia chlorotic ringspot viruses (CaCRSVs), Camellia yellow ringspot virus (CaYRSV), Camellia-associated badnavirus (CaBaV), and Camellia-associated marafivirus (CaMaV) based on comprehensive analyses. Among these viruses, CaYRSV, CaBaV, and CaMaV share similar genome organizations and clear sequence homology with known viruses in databases and could potentially be classified as new species of the genera *Badnavirus*, *Idaeovirus*, and *Marafivirus*, respectively. CaCRSVs comprise two distinct viruses, and each likely contains five genomic RNA segments that were found to be distantly related to viral RNAs of members in the genus *Emaravirus* (family *Fimoviridae*). The RNAs of CaCRSVs show conserved terminal sequences that differ markedly from those of emaraviral RNAs. These data, together with the phylogenetic analysis, suggest that the evolutionary status of CaCRSVs may represent a novel genus in the family *Fimoviridae*. In addition, two known viruses (geminivirus and blunervirus) and a mass of betaflexiviruses existing as heterogeneous mixtures were detected, and their roles in symptom formation were studied. Collectively, the information of the viral species and detection protocols that were developed can serve as a basis for better management of these viruses. Distinguishing the virus-related symptoms from genetic characteristics of *C. japonica* is also significant for breeding efforts.

Keywords: *Camellia japonica*, next-generation sequencing, virome, new viruses, RT-PCR detection, phylogenetic analysis

INTRODUCTION

Camellia spp. of the family *Theaceae* are economically important group of perennial evergreen flowering plants (Gao, 2005). This genus of approximately 280 species are native to East and Southeast Asia (Meegahakumbura et al., 2018). Most (238 species) are naturally distributed in China (http://www.iplant.cn/info/Camellia?t=z). *C. sinensis* is planted to produce popular tea beverages, while *C. japonica* (common camellia) is a well-known ornamental shrub. *C. japonica* and its hybrids are well-known ornamentals since they have large flowers of various colors and shapes, long and varied blossoming seasons and different growth habitats (Mondal, 2011). Ornamental camellias (chahua in Chinese) have been grown in China since Three Kingdoms Period (AD 220–265) and are the symbolic flowers of Chongqing and Yunnan. The common camellia was introduced to Japan where it was named Tsubaki to be distinguished from Sazannka (*C. sasanqua*), the Japanese camellia, over 1,000 years ago (Wu Y. et al., 2015). The ornamental camellias were brought to Europe and Americas in late 1870's (Bartholomew, 1986), and are now popular flowering and landscaping shrubs in many regions with mild climate in the world (Mondal, 2011). Additionally, camellias contain many bioactive compounds such as tea saponins with surface-active properties and pharmacological activities (Zhao et al., 2011).

Both biotic (fungal, bacterial, and viral diseases) and abiotic stresses affect ornamental camellias (Dickens and Cook, 1989; Taylor and Long, 2000; Zhang et al., 2014). Fungal pathogens such as those of leaf spots and gray blight are the primary concerns of camellias in China (Zhang et al., 2012; Yang S. et al., 2019), while viruses have not been well-studied, regardless of being suspected to be associated with some leaf-related diseases for decades (Milbrath and McWhorter, 1946; Gailhofer et al., 1988). The virus-like symptoms such as foliar mottle, mosaic, ringspots as well as foliar and flower variegations have been observed on *C. japonica* (Milbrath and McWhorter, 1946; Hildebrand, 1954; Ahlawat and Sardar, 1973; Gailhofer et al., 1988). These viral diseases could easily be transmitted across generations and spread between different regions by vegetative propagation (cutting and grafting) commonly used by commercial companies and individuals (Inouye, 1982). The variegation caused by the viruses may be confused with genetic variegation, which is valuable horticultural trait (Valverde et al., 2012). The putative viruses were associated with some viral diseases by biological and morphological studies (Plakidas, 1954; Inouye and Inouye, 1975; Hiruki, 1985; Gailhofer et al., 1988). With the application of next-generation sequencing (NGS) techniques, several new viruses have been recently identified from camellias with different symptoms (Hao et al., 2018; Zhang et al., 2018; Liu H. et al., 2019). However, the studies of the potentially implicated viruses are still inadequate, especially with respect to genome information.

Replication cycle of viruses with both RNA and DNA genome and viroids has an mRNA transcript and/or RNA replication stage. The enhancement during replication of their genomes inevitably increases the generation of double-stranded RNA (dsRNA), which can be degraded to the virus- or viroid-small RNAs (sRNAs) by the RNA silencing of the host plants (Ding, 2010). Therefore, the sequencing of plant total RNAs or sRNAs of the hosts is able to capture almost all sequence information of viruses and viroids in tested plant tissues (Wu Q. et al., 2015). The two sequencing techniques have some advantages and shortcomings (Pecman et al., 2017), and a combined utilization is also used in the virome analyses (Cao et al., 2019). Here, we used ribosome RNA-depleted RNA sequencing to analyze *C. japonica* plants displaying various symptoms, which allowed the identification of five new viruses, with several of them being exclusively associated with one distinct symptom based on comparative analysis.

MATERIALS AND METHODS

Plant Materials

Leaf samples of ten *C. japonica* trees, nine from Chongqing province (HC1, CRI1, CRI2, SWU1, SWU4, SWU11, SWU13, SWU14, and SWU20) and one from Jiangxi province (JX1), were collected during 2016–2018 (**Table S1**). According to similar leaf symptoms, these samples were divided into six groups designated as SC-HC (HC1; non-symptomatic), SC-JX (JX1; chlorotic ringspot), SC-CRI (CRI1, CRI2; malformation and mosaic), SC-L16 (SWU11; yellowing), SC-L17 (SWU1, SWU14; mosaic and chlorotic mottle), and SC-L18 (SWU4, SWU13, SWU20; yellow ringspot, yellow spot, yellow mottle, and yellowing) (**Figure 1**). The six sample groups were each tested by NGS. Ten grams of leaf tissues from each sample group was ground in liquid nitrogen to fine powder. One gram of the powder was used for RNA extraction, and the rest powder was stored at −80°C for future use.

RNA Extraction, NGS, and Data Processing

Total RNA was extracted using the EASY spin Plus Complex Plant RNA Kit (Aidlab, China), and then tested using the Nanodrop (Thermo Fisher Scientific, USA), Qubit 3.0 (Invitrogen, USA), and Agilent2100 (plant RNA Nano Chip, Agilent, USA) for purity, concentration, and integrity, respectively. After the removal of ribosome RNA by the Ribo-Zero Magnetic Kit (Epicenter, USA), the libraries were built using a TruSeq RNA Sample Prep Kit (Illumina, USA). An Illumina HiSeq X-ten platform (Illumina) set with length of 150-bp pair-end reads was then used for sequencing (Mega Genomics, China). Sequences of adaptor and low-quality trait were trimmed from raw reads, and the rest reads were mapped to the genome sequences of common tea (*C. sinensis*) (Wei et al., 2018), using the CLC Genomic Workbench 9.5 (Qiagen, USA). The reads with sequence similarities of >60% to the tea genome sequences were eliminated to reduce interference of the host background, and the remaining unique reads were *de novo* assembled using the Trinity program (Grabherr et al., 2013). The resulted contigs were subjected to BLASTx and BLASTn searches against viral (taxid:10239) and viroidal (taxid:2559587) sequences of local datasets retrieved from the National Center for Biotechnology Information (NCBI) databanks. These processes allowed the identification of the contigs with viral sequence attributes.

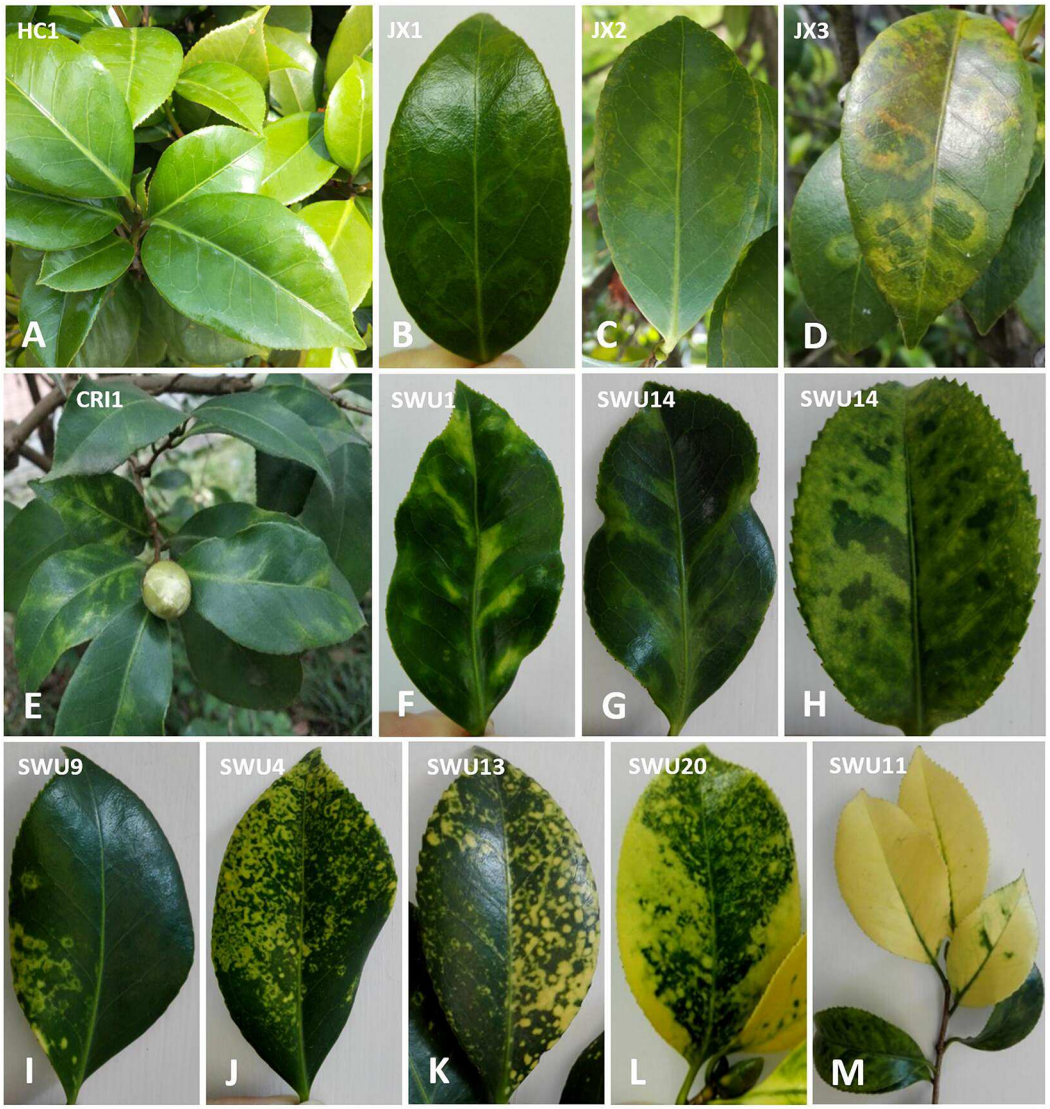


FIGURE 1 | Symptom observation of *Camellia japonica* leaves. **(A)** No obvious symptoms. **(B–D)** Chlorotic ringspot. **(E–G)** Malformation and mosaic. **(H)** Chlorotic mottle. **(I–L)** Yellow ringspot, yellow spot, yellow mottling, and yellowing. **(M)** Yellowing.

Recovery of Viral Genomes

A set of specific primers based on the viral contig sequences were designed using the Primer Premier 5 (Premier Biosoft, USA) to amplify overlapping fragments of each of the new *C. japonica* viruses (**Figure S1**). The primers are listed in **Table S2**. One-step reverse transcription-PCR (RT-PCR) assay was carried out using the PrimeScript One-Step RT-PCR Kit (Takara, Japan). Rapid amplification of cDNA ends-PCR (RACE-PCR) assay was conducted using the GeneRacer Core Kit (Invitrogen, USA). PCR assay was done with the 2 × Taq Master Mix Kit (Quick Load) (Novoprotein, China). The PCR amplicons were purified by the Gel Extraction Kit (Biomega, USA) and cloned into the pEASY-T1 Vector (TransGen, China). Sequence of each amplicon was determined from both directions of five clones by a biotechnology company (Tsingke, China). The full-length genome of each virus was assembled from all amplicons of the virus using the *de novo* assembly algorithm in SeqMan (DNAStar, USA).

Sequence Analysis and Read Assembly

Viral genome organizations were studied using the ORF finder (https://www.ncbi.nlm.nih.gov/orffinder) and the Conserved Domain Database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) websites in NCBI for opening reading frames (ORF) with a length >300-nucleotide (nt) and conserved amino acid (aa) domains with an e-value <0.05, respectively. The DRNApred (http://biomine.cs.vcu.edu/servers/DRNApred/), TMHMM (http://www.cbs.dtu.dk/services/TMHMM/), and PROMALS3D (http://prodata.swmed.edu/promals3d/promals3d.php) were used to predict DNA-binding

sites, transmembrane (TM) domains, and secondary structures of viral proteins inferred from ORFs, respectively. Nucleotide or aa sequence alignment and comparison were performed using the CLC Genomic Workbench 9.5.

A total of 9.54–11.35 G trimmed reads of six datasets were individually generated from the six independent leaf sample sets after a pipeline of data processing (**Table S1**). Subsequently, the reads (91.77–95.99%) mapped to the tea genomes as references were removed. Finally, assembly of the remaining 4.01–8.23% unique reads generated 13,583–34,687 contigs ranged from 200 to 8,789 nt in size. BLASTx analysis of the contigs using default parameters revealed the virus-related contigs that were homologous to several different taxa of viruses, including badnavirus, betaflexiviruses, blunervirus, emaravirus, geminivirus, idaeovirus, and marafivirus.

Phylogenetic Analysis

The genome (nt) or protein (aa) sequences of each of the new viruses identified by NGS and its closely related viruses retrieved from NCBI databases were aligned by the CLC Genomic Workbench 9.5. Phylogenic trees were constructed by the MEGA 7.0 (Kumar et al., 2016) using a neighbor-joining method with layouts of Jones-Taylor-Thornton (aa) or Maximum Composite Likelihood (nt, transitions + transversions) model substitution, complete deletion treatment of gaps, and 1,000 bootstrap replications.

Virome and PCR Analysis

Viral species of each sample group, RNA reads of each virus, and the proportion of viral reads in total reads were statistically analyzed. Venn diagrams were drawn using a website tool (http://bioinformatics.psb.ugent.be/webtools/Venn/). The copy number (average coverage) of viral RNA was calculated by multiplying the number of viral reads by the average length of total reads (about 150 nt) and dividing that result by the length of viral RNA.

The occurrence of viruses in 37 *C. japonica* trees (including 9 trees sequenced by NGS) from the Jiangxi and Chongqing provinces was investigated using the PCR or RT-PCR protocols (Cao et al., 2019), specific primers designed in previous studies (Hao et al., 2018; Zhang et al., 2018), and the primers designed by the DNAMAN 7 (Lynnon Biosoft, Canada) in this study (**Table S2**).

RESULTS

Identification of Viruses Infecting the Camellias

Among all the viral contigs, the betaflexivirus-related contigs accounted for 68% (59 of 87), which were detected in all the six sample groups (**Table S1**). Thus, these sequences were numerous and complicated, and the analysis below suggested that they were not associated with any observed symptoms. Therefore, the sequences of this taxon were not emphasized in the present work. We will focus on the molecular characterization of the five newly identified viruses related to badnavirus, emaravirus, idaeovirus, and marafivirus.

Two Known Camellia Viruses

The geminivirus- and blunervirus-related contigs shared more than 98% nt sequence identity with Camellia chlorotic dwarf-associated virus (CaCDaV; Zhang et al., 2018) and tea plant necrotic ring blotch virus (TPNRBV; Hao et al., 2018), respectively. These results confirmed the presence of the two viruses in *C. japonica*.

A New Monopartite Positive-Stranded RNA Virus

The monocistronic genome of the marafivirus-related virus (**Figure 2A**) is 6,878 nt long, excluding the poly (A) tails. It had the highest nt sequence identity (55.8%) to grapevine asteroid mosaic-associated virus (GAMaV, MK253012) (Vargas-Asencio et al., 2017). The 5′ untranslated region (5′ UTR, 140 nt) and 3′ UTR (123 nt) shared the highest 40 and 71.3% nt sequence identities with nectarine marafivirus M (NeVM, KT273413) (Villamor et al., 2016) and Citrus sudden death-associated virus (KY110735) (Maccheroni et al., 2005), respectively. Its genome organization is typical of marafiviruses, containing a single ORF (Igori et al., 2017). This ORF (nt 141–6,755) encodes a large putative polyprotein (2,204 aa, 242.2 kDa) consisting of a replication-associated polyprotein (RP) with a methyltransferase (Met, pfam01660, aa 145–426), a protease (Pro, cl05113, aa 840–939), a helicase (Hel, pfam01443, aa 1,030–1,262) and an RNA-dependent RNA polymerase (RdRp, cl03049, aa 1,600–1,836), and a coat protein (CP, cl03052, aa 2,029–2,188). The RP and the CP were most related to the NeVM (54.7% aa sequence identity) and the GAMaV (60.3% nt and 58.6% aa sequence identity), respectively. A 16-nt conserved nucleotide sequence stretch called "marafibox" [CA(G/A)GGUGAAUUGCUUC] (Izadpanah et al., 2002) was not found, but the RP amino acid sequences associated with the "marafibox" were partially identical to those of marafiviruses (**Figure 3A**).

Phylogenetic relationships constructed using the whole-genome sequences placed the marafivirus-related virus and maize rayado fino virus (KM523134) (Hammond and Ramirez, 2001) in a subgroup in the marafivirus group (**Figure 3B**). The results for the marafivirus-related virus satisfy the species demarcation criteria (<80% identical at whole genome sequence and <90% identical at coat protein sequence) of the genus *Marafivirus* (Dreher et al., 2011). Thus, this virus should be a new species of the genus.

A New Bipartite Positive-Stranded RNA Virus

The genome of the idaeovirus-related virus is composed of two genomic RNA components (RNA1 and RNA2) (**Figure 2B**). They shared the greatest nt sequence identities of 66.5 and 66.4% with RNA1 (KY399998) and RNA2 (KY399999) of black currant leaf chlorosis-associated virus (BCLCaV), respectively (James and Phelan, 2017). Like other idaeoviruses (Navarro et al., 2017), both RNAs start with a 5′ end tetranucleotide (AUAU), and end with 3′-terminal four stem-loop structures and cytidine (C) repeats (**Figure 3C**).

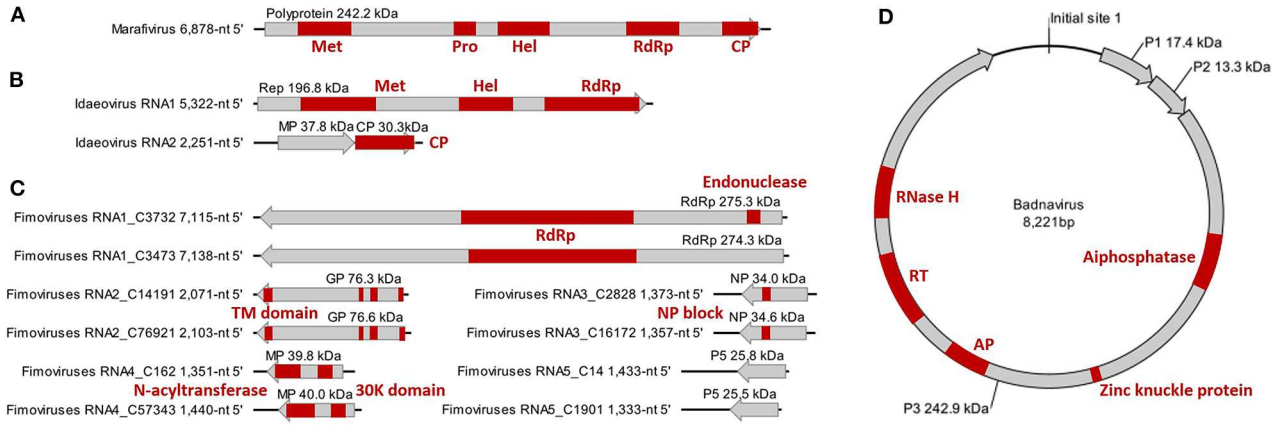


FIGURE 2 | Genome and protein features of Camellia-associated marafivirus (CaMaV) **(A)** and badnavirus (CaBaV) **(D)**, and Camellia yellow ringspot virus (CaYRSV, the idaeovirus) **(B)** and chlorotic ringspot viruses (CaCRSVs, the fimoviruses) **(C)**, with conserved aa domains or motifs indicated by red boxes. Met, methyltransferase; Pro, protease; Hel, helicase; RdRp, RNA-dependent RNA polymerase; CP, coat protein; Rep, replicase; MP, movement protein; GP, glycoprotein; TM, transmembrane; NP, nucleocapsid protein; AP, aspartate protease; RT, reverse transcriptase; RNase H, ribonuclease H.

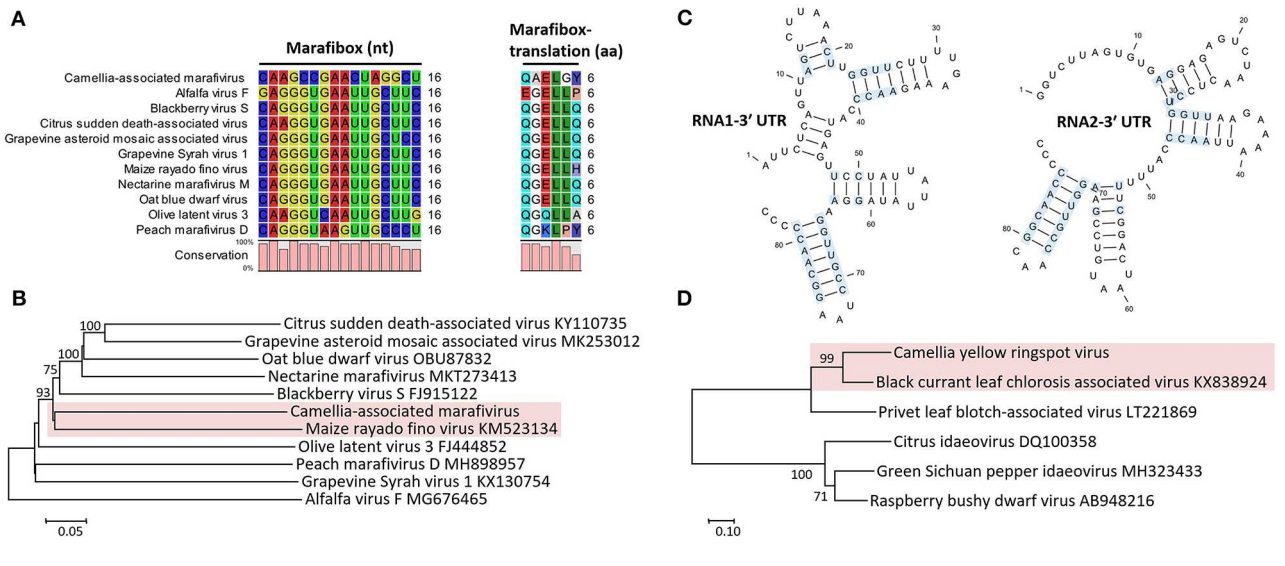


FIGURE 3 | (A) Multiple sequence alignment of Camellia-associated marafivirus (CaMaV) and representative marafiviruses at the marafibox region. **(B)** Phylogenetic analysis of CaMaV and the marafiviruses based on the whole genome sequences; bootstrap values (1,000 replications) under 50% are pruned. The CaMaV and its closest relative are indicated by the red background. **(C)** Phylogenetic relationships inferred from the conserved coat protein (CP) amino acid sequences of Camellia yellow ringspot virus (CaYRSV, the idaeovirus) and representative idaeoviruses, with CaYRSV and the closest relative backgrounded by red. Significant bootstrap values that were greater than 50% (1,000 replications) are shown. **(D)** Secondary structure of 3′ untranslated regions (3′ UTR) of genomic RNAs of CaYRSV. The blue background highlights the identical nucleotides between the stem-loops of RNA1 and RNA2.

RNA1 (5,322-nt, ORF in nt 52–5,238) encodes a putative replicase (Rep) protein (1,728 aa and 196.8 kDa) consisting of conserved Met (pfam01660, aa 195–543), Hel (pfam01443 aa 895–1,146), and RdRp (pfam00978, aa 1,276–1,709) domains. The Rep was most homologous to the BCLCaV (74.2% aa sequence identity). The 5′ and 3′ UTRs shared the highest nt sequence identities with privet leaf blotch-associated virus (54.6%) (Navarro et al., 2017) and the BCLCaV (53.6%), respectively.

RNA2 (2,251-nt) contains two ORFs with one nucleotide overlap that encode a putative movement protein (MP, 343 aa and 37.8 kDa) at nt 319–1,347 and a putative CP (cl05884, 270 aa and 30.3 kDa) at nt 1,350–2,159. It was most related to the BCLCaV at 5′ UTR, 3′ UTR, MP gene, and CP gene, for which the sequence identities shared were 54.5% (nt), 52.6% (nt), 69% (aa), and 71.2% (aa), respectively. The phylogenetic analysis based on the CP gene also suggested the closest relationship with the BCLCaV (**Figure 3D**). Given the differences with the BCLCaV in

sequence and host, the idaeovirus-related virus was deemed as a putative new species in the genus *Idaeovirus*.

Two Novel Multipartite Negative-Stranded RNA Viruses Associated With Emaraviruses

Ten contigs related to the genus *Emaravirus* (family *Fimoviridae*) were identified in the JX1 tree. The complete sequences of these RNA fragments were determined by Sanger sequencing (**Figure S1**). These ten viral RNAs could be divided into two groups according to significant aa sequence differences (25.2–57.3%) (**Table S3**) and difference of the RNA copy numbers [3-digit vs 2-digit (4 out of 5)] between the two groups (**Table S4**). Each group harbors five RNAs that encode core proteins with similarities to those of the emaraviruses, suggesting the existence of two putative fimoviruses.

The 5′ and 3′ ends were highly complementary in all the RNAs (**Figure 4A**), but a C residue that invariably occurs at the 10th nt position of the 3′ end (counting from 3′ to 5′) was exceeding and non-complementary, which was different from the emaraviruses (Mielke and Muehlbach, 2007; Mielke-Ehret and Mühlbach, 2012). For all of the RNAs, the 5′ and 3′ termini were conserved in the 11-nt (AGUAGUUWUCU, W = A/U) and 12-nt (AGCAAAACUACU), respectively (**Figure 4B**). The terminal consensuses were unique since the emaraviruses had a 13-nt consensus at each of the termini (5′- AGUAGUGUUCUCC……GGAGUUCACUACU-3′, the identical nt between the putative fimoviruses and the emaraviruses were underlined) (Mielke and Muehlbach, 2007). Furthermore, the GC content of 30% at the termini of the two putative fimoviruses was lower than average of 46% for the emaraviruses.

An AUG-initiated ORF that encodes a hypothetical protein was predicted for each of the ten RNA segments named RNA1-C3732, RNA1-C3473, RNA2-C14191, RNA2-C76921, RNA3-C2828, RNA3-C16172, RNA4-C162, RNA4-C57343, RNA5-C14, and RNA5-C1901 (**Figure 2C**). The lengths of the genomic 5′ UTRs (42- to 741-nt) and 3′ UTRs (38- to 154-nt) are variable, similar to those reported for the emaraviruses (Yang C. et al., 2019). All of the putative proteins were also related to the emaraviruses based on BLASTp analysis.

RNA1-C3732 of 7,115 nt and RNA1-C3473 of 7,138 nt contain an ORF (nt 7,050–88 for C3732; nt 7,138–92 for C3473) that encodes a putative RdRp (2,320 aa and 275.3 kDa for C3732; 2,325 aa and 274.3 kDa for C3473). The CDD search revealed a Bunya_RdRp superfamily domain (cl20265) for both the proteins and an endonuclease domain (cl20011) for the protein of the C3473 (**Figure 2C**). The aa sequences of these two proteins were 58.1% identical to one another, while only 24.1–28% identical to that of RdRps of the emaraviruses (**Table S3**).

RNA2-C14191 and RNA2-C76921 are 2,071 nt and 2,103 nt long, respectively. Their ORFs (nt 2,002–47 for C14191; nt 2,025–43 for C76921) encode putative glycoproteins (GP) of 651 aa (76.3 kDa) and 660 aa (76.6 kDa), respectively. The GP aa sequences were 16.7–20.8% identical to those of the emaraviruses and 44.3% with one another (**Table S3**). Three N-terminal TM domains and a C-terminal TM domain that were akin to those of the emaraviruses (Yang C. et al., 2019) were predicted in each of the GPs (**Figure 2C**).

RNA3-C2828 and RNA3-C16172 are 1,373 and 1,357 nt, respectively. They contain one ORF (nt 1,258–371 for C2828; nt 1,241–342 for C16172) that was predicted to encode putative nucleocapsid proteins (NP) of 295 aa (34 kDa) for C2828 and 299 aa (34.6 kDa) for C16172. An amino acid block, NXL-GXEX6PXE, conserved in the emaraviruses was identified in the two putative fimoviruses (**Figure 2C**), whereas another conserved block (NX2SX5A) was absent (Elbeaino et al., 2009). The NPs of the two putative fimoviruses shared very limited aa sequence identities of 11.6–18.3% with those of the emaraviruses and 43.7% with one another (**Table S3**).

RNA4-C162 (1,351 nt) and RNA4-C57343 (1,440 nt) have one ORF at nt 1,197–175 and nt 1,351–326, respectively. This ORF encode a putative movement protein (MP) of 340 aa (39.8 kDa) for C162 or 341 aa (40 kDa) for C57343. The 30K-MP structural signatures, including a putative catalytic Asp (D) residue and a series of alpha-helixes and beta-strands, were present based on the secondary structure analysis (**Figure S2**; Yu et al., 2013). The 30K domain was followed by an N-acyltransferase superfamily (cl17182) (**Figure 2C**). The identities of the amino acid sequences of the MP were 11.8–21.7% between the two putative fimoviruses and the emaraviruses and 74.8% between the two putative fimoviruses (**Table S3**).

RNA5-C14 of 1,433 nt and RNA5-C1901 of 1,333 nt contain a single ORF (nt 1,395–742 for C14 and nt 1,295–642 for C1901) coding for putative proteins of the same size (217 aa) with molecular weight of 28.8 and 25.5 kDa, respectively. The two proteins shared 53.9% aa sequence identify with each other, and approximately 21% aa sequence identify with the putative protein encoded by RNA7 of high plains wheat mosaic virus (KJ939630) (**Table S3**; Tatineni et al., 2014). These protein homologs may play similar roles fighting against the RNA silencing defenses of the host (Gupta et al., 2018, 2019).

The proteins encoded by RNA1–RNA4 were considered as the core elements because they are conserved for all assigned and unclassified members of the genus *Emaravirus* in the family *Fimoviridae* (Elbeaino et al., 2018). Phylogenetic analyses using the aa sequences of three of these proteins (RNA1–RNA3) all placed the two putative fimoviruses in a cluster distinct from the two subclusters formed by the emaraviruses (**Figures 4C–F**), supporting that they are new members of the family with an extraordinary evolutionary path.

Based on the facts that the two putative fimoviruses have moderate aa sequence identities (<74.8%) shared between them, unique termini at the two ends of the all five RNAs, low aa sequence identities (<28%) of their deduced proteins with the emaraviruses, and the evolutionary status representing a special clade of the family *Fimoviridae*, we propose these viruses as two putative species of a new taxon (genus) in the recently established family *Fimoviridae* (Elbeaino et al., 2018).

A New Double-Stranded Circular DNA Virus

The badnavirus-related virus has a circular DNA genome of 8,221 bp, which contains three ORFs on the plus strand

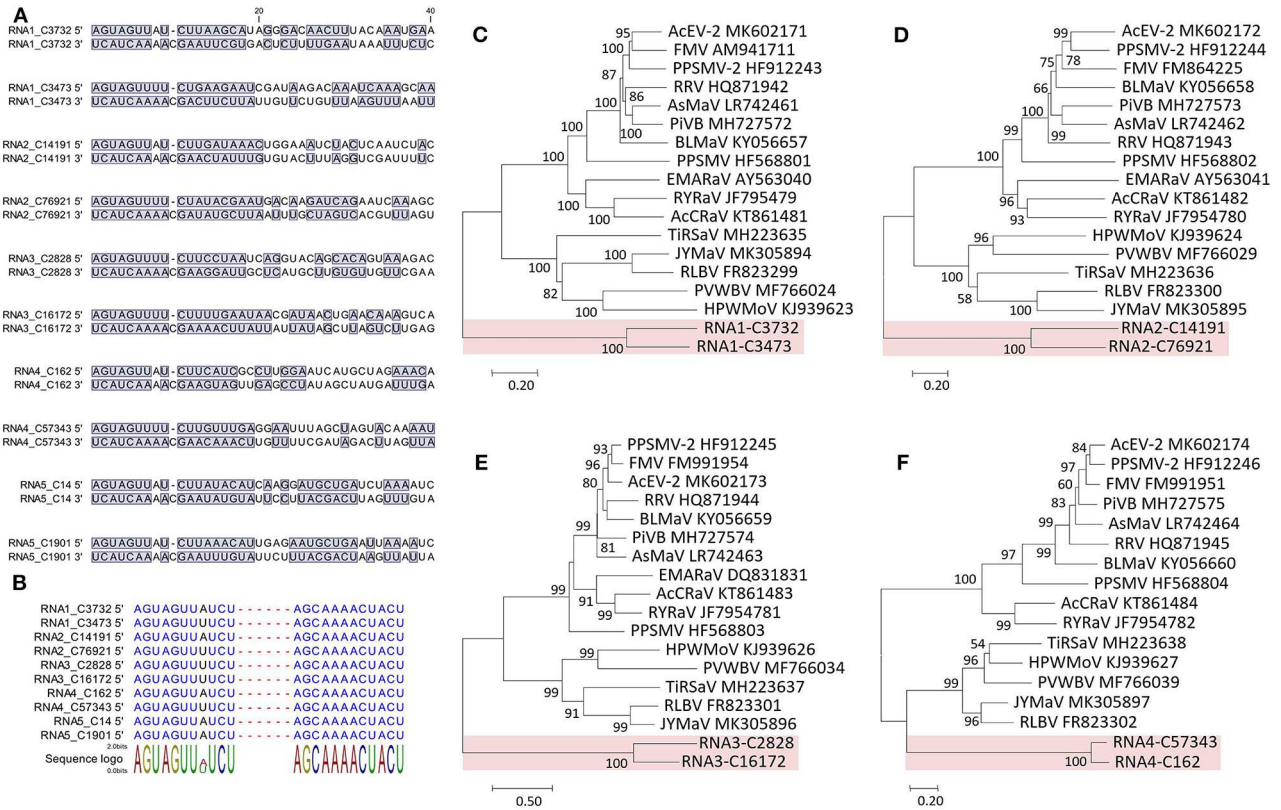


FIGURE 4 | Sequence comparison of the 5′ and 3′ genomic ends of Camellia chlorotic ringspot viruses (CaCRSVs, the fimoviruses) and the complementary nucleotides are indicated with a blue background **(A)**. Multiple sequence alignment of both genomic termini of viral RNAs of CaCRSVs, and the identical nucleotides are displayed in blue **(B)**. Evolutionary analysis of CaCRSVs and representative emaraviruses based on amino acid sequences of RdRp **(C)**, GP **(D)**, NP **(E)**, and MP **(F)** genes. The genes of CaCRSVs are indicated by the red background. Bootstrap values (1,000 replications) below 50% are not shown.

(**Figure 2D**). The RNA reads mapping analysis (**Figure S3**) showed that the mapped reads in the viral genome were overlapping and continuous, suggesting the episomal form of the virus rather than fragments integrate into host genomes. Multiple sequence comparisons at the whole genome level showed 31.3–37.4% nt identities between this virus and classified members of the genus *Badnavirus*. The genome contains the tRNAmet-binding site (TGGTATCAGAGCTTCGGC, nt 1–18), the TATA boxes (nt 109–112, 393–396, and 398–401), and the polyadenylation signal (AATAAA, nt 8,139–8,144), which resembled those of badnaviruses (Bouhida et al., 1993).

ORF1 (nt 421–870) encodes a putative protein P1 (149 aa, 17.4 kDa), which shared the highest aa sequence identity of 53.6% with the P1 of cacao swollen shoot Togo A virus (AJ781003) (Oro et al., 2012). A DUF1319 superfamily (cl06184) of unknown function that was possibly virion-associated was found in the P1 (Cheng et al., 1996).

ORF2 (nt 870–1,226) encodes a putative nucleic acid-binding protein, P2 (118 aa, 13.3 kDa) that had the highest aa sequence identity (34.3%) with Dioscorea bacilliform ES virus (KY827394) (Sukal et al., 2017). The P2 was predicted to have a DNA-binding region at aa 28–43 (Jacquot et al., 1996).

ORF3 (nt 1,223–7,777) encodes a putative polyprotein P3 (2,184 aa, 242.9 kDa). The P3 shared the highest aa sequence identity of 33.6% with that of Dioscorea bacilliform AL virus 2 (DBALV2, MH404155) (Sukal et al., 2020). The domains (**Figure 2D**) identified in the P3 include zinc knuckle protein (pfam00098, aa 999–1,016), aspartate protease (AP, cl11403, aa 1,295–1,389), reverse transcriptase (RT, cd01647, aa 1,511–1,694), and ribonuclease H (RNase H, cl14782, aa 1,793–1,921), which are typical of the genus *Badnavirus* (MacFarlane, 2011). In addition, a trimeric dUTP diphosphatase (cl00493, aa 505–640) was found in the P3.

The phylogenetic tree constructed by the whole-genome sequences of the badnavirus-related virus and representative badnaviruses grouped it with cacao mild mosaic virus (KX276640) (Chingandu et al., 2017) and sweet potato pakakuy virus (FJ560943) (Kreuze et al., 2009) in the same subcluster (**Figure 5**). Despite the close relationship with badnaviruses, the highest nt sequence identity of 66% shared between the virus and badnaviruses (DBALV2) at the regions combined with the RT and RNase H domains did not reach the species demarcation level (80%) of the genus *Badnavirus* (Geering and Hull, 2011), suggesting that the virus should be considered a new, distinct badnavirus species.

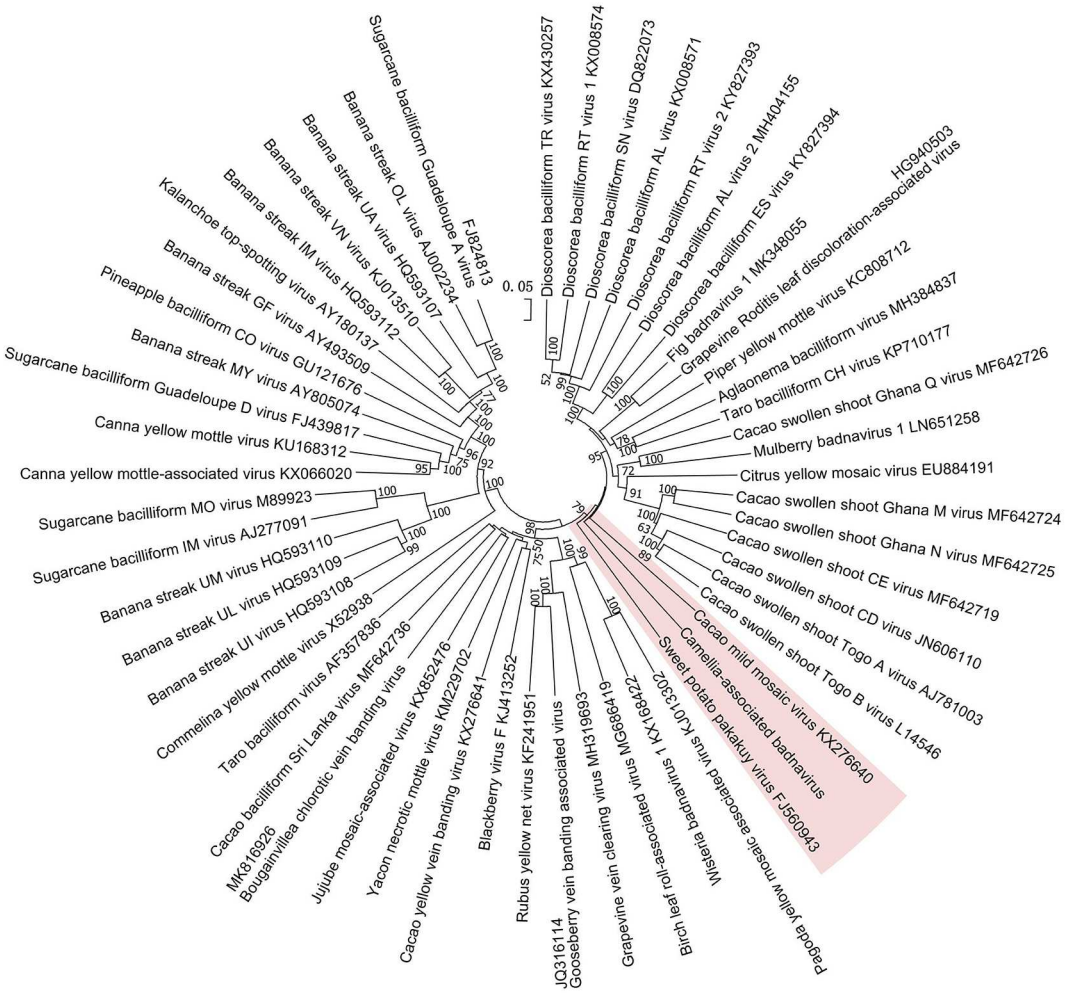


FIGURE 5 | Phylogenetic analysis of full-genome sequences of Camellia-associated badnavirus (CaBaV) and representative badnaviruses. The CaBaV and its most associated viruses are indicated by the red background. Bootstrap values (1,000 replications) are shown only if they were above 50%.

Virome and Symptomatology Analysis

The 59 betaflexivirus-like contigs, the sequences and taxonomy of which will not be discussed in this study, were categorized according to 80% (considered as single putative virus as contigs shared >80% nt sequence identity). Then, based on the BLASTn identity of each putative virus (the contigs) shared with its closet relative available in databases, they were identified as apple stem grooving virus (ASGV, KR106996; 1 contigs, 78.6%), Camellia ringspot associated virus 1 (CRSaV-1, MK050792; 8 contigs, 79–96%), CRSaV-2 (MK050793 and MK050794; 16 contigs, 80–94%), CRSaV-3_MK050795 (3 contigs, 88–97%), CRSaV-3_MK050796 (6 contigs, 92–99%), and three potential new chordoviruses indicated by chordovirus-1 (8 contigs, 68–72%), -2 (16 contigs, 68–76%) and -3 (1 contigs, 66.6%). The comparative analysis (**Figure 6**) suggested that the betaflexiviruses and the marafivirus had minor effects on the development of the different symptoms observed on *C. japonica* plants.

PCR and RT-PCR using specific primers showed that CCaDaV was found in 18 camellias exhibiting mosaic and/or malformation, the idaeovirus was detected in 13 camellias with yellowing, yellow spots or yellow ringspots, the fimoviruses were present in 8 camellias of chlorotic ringspots, while some of the symptomatic camellias might be only infected by one of these three viruses (**Table 1**). The RT-PCR assay for the fimoviruses could not distinguish them from each other since single infection of either one was not available in this study. In contrast to these viruses, the badnavirus, the marafivirus, and TPNRBV were not consistently associated with any visible symptoms.

From the perspective of a viral population (**Table S4**), the viral copy numbers of the fimoviruses were 23 times greater than that of the badnavirus in the SC-JX group. The reads of CaCDaV accounted for around 0.54% of the total reads of the SC-CRI group, which was much higher than those (<0.12%) of the other viruses. The copy number of the idaeovirus in the SC-L18 was 8 times higher than that of TPNRBV. These findings further

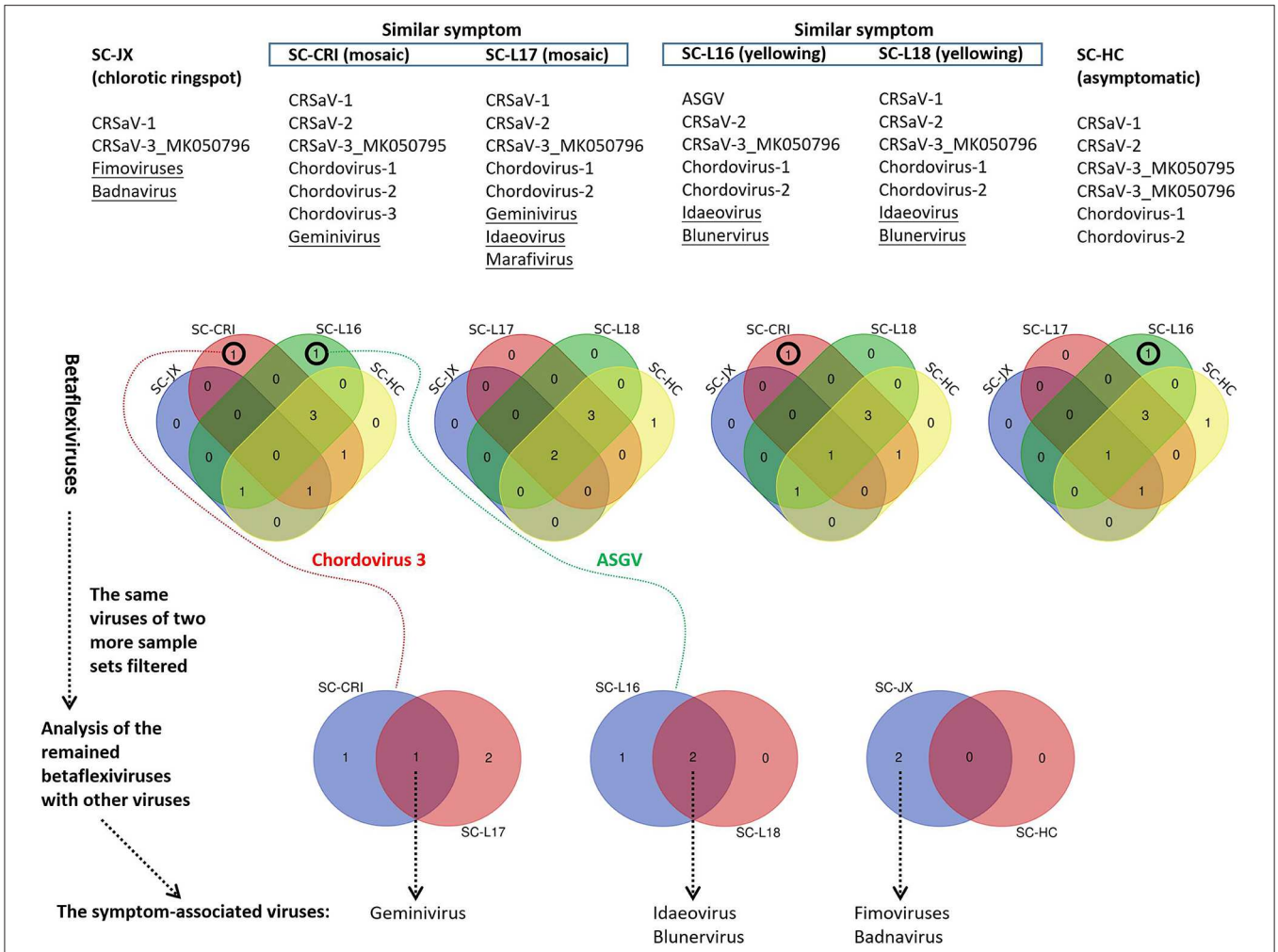


FIGURE 6 | Comparative analysis of viruses in six sample sets (SC-HC, SC-JX, SC-CRI, SC-L16, SC-L17, and SC-L18), using Venn diagrams. At first, the similar betaflexiviruses of two more sample sets showing distinct symptoms were removed. Then, the same viruses of two sample sets of similar symptoms were considered to be symptom-related. ASGV (apple stem grooving virus), CRSaV-1, -2, and -3 (Camellia ringspot associated virus 1, 2, and 3), chordovirus-1, -2, and -3 (the three potential new chordoviruses identified in this study).

suggested that the fimoviruses, the idaeovirus, and CaCDaV were associated with the symptom expressions in the host trees.

Based on the collective analysis of the presented data, the new viruses were provisionally named Camellia chlorotic ringspot viruses (CaCRSVs, the fimoviruses), Camellia yellow ringspot virus (CaYRSV, the idaeovirus), Camellia-associated badnavirus (CaBaV), and Camellia-associated marafivirus (CaMaV).

DISCUSSION

The foliar symptoms that were observed on *C. japonica* in this study resemble those that have been previously reported (Milbrath and McWhorter, 1946; Hildebrand, 1954; Ahlawat and Sardar, 1973; Gailhofer et al., 1988), but were more variable and complicated, especially the ringspot-associated symptoms occurring on either the same or different trees. These were yellow and chlorotic ringspots or spots with a diameter reaching the millimeter or centimeter levels (**Figures 1B–D,I–L**). For each type of the symptoms, it is important to explore how many viruses may be involved in development of the symptom and whether the culprit of each symptom is a sole virus or multiple viruses. To address these issues, samples from the *C. japonica* plants displaying different symptoms were analyzed by NGS coupled with homology-based method using BLAST programs which have been widely utilized for virus discovery (Wu Q. et al., 2015).

The NGS techniques have been used to explode in the discovery of new viral species associated with plant diseases (Adams et al., 2009; Hadidi et al., 2016). Experimental evidence from metagenomics based on NGS has revealed the natural biodiversity of plant viruses (Roossinck, 2011; Roossinck et al., 2015). A cryptic virus kingdom has yet to be explored since the research emphasis still largely focuses on the cultivated crops (Khoshbakht and Hammer, 2008), beyond which there

TABLE 1 | PCR and RT-PCR analysis of viruses (except betaflexiviruses) in *C. japonica* trees collected in three independent locations from the Jiangxi and Chongqing provinces.

Tree	Symptom	CaCRSVs[g]	CaCRSVs[h]	CaCDaV	CaYRSV	TPNRBV	CaBaV	CaMaV
JX1[a]	Chlorotic ringspot	+[d]	+	–[e]	–	–	+	–
JX2[f]	Chlorotic ringspot	+	+	–	–	–	–	–
JX3	Chlorotic ringspot	+	+	–	–	–	–	–
JX4	Chlorotic ringspot	+	+	–	–	–	+	–
JX5	Chlorotic ringspot	+	+	–	–	–	–	–
JX6	Chlorotic ringspot	+	+	–	–	–	–	–
JX7	Chlorotic ringspot	+	+	–	–	–	–	–
JX8	Chlorotic ringspot	+	+	–	–	–	–	–
JX9	Non-symptomatic	–	–	–	–	–	–	–
CRI1[b]	Malformation, mosaic	–	–	+	–	–	–	–
CRI2	Malformation, mosaic	–	–	+	–	–	–	–
CRI3	Asymptomatic	–	–	+	–	–	–	–
SWU1[c]	Malformation, mosaic	–	–	+	–	–	–	+
SWU2	Yellowing, mosaic	–	–	+	+	–	–	–
SWU3	Mosaic	–	–	+	+	–	–	–
SWU4	Yellow ringspot	–	–	–	+	+	–	–
SWU5	Yellowing, malformation	–	–	+	+	+	–	–
SWU6	Malformation, mosaic	–	–	+	–	+	–	–
SWU7	Malformation, mosaic	–	–	+	–	–	–	–
SWU8	Yellowing, malformation	–	–	+	+	+	–	–
SWU9	Yellow ringspot, malformation	–	–	+	+	+	–	–
SWU10	Malformation	–	–	+	+	+	–	–
SWU11	Yellowing	–	–	–	+	+	–	–
SWU12	Yellow spot	–	–	+	+	+	–	–
SWU13	Yellow ringspot	–	–	–	+	+	–	–
SWU14	Mosaic, chlorotic mottle	–	–	+	+	–	–	–
SWU15	Malformation	–	–	+	–	+	–	–
SWU16	Malformation	–	–	+	–	–	–	–
SWU17	Yellow ringspot, malformation	–	–	+	+	–	–	–
SWU18	Asymptomatic	–	–	+	+	–	–	–
SWU19	Malformation, mosaic	–	–	+	+	–	–	–
SWU20	Yellow ringspot	–	–	–	+	–	–	–
SWU21	Yellow ringspot	–	–	+	+	–	–	–
SWU22	Yellow ringspot	–	–	+	+	–	–	–
SWU23	Asymptomatic	–	–	+	–	–	–	–
SWU24	Yellow ringspot, malformation	–	–	+	+	–	–	–
SWU25	Malformation	–	–	+	+	–	–	–

[a] *JX trees from the Jiangxi (JX) province.*
[b] *CRI trees from Citrus Research Institute (CRI) in Chongqing province.*
[c] *SWU trees from Southwest University (SWU) in Chongqing province.*
[d] *Positive to a virus (+).*
[e] *Negative to a virus (–).*
[f] *Trees underlined showing a strong correlation of symptoms to a virus.*
[g] *Detection of RNA1-C3732.*
[h] *Detection of RNA1-C3473.*

are plentiful plant species distributed over the world (Pimm and Joppa, 2015). In this study, the NGS analyses of the ornamental camellias revealed the presence of the viruses related to the genus *Idaeovirus* and the families *Betaflexiviridae*, *Caulimoviridae*, *Fimoviridae*, *Geminiviridae*, *Kitaviridae*, and *Tymoviridae* (Adams et al., 2011; Dreher et al., 2011; Geering and Hull, 2011; MacFarlane, 2011; Zerbini et al., 2017; Elbeaino et al., 2018; Walker et al., 2019). The identification of ASGV and TPNRBV which are the known viruses infecting other economically important crops (Hao et al., 2018; Liu Q. et al., 2019) hinted at the potential roles of the infected *C. japonica* trees as viral reservoirs. Based on informatic analyses of the genomic features and phylogeny, the five new viruses were proposed to be new members of the demarcated taxa or even of a novel

taxon (CCRaVs). These data indicated a rich diversity of viruses infecting the *C. japonica* plants.

The new fimoviruses (CCRaVs) infecting the camellias were validated to be consistent in genome architectures with the related emaraviruses infecting other plant species. The genomic RNA components of CCRaVs are likely to be at least pentapartite. RNA recombination, reassortment, and gene duplication that increase sequence variation or genome segmentation would contribute to the uncertainty in the acquisition of definite full genomes of emaraviruses (Tatineni et al., 2014; Di Bello et al., 2015; Lu et al., 2015; Patil et al., 2017; Yang C. et al., 2019). Until recently, two novel RNA segments of an emaravirus, European mountain ash ringspot-associated virus, were sequenced, in additional to the four known genomic RNAs (von Bargen et al., 2019). For CCRaVs, it is possible the additional viral RNA segments that are highly divergent from sequences of the available emaraviruses are present, and thus, they are undetectable in database-backed homology annotation.

Like other woody plants, ornamental camellias are connaturally perennial, which facilitates virus-plant symbiosis and symbiogenesis (Roossinck, 2008). In parallel to being limited to a single plant, viruses are capable of being transmitted from one plant to another in nature through vegetative propagation and vectors that play vital roles in the long-distance virus movement and increase the influence of the environmental changes exerted upon the course of virus diversification (Elena et al., 2014; Lefeuvre et al., 2019). Most of the extant viruses related to those identified from camellias in this study have specific vectors for their dispersals. For instance, badnaviruses, betaflexiviruses, and marafiviruses are transmitted by insect species infesting plants (Adams et al., 2011; Dreher et al., 2011; Bhat et al., 2016). Some phytophagous mites also contribute to the spread of emaraviruses and blunerviruses (Tatineni et al., 2014; Walker et al., 2019). For idaeoviruses, pollens transmission might be an effective way to transmit them (Bulger et al., 1990).

It is interesting that betaflexiviruses were detected in all the six sequenced sample groups independently of the geographic locations, while other viruses were not, suggesting there might be a long-term co-evolution of the betaflexiviruses with *C. japonica*. Given the possibility of natural transmission of these viruses, it is still not known whether *C. japonica* or other plants are the original hosts. Moreover, the global transport of massive plant materials and the unknown state of the plants carrying viruses make it difficult to trace the time when viruses of the source plant began to spread to native species or native viruses began to infect these plants.

Infections of multiple viruses in the same plant may not cause pathogenic effects on plants (Büttner et al., 2015), probably due to the balance kept among viruses or the convergent evolution of viruses toward mild interactions with the host (Roossinck, 2008). Otherwise, viral sequence variation, host genetic background, and environments all play generic roles in symptom development of plant viruses. These could be invoked to explain why the HC1 sample infected with many betaflexiviruses was asymptomatic, whereas the camellias infected with some of them showed some ringspot symptoms (Liu H. et al., 2019). It is also reasonable to characterize CaCDaV, CaCRSVs, and CaYRSV as symptom-associated, whereas TPNRBV, CaBaV, and CaMaV are not. However, the investigations must be widened to more regions and infected plants for confirmation.

In a natural setting, mixed infection of different viruses in a plant is the rule rather than the exception (Elena et al., 2014). To create a variety of flowers in a single *C. japonica* plant, scions of different origins are usually grafted by gardeners onto the same tree. This horticultural practice, as opposed to natural events, is considered as one of the sources for the viral coinfections. With respect to a strict Koch's rule required to support the opinion, transmission trials of the viruses infecting camellias are on the way.

In conclusion, this virome analysis of the *C. japonica* trees provides basic information of some viruses associated with the symptoms observed in this study and for evaluating the potential risk and management of the known or new viruses. The symptoms sporadically distributed among branches or/and plants were virus-related rather than derived from the *C. japonica* plants themselves.

DATA AVAILABILITY STATEMENT

The sequence information generated in this study can be retrieved from the NCBI database GenBank at CaMaV (MT036048), CaBaV (MT036049), CaCRSVs (MT040095–MT040104), and CaYRSV (MT036046 and MT036047).

AUTHOR CONTRIBUTIONS

MC conceived and designed the experiments. SZ, LY, and XT collected the samples and conducted the experiments. MC, LM, and SZ analyzed data. CZ, MC, RL, and SZ discussed the results and drafted and revised the manuscript. All authors read and approved the final draft of the manuscript.

FUNDING

This research was supported by the National Key R&D Program of China (2019YFD1001800), Fundamental Research Funds for the Central Universities (XDJK2018AA002 and XDJK2020D015), Chongqing Research Program of Basic Research and Frontier Technology (cstc2017jcyjBX0016), and Innovation Program for Chongqing's Overseas Returnees (cx2019013) and 111 Project (B18044).

ACKNOWLEDGMENTS

We thank the editor and two reviewers for their constructive comments and suggestions. We 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.2020.00945/full#supplementary-material

Figure S1 | The sequencing strategies of the viral genomes. Black, red, and blue lines represent the viral full-length genome, contigs, and overlapping fragments amplified by PCR and RT-PCR, respectively. The contigs of the idaeovirus (CaYRSV) were derived from SC-L16, SC-L17, and SC-L18 sample sets.

Figure S2 | Amino acid sequence alignment of two putative movement proteins (MP) of the fimoviruses (CaCRSVs) using the PROMALS3D. The 30K-MP domain is indicated by green bars. Alpha-helixes and beta-strands are indicated by red and blue backgrounds, respectively. The conserved catalytic amino acid residue Asp (D) appears in the red box.

Figure S3 | The RNA reads (2,128) mapped in the viral genome of the badnavirus (CaBaV). The highest read was 85. The gray arrows represent ORFs.

Table S1 | Next-generation sequencing analysis of six sample sets of 10 specimens collected from *C. japonica* plants grown in the Jiangxi and Chongqing provinces.

Table S2 | List of primers used in this study.

Table S3 | Pairwise aa sequence identity (%) shared between the fimoviruses (CaCRSVs) and representative emaraviruses at proteins encoded by RNA1 to RNA5. [a]CaCRSVs, Camellia chlorotic ringspot viruses; AcEV-2, Actinidia emaravirus 2; AcCRaV, Actinidia chlorosis ringspot-associated virus; AsMaV, aspen mosaic-associated virus; BLMaV, blackberry leaf mottle associated virus; EMARaV, European mountain ash ringspot-associated virus; FMV, fig mosaic virus; HPWMoV, high plains wheat mosaic virus; JYMaV, jujube yellow mottle-associated virus; PiVB, Pistacia emaravirus; PPSMV and PPSMV-2, pigeonpea sterility mosaic virus and -2; PVWBV, palo verde witches broom virus; RLBV, raspberry leaf blotch virus; RRV, rose rosette virus; RYRaV, redbud yellow ringspot-associated virus; TiRSaV, ti ringspot-associated virus. [b]na, not available.

Table S4 | Virome analysis of the viruses (except the betaflexiviruses) identified from the symptomatic *C. japonica* plants that were sequenced.

REFERENCES

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

Adams, M. J., Candresse, T., Hammond, J., Kreuze, J. F., Martelli, G. P., Namba, S., et al. (2011). "Family betaflxiviridae," 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), 920–941.

Ahlawat, Y. S., and Sardar, K. K. (1973). Insect and dodder transmission of tea rose yellow mosaic virus. *Curr. Sci.* 42:181.

Bartholomew, B. (1986). The Chinese species of *Camellia* in cultivation. *Arnoldia* 46, 3–15.

Bhat, A. I., Hohn, T., and Selvarajan, R. (2016). Badnaviruses: the current global scenario. *Viruses* 8:177. doi: 10.3390/v8060177

Bouhida, M., Lockhart, B. E. L., and Olszewski, N. E. (1993). An analysis of the complete sequence of a sugarcane bacilliform virus genome infectious to banana and rice. *J. Gen. Virol.* 74, 15–22. doi: 10.1099/0022-1317-74-1-15

Bulger, M. A., Stace-Smith, R., and Martin, R. R. (1990). Transmission and field spread of raspberry bushy dwarf virus. *Plant Dis.* 74, 514–517. doi: 10.1094/PD-74-0514

Büttner, C., Bargen, S. V., and Bandte, M. (2015). "Phytopathogenic viruses," in *Principles of Plant-Microbe Interactions,* eds B. Lugtenberg (New York, NY: Springer International Publishing), 115–122. doi: 10.1007/978-3-319-08575-3_13

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

Cheng, C.-P., Lockhart, B. E. L., and Olszewski, N. E., (1996). The ORF I and II proteins of commelina yellow mottle virus are virion-associated. *Virology* 223, 263–271. doi: 10.1006/viro.1996.0478

Chingandu, N., Sreenivasan, T. N., Surujdeo-Maharaj, S., Umaharan, P., Gutierrez, O. A., and Brown, J. K. (2017). Molecular characterization of previously elusive badnaviruses associated with symptomatic cacao in the New World. *Arch. Virol.* 162, 1363–1371. doi: 10.1007/s00705-017-3235-2

Di Bello, P. L., Ho, T., and Tzanetakis, I. E. (2015). The evolution of emaraviruses is becoming more complex: seven segments identified in the causal agent of Rose rosette disease. *Virus Res.* 210, 241–244. doi: 10.1016/j.virusres.2015.08.009

Dickens, J. S. W., and Cook, R. T. A. (1989). *Glomerella cingulata* on camellia. *Plant Pathol.* 38, 75–85. doi: 10.1111/j.1365-3059.1989.tb01430.x

Ding, S.-W. (2010). RNA-based antiviral immunity. *Nat. Rev. Immunol.*10, 632–644. doi: 10.1038/nri2824

Dreher, T. W., Edwards, M. C., Gibbs, A. J., Haenni, A.-L., Hammond, R. W., Jupin, I., et al. (2011). "Family *Tymoviridae*," 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), 944–952. doi: 10.1016/B978-0-12-384684-6.00080-X

Elbeaino, T., Digiaro, M., and Martelli, G. P. (2009). Complete nucleotide sequence of four RNA segments of fig mosaic virus. *Arch. Virol.* 154, 1719–1727. doi: 10.1007/s00705-009-0509-3

Elbeaino, T., Digiaro, M., Mielke-Ehret, N., Muehlbach, H.-P., and Martelli, G. P. (2018). ICTV virus taxonomy profile: *Fimoviridae*. *J. Gen. Virol.* 99, 1478–1479. doi: 10.1099/jgv.0.001143

Elena, S. F., Fraile, A., and García-Arenal, F. (2014). "Evolution and emergence of plant viruses," in *Advances in Virus Research,* eds K. Maramorosch and F. A. Murphy (San Diego, CA: Elsevier Academic Press), 161–191. doi: 10.1016/B978-0-12-800098-4.00003-9

Gailhofer, M., Thaler, I., and Milicic, D. (1988). Occurrence of camellia leaf yellow mottle virus (CLYMV) on east adriatic coast. *Acta Hort.* 234, 385–392. doi: 10.17660/ActaHortic.1988.234.46

Gao, J. (2005). *Collected Species of the Genus Camellia-An Illustrated Outline*. Hangzhou: Zhejiang Science and Technology Publishing House.

Geering, A. D. W., and Hull, R. (2011). "Family caulimoviridae," 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), 429–443. doi: 10.1016/B978-0-12-384684-6.00040-9

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

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

Gupta, A. K., Hein, G. L., and Tatineni, S. (2019). P7 and P8 proteins of high plains wheat mosaic virus, a negative-strand RNA virus, employ distinct mechanisms of RNA silencing suppression. *Virology* 535, 20–31. doi: 10.1016/j.virol.2019.06.011

Hadidi, A., Flores, R., Candresse, T., and Barba, M. (2016). Next-generation sequencing and genome editing in plant virology. *Front. Microbiol.* 7:1325. doi: 10.3389/fmicb.2016.01325

Hammond, R. W., and Ramirez, P. (2001). Molecular characterization of the genome of maize rayado fino virus, the type member of the genus *Marafivirus*. *Virology* 282, 338–347. doi: 10.1006/viro.2001.0859

Hao, X., Zhang, W., Zhao, F., Liu, Y., Qian, W., Wang, Y., et al. (2018). Discovery of plant viruses from tea plant (*Camellia sinensis* (l.) o. kuntze) by metagenomic *sequencing. Front. Microbiol.* 9:2175. doi: 10.3389/fmicb.2018.02175

Hildebrand, E. M. (1954). Camellia variegation in texas. *Plant Dis. Rep.* 38, 566–567.

Hiruki, C. (1985). A preliminary study on infectious variegation of camellia. *Acta Hort.* 164, 55–62. doi: 10.17660/ActaHortic.1985.164.5

Igori, D., Lim, S., Baek, D., Kim, S. Y., Seo, E., Cho, I.-S., et al. (2017). Complete nucleotide sequence and genome organization of peach virus D, a putative new member of the genus *Marafivirus*. *Arch. Virol.* 162, 1769–1772. doi: 10.1007/s00705-017-3255-y

Inouye, T. (1982). Graft transmission of virus-like symptoms of camellias. *Ann. Phytopathol. Soc. Jpn.* 48:117.

Inouye, T., and Inouye, N. (1975). Rod-shaped particles found in *Camellia* leaves with necrotic ring spots. *Ann. Phytopathol. Soc. Jpn.* 40:133.

Izadpanah, K., Yun, P. Z., Daubert, S., Masumi, M., and Rowhani, A. (2002). Sequence of the coat protein gene of Bermuda grass etched-line virus, and of the adjacent 'marafibox' motif. *Virus Genes* 24, 131–134. doi: 10.1023/A:1014516515454

Jacquot, E., Hagen, L. S., Jacquemond, M., and Yot, P. (1996). The open reading frame 2 product of cacao swollen shoot badnavirus is a nucleic acid-binding protein. *Virology* 225, 191–195. doi: 10.1006/viro.1996.0587

James, D., and Phelan, J. (2017). Complete genome sequence and analysis of blackcurrant leaf chlorosis associated virus, a new member of the genus *Idaeovirus*. *Arch. Virol.* 162, 1705–1709. doi: 10.1007/s00705-017-3257-9

Khoshbakht, K., and Hammer, K. (2008). How many plant species are cultivated? *Genet. Resour. Crop Evol.* 55, 925–928. doi: 10.1007/s10722-008-9368-0

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., et al. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1–7. doi: 10.1016/j.virol.2009.03.024

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

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. doi: 10.1038/s41579-019-0232-3

Liu, H., Wu, L., Zheng, L., Cao, M., and Li, R. (2019). Characterization of three new viruses of the family *Betaflexiviridae* associated with camellia ringspot disease. *Virus Res.* 272:197668. doi: 10.1016/j.virusres.2019.197668

Liu, Q., Xuan, Z., Wu, J., Qiu, Y., Li, M., Zhang, S., et al. (2019). Loquat is a new natural host of apple stem grooving virus and apple chlorotic leaf spot virus in China. *Plant Dis.* 103. doi: 10.1094/PDIS-04-19-0721-PDN

Lu, Y., McGavin, W., Cock, P. J., Schnettler, E., Yan, F., Chen, J., et al. (2015). Newly identified RNAs of raspberry leaf blotch virus encoding a related group of proteins. *J. Gen. Virol.* 96, 3432–3439. doi: 10.1099/jgv.0.000277

Maccheroni, W., Alegria, M. C., Greggio, C. C., Piazza, J. P., Kamla, R. F., Zacharias, P. R. A., et al. (2005). Identification and genomic characterization of a new virus (*Tymoviridae* family) associated with citrus sudden death disease. *J. Gen. Virol.* 79, 3028–3037. doi: 10.1128/JVI.79.5.3028-3037.2005

MacFarlane, S. A. (2011). "Genus idaeovirus," 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), 1073–1075.

Meegahakumbura, M. K., Wambulwa, M. C., Li, M.-M., Thapa, K. K., Sun, Y.-S., Moller, M. et al. (2018). Domestication origin and breeding history of the tea plant (*Camellia sinensis*) in China and India based on nuclear microsatellites and cpDNA sequence data. *Front. Plant Sci.* 8:2270. doi: 10.3389/fpls.2017.02270

Mielke, N., and Muehlbach, H.-P. (2007). A novel, multipartite, negative-strand RNA virus is associated with the ringspot disease of European mountain ash (*Sorbus aucuparia* L.). *J. Gen. Virol.* 88, 1337–1346. doi: 10.1099/vir.0.82715-0

Mielke-Ehret, N., and Mühlbach, H.-P. (2012). *Emaravirus*: a novel genus of multipartite, negative strand RNA plant viruses. *Viruses* 4, 1515–1536. doi: 10.3390/v4091515

Milbrath, J. A., and McWhorter, F. P. (1946). Yellow mottle leaf, a virus disease of camellia. *Am. Camellia Soc. Yearb.* 51–53.

Mondal, T. K. (2011). "Camellia," in *Wild Crop Relatives: Genomic and Breeding Resources,* ed C. Kole (Heidelberg: Springer Press), 15–39. doi: 10.1007/978-3-642-21201-7_2

Navarro, B., Loconsole, G., Giampetruzzi, A., Aboughanem-Sabanadzovic, N., Ragozzino, A., Ragozzino, E., et al. (2017). Identification and characterization of privet leaf blotch-associated virus, a novel idaeovirus. *Mol. Plant Pathol.* 18, 925–936. doi: 10.1111/mpp.12450

Oro, F., Mississo, E., Okassa, M., Guilhaumon, C., Fenouillet, C., Cilas, C., et al. (2012). Geographical differentiation of the molecular diversity of cacao swollen shoot virus in Togo. *Arch. Virol.* 157, 509–514. doi: 10.1007/s00705-011-1158-x

Patil, B. L., Dangwal, M., and Mishra, R. (2017). Variability of emaravirus species associated with sterility mosaic disease of pigeonpea in India provides evidence of segment reassortment. *Viruses* 9:183. doi: 10.3390/v9070183

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

Pimm, S. L., and Joppa, L. N. (2015). How many plant species are there, where are they, and at what rate are they going extinct? *Ann. Mo. Bot. Gard.* 100, 170–176. doi: 10.3417/2012018

Plakidas, A. G. (1954). Leaf and flower variegation in camellias by grafting. *Phytopathology* 44, 14–18.

Roossinck, M. J. (2011). The big unknown: plant virus biodiversity. *Curr. Opin. Virol.* 1, 63–67. doi: 10.1016/j.coviro.2011.05.022

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

Roossinck, M. J. *(eds.).* (2008). *Plant Virus Evolution*. Berlin: Springer Science & Business Media. doi: 10.1007/978-3-540-75763-4

Sukal, A., Kidanemariam, D., Dale, J., James, A., and Harding, R. (2017). Characterization of badnaviruses infecting *Dioscorea* spp. in the Pacific reveals two putative novel species and the first report of dioscorea bacilliform RT virus 2. *Virus Res.* 238, 29–34. doi: 10.1016/j.virusres.2017.05.027

Sukal, A. C., Kidanemariam, D. B., Dale, J. L., Harding, R. M., and James, A. P., (2020). Characterization and genetic diversity of dioscorea bacilliform viruses present in a Pacific yam germplasm collection. *Plant Pathol.* 69, 576–584. doi: 10.1111/ppa.13133

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

Taylor, C. H., and Long, P. G. (2000). Review of literature on camellia flower blight caused by *Ciborinia camelliae*. *N. Z. J. Crop and Hort. Sci.* 28, 123–138. doi: 10.1080/01140671.2000.9514132

Valverde, R. A., Sabanadzovic, S., and Hammond, J. (2012). Viruses that enhance the aesthetics of some ornamental plants: beauty or beast? *Plant Dis.* 96, 600–611. doi: 10.1094/PDIS-11-11-0928-FE

Vargas-Asencio, J., Wojciechowska, K., Baskerville, M., Gomez, A. L., Perry, K. L., and Thompson, J. R. (2017). The complete nucleotide sequence and genomic characterization of grapevine asteroid mosaic associated virus. *Virus Res.* 227, 82–87. doi: 10.1016/j.virusres.2016.10.001

Villamor, D. E. V., Mekuria, T. A., Pillai, S. S., and Eastwell, K. C. (2016). High-throughput sequencing identifies novel viruses in nectarine: insights to the etiology of stem-pitting disease. *Phytopathology* 106, 519–527. doi: 10.1094/PHYTO-07-15-0168-R

von Bargen, S., Dieckmann, H.-L., Candresse, T., Mühlbach, H.-P., Roßbach, J., and Büttner, C. (2019). Determination of the complete genome sequence of European mountain ash ringspot-associated emaravirus from *Sorbus intermedia* reveals two additional genome segments. *Arch. Virol.* 164, 1937–1941. doi: 10.1007/s00705-019-04275-0

Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Dempsey, D. M., Dutilh, B. E., et al. (2019). Changes to virus taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2019). *Arch. Virol.* 164, 2417–2429. doi: 10.1007/s00705-019-04306-w

Wei, C., Yang, H., Wang, S., Zhao, J., Liu, C., Gao, L., et al. (2018). Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4151–E4158. doi: 10.1073/pnas.1719622115

Wu, Q., Ding, S.-W., Zhang, Y., and Zhu, S. (2015). Identification of viruses and viroids by next-generation sequencing and homology-dependent and

homology-independent algorithms. *Ann. Rev. Phytopathol.* 53, 425–444. doi: 10.1146/annurev-phyto-080614-120030

Wu, Y., Ning, Z., Tian, B., Wang, Y., and Zhen, L. (2015). Origin and differences of *Camellia* between China and Japan. *World Forestry Res.* 28, 81–84. Available online at: http://en.cnki.com.cn/Article_en/CJFDTotal-SJLY201504014.htm

Yang, C., Zhang, S., Tang, T., Fu, J., Di Serio, F., and Cao, M. (2019). Identification and characterization of a novel emaravirus associated with jujube (*Ziziphus jujuba* Mill.) yellow mottle disease. *Front. Microbiol.* 10:1417. doi: 10.3389/fmicb.2019.01417

Yang, S., Wang, H., Yi, Y., and Tan, L. (2019). First report that *Colletotrichum aenigma* causes leaf spots on *Camellia japonica* in China. *Plant Dis.* 103, 2127–2127. doi: 10.1094/PDIS-01-19-0224-PDN

Yu, C., Karlin, D. G., Lu, Y., Wright, K., Chen, J., and MacFarlane, S. (2013). Experimental and bioinformatic evidence that raspberry leaf blotch emaravirus P4 is a movement protein of the 30K superfamily. *J. Gen. Virol.* 94, 2117–2128. doi: 10.1099/vir.0.053256-0

Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., et al. (2017). ICTV virus taxonomy profile: *Geminiviridae*. *J. Gen. Virol.* 98, 131–133. doi: 10.1099/jgv.0.000738

Zhang, S., Shen, P., Li, M., Tian, X., Zhou, C., and Cao, M. (2018). Discovery of a novel geminivirus associated with camellia chlorotic dwarf disease. *Arch. Virol.* 163, 1709–1712. doi: 10.1007/s00705-018-3780-3

Zhang, W., Wang, L., Kong, W., and Cai, P. (2014). Occurrence and prevention of garden plant sunscald. *J. Landsc. Res.* 6:37. Available online at: https://search.proquest.com/openview/a4c9fc0352ef3f26002d9d749c2ac1fb/1?pq-origsite=gscholar&cbl=1596366

Zhang, Y. M., Maharachchikumbura, S. S. N., Wei, J. G., McKenzie, E. H. C., and Hyde, K. D. (2012). *Pestalotiopsis camelliae*, a new species associated with grey blight of *Camellia japonica* in China. *Sydowia* 64, 335–344. Available online at: http://www.sydowia.at/syd64-2/T13-Zhang.html

Zhao, P., Gao, D.-F., Xu, M., Shi, Z.-G., Wang, D., Yang, C.-R., and Zhang, Y.-J. (2011). Triterpenoid saponins from the genus *Camellia*. *Chem. Biodivers.* 193, 1931–1942. doi: 10.1002/cbdv.201000265

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.

Copyright © 2020 Zhang, Yang, Ma, Tian, Li, Zhou and Cao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 05 June 2020
doi: 10.3389/fmicb.2020.01232

Robust Virome Profiling and Whole Genome Reconstruction of Viruses and Viroids Enabled by Use of Available mRNA and sRNA-Seq Datasets in Grapevine (*Vitis vinifera* L.)

*V. Kavi Sidharthan[1], Amitha Mithra Sevanthi[2], Sarika Jaiswal[3] and V. K. Baranwal[1]**

[1] Division of Plant Pathology, Indian Council of Agricultural Research-Indian Agricultural Research Institute, New Delhi, India, [2] Indian Council of Agricultural Research-National Institute for Plant Biotechnology, New Delhi, India, [3] Centre for Agricultural Bioinformatics, Indian Council of Agricultural Research-Indian Agricultural Statistics Research Institute, New Delhi, India

OPEN ACCESS

Edited by:
Ahmed Hadidi, Agricultural Research Service, United States Department of Agriculture, United States

Reviewed by:
John Hammond, Agricultural Research Service, United States Department of Agriculture, United States
Teruo Sano, Hirosaki University, Japan

***Correspondence:**
V. K. Baranwal
vbaranwal2001@yahoo.com

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

Received: *04 March 2020*
Accepted: *14 May 2020*
Published: *05 June 2020*

Citation:
Sidharthan VK, Sevanthi AM, Jaiswal S and Baranwal VK (2020) Robust Virome Profiling and Whole Genome Reconstruction of Viruses and Viroids Enabled by Use of Available mRNA and sRNA-Seq Datasets in Grapevine (Vitis vinifera L.). Front. Microbiol. 11:1232. doi: 10.3389/fmicb.2020.01232

Next-generation sequencing (NGS) based virome analyses of mRNA and sRNA have recently become a routine approach for reliable detection of plant viruses and viroids. In the present study we identified the viral/viroidal spectrum of several Indian grapevine cultivars and reconstructed their whole genomes using the publically available mRNAome and sRNAome datasets. Twenty three viruses and viroids (including two variants of grapevine leafroll associated virus 4) were identified from two tissues (fruit peels and young leaves) of three cultivars among which nine unique grapevine viruses and viroids were identified for the first time in India. Irrespective of the assemblers and tissues used, the mRNA based approach identified more acellular pathogens than the sRNA based approach across cultivars. Further, the mRNAome was on par with the whole transcriptome in viral identification. Through *de novo* assembly of transcriptomes followed by mapping against reference genome, we reconstructed 19 complete/near complete genomes of identified viruses and viroids. The reconstructed viral genomes included four larger RNA genomes (>13 kb), a DNA genome (RG grapevine geminivirus A), a divergent genome (RG grapevine virus B) and a genome for which no reference is available (RG grapevine virus L). A large number of SNPs detected in this study ascertained the quasispecies nature of viruses. Detection of three recombination events and phylogenetic analyses using reconstructed genomes suggested the possible introduction of viruses and viroids into India from several continents through the planting material. The whole genome sequences generated in this study can serve as a resource for reliable indexing of grapevine viruses and viroids in quarantine stations and certification programs.

Keywords: virome, mRNAome, sRNAome, grapevine, India

INTRODUCTION

Grapevine (*Vitis vinifera* L.) is an important cash crop grown worldwide (McGovern, 2003). Being a clonally propagated crop, grapevine is amenable for coinfection by different viruses and viroids (Jo et al., 2018). It is reported to be susceptible to the largest number of acellular pathogens compared to other crop species (Beuve et al., 2018; Hily et al., 2018). Till date, more than 70 viruses and 7 viroids have been reported to infect grapevine (Singhal et al., 2019). Many a time, grapevine viruses deviate from the classical 'one pathogen – one disease' concept, i.e., interaction among more than one viral agent leads to disease development (Byrd and Segre, 2016).

Planting pathogen free crop propagules is of paramount importance in grapevine for increasing the productive life of vineyards (Kumar et al., 2015). Traditional detection techniques like ELISA, PCR and their variants are employed for indexing of a few selected viruses of grapevine while certifying the planting material for commercial planting. But these methods can only answer whether the pathogen(s) under investigation is present or not, leaving the status of all other untested viruses in the planting material unknown (Czotter et al., 2018). To assure the health of the planting material of grapevine that remains productive in the field for an average of 15 years in India, it would be necessary to subject it to rigorous indexing for all possible grapevine infecting viruses/viroids. For this, it would be essential to study the virome (total viral population) of the mother stock, the results of which can then be used for developing appropriate detection assays of all pathogens for screening the clonal propagules. Next-generation sequencing (NGS) approaches can provide us with a snapshot of the virome present in the propagule as they are effective not only in detecting the known viral pathogens and their variants, if any, but also in unravelling unknown one(s) (Jo et al., 2018). Among the various NGS based approaches, sRNA (sRNAome) and mRNA (mRNAome) sequencing are commonly used to reveal the virome of a given sample (Pantaleo et al., 2010; Pirovano et al., 2015; Jones et al., 2017; Maliogka et al., 2018; Pooggin, 2018; Massart et al., 2019). Recently, a few studies attempted to reveal the virome of different crops like grapevine, apple, and pepper from publically available mRNAome data (Jo et al., 2015, 2016, 2017).

Both sRNA and mRNA pools can effectively capture single as well as double strand RNA viruses and some DNA viruses (Seguin et al., 2014; Roossinck et al., 2015; Jo et al., 2017). However, the relatively lower representation of viral RNA in the background of total plant RNA limits the use of mRNAome compared to sRNAome for viral detection (Beuve et al., 2018; Maliogka et al., 2018). As mRNA based methods can give longer contigs, they are more useful for variant detection, especially when significant genetic diversity exists as found in some of the grapevine viruses such as grapevine leafroll associated virus 3 (GLRaV3) (Xiao et al., 2019). Thus, it would be worthwhile to study the virome using both these methods for robust identification of entire virome of a plant species.

Though India grows grapevine on 137,000 hectares and exports 185,172 tonnes of grapes annually (FAOSTAT, 2017), only a few studies have been attempted to detect grapevine viruses in India. All these studies targeted only one/few virus(es)/viroid(s) at a time using traditional detection methods (Kumar et al., 2012, 2013; Sahana et al., 2013; Adkar-Purushothama et al., 2014; Rai et al., 2017; Marwal et al., 2019; Singhal et al., 2019). The current study is the first virome report of grapevines from India using sRNA and mRNA datasets of three Indian grapevine cultivars available in the public domain (Tirumalai et al., 2019) identifying a large number of viruses and viroids.

MATERIALS AND METHODS

Plant Materials and Library Construction

Detailed information on plant materials and library construction is available in Tirumalai et al. (2019). In brief, total RNA was isolated from fruit peels (FP) and young leaves (YL) of three grapevine cultivars- Bangalore Blue (BB), Dilkush (DK), and Red Globe (RG). mRNA-seq and sRNA-seq libraries with two biological replicates, 24 in total, were constructed from isolated total RNA according to the NEXT flex Rapid directional mRNA-seq bundle library protocol (Trapnell et al., 2012) and the TruSeq Small RNA Sample Preparation Guide (Illumina, San Diego, CA, United States) respectively. Sequencing was performed on the Illumina NextSeq500 platform which yielded 75 bp single end reads. Thus, a total, of 12 mRNA and 12 sRNA libraries obtained from two tissues (FP, YL) of three grapevine cultivars (BB, DK, RG) in two biological replicates were analyzed in the current study. The details of the materials used and the complete processing pipeline are indicated in **Figure 1**.

Raw Data Pre-processing and *de novo* Assembly of Pre-processed Reads

The bioinformatics analyses were performed using Advanced Super Computing Hub for Omics Knowledge in Agriculture (ASHOKA) facility at ICAR-IASRI, New Delhi, India. Raw data of 24 libraries were downloaded from SRA database and converted to FASTQ files using the SRA toolkit version 2.9.6 (Leinonen et al., 2010). Three approaches were followed for *de novo* assembly. In the first approach, mRNA and sRNA libraries, 12 each, were individually assembled using Trinity version 2.5.1 (Grabherr et al., 2011) and CLC genomics workbench 12 *de novo* assembly tool, respectively with default parameters. For the second approach, combined mRNAome or sRNAome for each cultivar was obtained by aggregating corresponding mRNA or sRNA reads, respectively from four libraries (including two tissues and two replicates) of each individual cultivar. Similarly, whole transcriptome of each cultivar was obtained by aggregating both mRNA and sRNA reads from eight libraries of individual cultivar in the third approach. Trinity (k = 25), SPAdes (k = 21,23,25) version 3.13.1 (Bushmanova et al., 2019) and CLC (automatic word size = 20), Velvet (k = 17) version 1.2.10 (Zerbino and Birney, 2008) were used to assemble combined mRNAomes and sRNAomes, respectively while whole transcriptomes were assembled using SPAdes (k = 17,19,21) and Velvet (k = 21) assemblers.

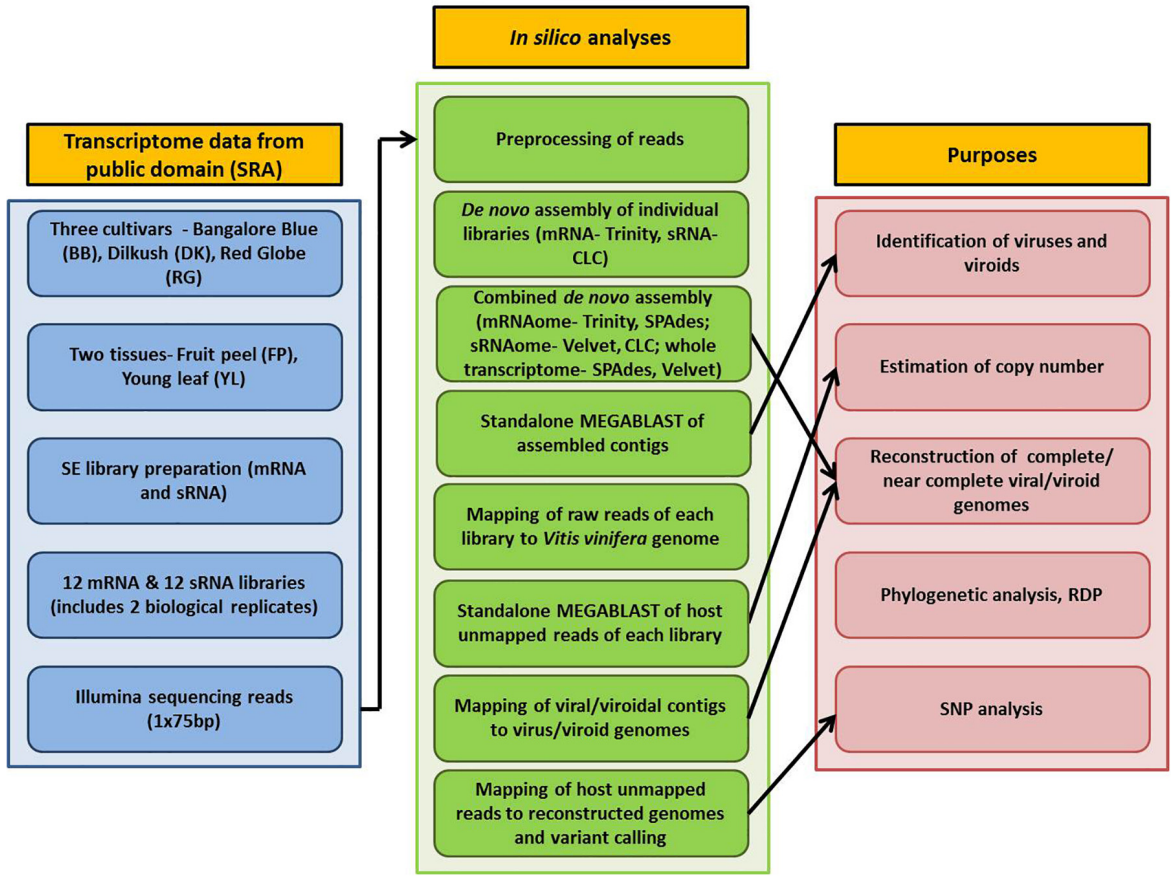


FIGURE 1 | Schematic representation of the viromic study of three Indian grapevine cultivars.

Identification of Viruses and Viroids and Copy Number Estimation

All the assembled contigs were subjected to standalone MEGABLAST analysis (e-value cut off: 1e-5; query coverage: ≥ 80%) against the complete reference sequences of viruses and viroids[1] using NCBI blast+ version 2.9.0. Only contigs of greater than 50 (for sRNAome) and 200 nucleotides (mRNAome and whole transcriptome) were considered for analyses. To validate the viruses/viroids identified through assembly, the reads of each mRNA/ sRNA library were first mapped to the *Vitis vinifera* genome (GCF_000003745.3) using CLC workbench mapping tool with default parameters (match score-1, mismatch cost-2, length fraction-0.5, similarity fraction-0.8). The unmapped reads were then analyzed using MEGABLAST algorithm (e-value cut off: 1e-5; query coverage: ≥80%) against the reference genomes of viruses and viroids. Only those viruses/vioids that were detected through assembly (from sRNAome/mRNAome/whole transcriptome) and BLAST analysis of reads from at least two libraries (derived from the particular nucleic acid pool from which the contigs were obtained) of the corresponding cultivar were considered. To arrive at the copy number for a virus/viroid, the number of reads associated with either RdRp ORF (in case of viruses that use sub-genomic RNA (sgRNA) for translation) or the entire polyprotein [in grapevine fleck virus (GFkV), grapevine rupestris vein feathering virus (GRVFV)] or the entire genome (in viroids) was multiplied with 75 (for mRNA)/ 24 (for sRNA) followed by division with the size (bp) of the corresponding genomic region of the virus/viroid. Intact mRNA reads were used for copy number estimation while the pre-processed reads were used in case of sRNA. The average length of pre-processed sRNA reads in all libraries was near to 24. Hence the factors 75 and 24 were used for mRNA and sRNA libraries, respectively. As a reference genome for grapevine virus L (GVL) was not available in NCBI, we included the *de novo* assembled GVL genome of the present study (that was identified by performing BLASTn analysis of larger contigs against "non-redundant" (nr) (NCBI) database) for MEGABLAST analysis.

Reconstruction of Whole Genomes of Viruses and Viroids

Virus/viroid associated contigs were filtered from the total contigs using SAM tools version 1.9 (Li et al., 2009). The detailed procedure followed for genome reconstruction is given

[1]http://www.ncbi.nlm.nih.gov/genome/viruses/

in **Supplementary Figure S1**. In brief, the Trinity assembled longer contigs from combined mRNAomes were examined for the presence of intact viral/viroidal genome. Further, the SPAdes assembled longer contigs from combined mRNAomes and whole transcriptomes were examined followed by inspection of Trinity assembled larger contigs in individual mRNA libraries. Next, the Trinity assembled viral/viroidal contigs from combined mRNAomes were mapped against the NCBI designated reference genomes of identified viruses and viroids (CLC workbench mapping tool). In cases where the Trinity assembled contigs were insufficient to reconstruct the entire genome, SPAdes assembled contigs from combined mRNAomes and whole transcriptomes were supplemented during mapping. Still, if the genome could not be obtained, the most closely related genome was used as reference during mapping. The full length consensus sequence, if obtained, after mapping/directly by *de novo* assembly was considered as the complete/near complete genome for a particular virus/viroid. To find ORFs in assembled viral genomes, we used NCBI ORF finder[2].

Pairwise Distance and Phylogenetic Analyses

The complete genomes retrieved from NCBI along with the viral/viroid genomes reconstructed in this study were aligned using CLUSTALW tool in MEGA7 software version 7.0.26 (Kumar et al., 2016). Aligned sequences were subjected to pairwise distance analysis and phylogenetic tree construction using neighborhood joining (NJ) method and Kimura 2-parameter (K2P) model with 1000 bootstrap replicates. For grapevine geminivirus A (GGVA), grapevine latent viroid (GLVd), grapevine leafroll associated virus 4 (GLRaV4), grapevine virus B (GVB), GVL, grapevine rootstock stem lesion associated virus (GRSLaV) and GRVFV, all the respective complete genomes available in NCBI were used for analysis. Owing to the availability of a large number of genome sequences for GLRaV3, only those sequences showing 100% query coverage in BLASTn analysis against nr (NCBI) database were taken for analysis. Similarly, in cases of Australian grapevine viroid (AGVd), grapevine yellow speckle viroid-1, -2 (GYSVd1, GYSVd2), and hop stunt viroid (HSVd), only 10 non-redundant genomes that were highly similar to each isolate of a viroid were used. In all the cases, an outgroup (except for pairwise distance analysis) and the NCBI designated reference genome (except GVL, for which there is no designated reference sequence) were included.

Single Nucleotide Polymporphism (SNP) Analyses

The host unmapped reads of individual cultivars were mapped against the complete/near complete viral/viroid genomes assembled from the corresponding cultivar using the mapping tool available in CLC workbench using default parameters (match score-1, mismatch cost-2, length fraction-0.5, similarity fraction-0.8). For SNP detection, the mapped files were subjected to fixed ploidy variant detection using CLC workbench. As viral genomes are haploid, ploidy value was considered as one throughout the analyses.

Recombination Analyses

Using CLUSTALW aligned MEGA file as input, recombination analysis was performed using RDP4 package version 4.39 (Martin et al., 2015) employing nine different algorithms. Only recombination events detected by at least five algorithms in the reconstructed viral genomes were considered. Only viral sequences used for phylogenetic analyses were used for detection of recombinants.

RESULTS

Pre-processing of Raw Data

The number of raw reads ranged from 10.5 to 40.2 million with an average of 23.3 million for mRNA and 2.9 to 8.3 million with an average of 4.3 million for sRNA libraries (**Table 1**). As mRNA reads were of acceptable quality (without adapter sequences; phred-score > 20), we proceeded directly for *de novo* assembly while sRNA reads were filtered to remove adapter sequences and poor quality reads (quality scores < 0.05).

Identification of Viruses and Viroids From Grapevine mRNAome and sRNAome

We identified more viruses and viroids from mRNAome (23) than sRNAome (7) across cultivars and tissues (**Supplementary Table S1**). The only exception for this was the FP-specific sRNA datasets of cv. DK which identified six viruses and viroids while the corresponding mRNAome could identify only five. Among the two tissues, relatively more viruses/viroids were identified in FP than YL in all cultivars except DK from mRNA libraries. However, nearly similar number of viruses/viroids was identified from sRNA libraries across tissues and cultivars (**Supplementary Figure S2** and **Supplementary Table S1**). Combined mRNAome assembly using Trinity identified the same number of viruses and viroids (23) across cultivars as compared to the individual mRNA libraries (23). However, on a closer look, we found that combining the reads of the two tissues of each cultivar did offer some advantage in case of mRNAome, since additional virus(es)/viroid(s) were identified in BB (1), DK (2) and RG (1). The only exception to this is GVE that was detected in individual mRNA libraries but not in combined mRNAome (**Figures 2A,B**, **Supplementary Figures S3A,B**, and **Supplementary Table S2**). Similarly, combined sRNAome assembly using CLC was more effective as it could identify two unique viruses (GVF in DK and GRSLaV in RG) in addition to the seven viruses and viroids identified by the individual library approach across cultivars (**Figures 2D,E**, **Supplementary Figures S4A,B**, and **Supplementary Table S3**). Between the combined sRNAome and combined mRNAome, the former could identify only a fraction of viruses and viroids (12) identified by the latter even after accounting for the viruses and viroids identified by all the assemblers. Interestingly, from combined

[2]www.ncbi.nlm.nih.gov/orffinder

TABLE 1 | Raw reads, host unmapped reads, and viral/viroid associated reads of each library.

Library	Raw mRNA reads	Raw sRNA reads	Host unmapped mRNA reads	Host unmapped sRNA reads	Viral/ viroidal mRNA reads	Viral/ viroidal sRNA reads	Proportion of host unmapped reads from mRNAome	Proportion of host unmapped reads from sRNAome	Proportion of viral/viroid reads as the percentage of unmapped mRNA reads[a]	Proportion of viral/viroid reads as the percentage of unmapped sRNA reads[a]
BBFPR1	10544229	3726821	449774	389040	8699	11	4.27	10.44	1.93	0.003
BBFPR2	23438185	5250400	989202	515220	19392	17	4.22	9.81	1.96	0.003
BBYLR1	13569098	3115864	414503	274662	1477	2	3.05	8.81	0.36	0.001
BBYLR2	40239577	4317810	1225095	284457	4180	11	3.04	6.59	0.34	0.004
DKFPR1	25945875	3789502	840303	430384	19303	113	3.24	11.36	2.30	0.026
DKFPR2	23865863	8263827	781885	697282	16383	413	3.28	8.44	2.10	0.059
DKYLR1	23773201	2962599	899366	179017	5003	11	3.78	6.04	0.56	0.006
DKYLR2	26500597	5197911	1003686	246769	5467	44	3.79	4.75	0.54	0.018
RGFPR1	22622837	4936802	1803089	689055	78225	73	7.97	13.96	4.34	0.011
RGFPR2	21927732	3251448	1795168	599065	73229	19	8.19	18.42	4.08	0.003
RGYLR1	26815464	2942096	892189	253236	26128	175	3.33	8.61	2.93	0.069
RGYLR2	20471622	3421076	643008	336020	21035	126	3.14	9.82	3.27	0.037
Average	23309523.3	4264679.67	978105.67	407850.58	23210.08	84.58	4.27	9.75	2.06	0.020

[a]*Only reads with qcov ≥ 80 were considered.*

mRNAomes and whole transcriptomes exactly the same number of viruses/viroids was identified in BB, DK and RG cultivars (6, 10, and 21), representing a total of 23 viruses/viroids though the identities of a few differed in cvs. DK and RG. The identified acellular pathogens included 14 grapevine viruses (including two GLRaV4 variants), four mitoviruses and five viroids – Alternaria alternata chrysovirus 1 (AaCV1), Alternaria arborescens mitovirus 1 (AaMV1), AGVd, Erysiphe necator mitovirus 1 (EnMV1), Erysiphe necator mitovirus 3 (EnMV3), GFkV, GGVA, GLVd, GLRaV3, grapevine leafroll associated virus -4, -5, -6 (GLRaV4, GLRaV5, GLRaV6), GVA, GVB, GVL, GVE, GVF, GRSLaV, GRVFV, GYSVd1, GYSVd2, HSVd and tobacco streak virus (TSV). It is worthy of mention that while TSV was detected from cvs. BB, RG in the mRNAome, none of the sRNA libraries could detect it (**Supplementary Tables S2, S3**). Interestingly, GLRaV3, GVA, GVB, GYSVd1, and HSVd were identified in all the cultivars (**Figures 3A–D** and **Supplementary Table S4**).

Performance of Different Assemblers in Identification of Viruses and Viroids From mRNAome, sRNAome, and Whole Transcriptomes

In case of combined mRNAomes, the number of viruses and viroids identified by both the assemblers were similar except in cv. RG where Trinity identified one additional viroid (GYSVd1) compared to SPAdes (**Figures 2B,C** and **Supplementary Table S2**). However, both the assemblers identified almost similar number of viral contigs for all but one cultivar (**Supplementary Figures S3B,C**). In case of combined sRNAomes, Velvet identified one additional virus in BB (GVA) and RG (GVE) than CLC but CLC detected two Velvet undetected virus/viroid in DK (GVF, GYSVd1). Interestingly, GVB was identified in cv. DK by Velvet but not by CLC (**Figures 2E,F** and **Supplementary Table S3**). Considering the number of viral and viroid contigs, Velvet identified 67.3, 14.7, and 52.3% more contigs than CLC in BB, DK, and RG (**Supplementary Figures S4B,C**). In case of whole transcriptome assembly, SPAdes identified more viruses (2, 3, and 6 additional viruses/viroids in cvs. BB, DK, and RG, respectively) and viral contigs as compared to Velvet in all cultivars. Notably Velvet based assembly failed to identify HSVd from any whole transcriptome, or GYSVd2 from DK or GYSVd1,2 from RG, whereas SPAdes identified HSVd in BB and RG, and GYSVd2 in DK and GYSVd1,2 in RG from the corresponding whole transcriptomes. However, Velvet did detect HSVd in each of the combined sRNAomes and GYSVd2 from DK and GYSVd1,2 from RG (**Figures 2G,H**, **Supplementary Figures S5A,B**, and **Supplementary Tables S3, S5**).

Copy Number Estimation for Identified Viruses and Viroids in Each mRNA and sRNA Library

The number of host unmapped mRNA and sRNA reads ranged from 0.41 to 1.80M and 0.18 to 0.70M, respectively across

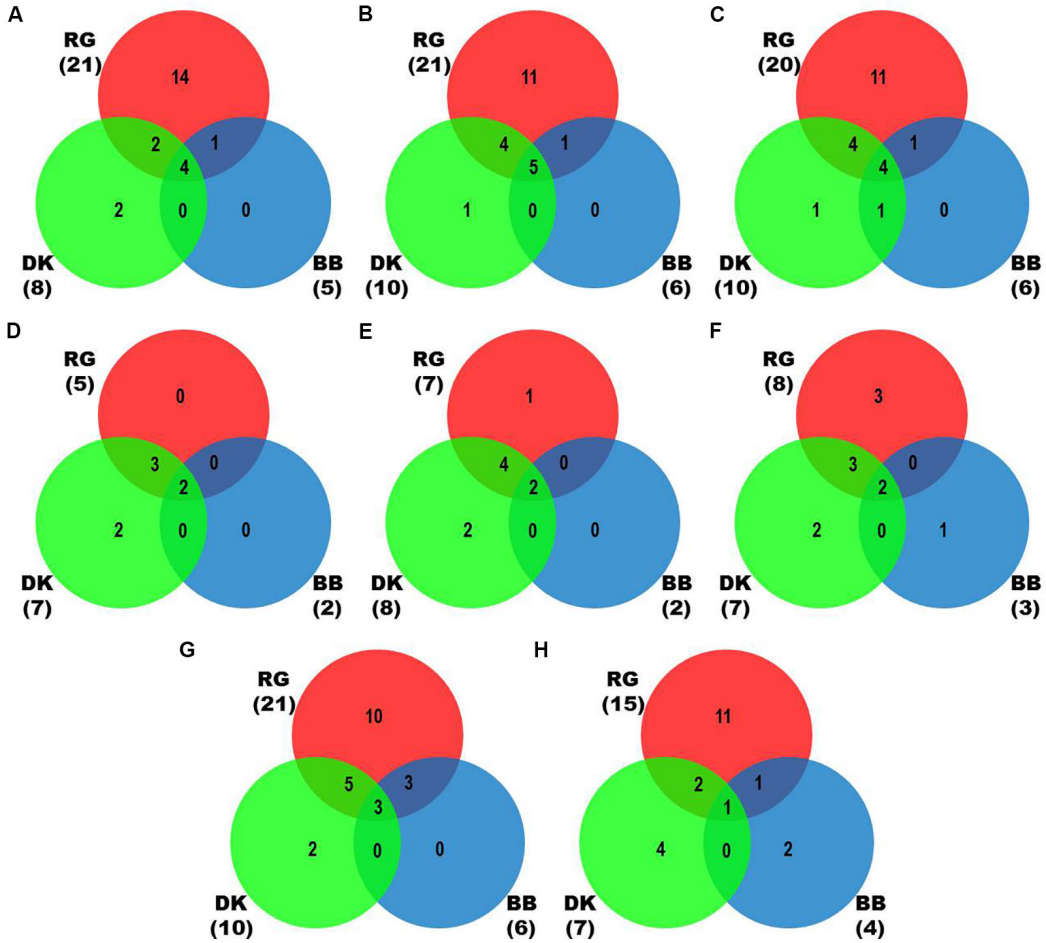


FIGURE 2 | Identification of viruses and viroids in three Indian grapevine cultivars – BB, DK, RG. Venn diagrams display the number of identified viruses and viroids based on Trinity assembled contigs from individual mRNA libraries of each cultivar **(A)**, Trinity **(B)**, SPAdes (*k* = 21,23,25) assembled contigs from combined mRNAome of each cultivar **(C)**, CLC assembled contigs from individual sRNA libraries of each cultivar **(D)**, CLC **(E)**, Velvet (*k* = 13,15,17) assembled contigs from combined sRNAome of each cultivar **(F)**, SPAdes (*k* = 17,19,21) **(G)** and Velvet (*k* = 17,19,21) **(H)** assembled contigs from whole transcriptome of each cultivar. Blue, green and red circles in the Venn diagrams represent the cvs. BB, DK, and RG, respectively.

libraries. Though the number of host unmapped reads was higher (0.98M) in case of mRNA compared to sRNA (0.41M), the proportion of unmapped reads to total reads was higher in the latter (9.75%) than the former (4.27%). On average, 2.06 and 0.02% of host-unmapped reads from mRNA and sRNA libraries mapped to viral/viroidal genomes (**Supplementary Figures S6A,B** and **Table 1**). In general, the proportion of virus/viroid associated reads was relatively higher in mRNA libraries constructed from FP than YL while no such trend was observed in case of sRNA libraries. Based on copy number estimates, HSVd (94–100%) predominated in cv. BB in both mRNA and sRNA libraries. In case of mRNA libraries of cvs. DK and RG, HSVd and GYSVd2 were predominant in FP and YL, respectively. In sRNA libraries of cv. DK and in all but one sRNA libraries of cv. RG, GYSVd2 was predominant irrespective of tissue type (**Supplementary Figures S7A,B**). Further, both the replicates in each tissue of a cultivar were highly similar not only in detecting the viromes but also in estimating their copy number.

Viral/Viroid Genome Reconstruction From *de novo* Assembled Contigs

By mapping, the viral/viroid associated contigs from combined mRNAome and whole transcriptome of each cultivar against the NCBI designated reference genomes of identified viruses and viroids we obtained complete or near complete (>99%) genomes of 15 viruses and viroids from three cultivars (**Table 2**). Some other viral contigs could not be assembled into full genomes using the reference genomes as scaffolds. For the assembly of GLRaV3 and GLRaV4 genomes from cv. RG, the longest Trinity assembled contig of each virus was first blasted against the nr (NCBI) database. The complete genome of the most highly similar isolate was then used as a reference during mapping in each case. Trinity assembly of library RGFPR2 directly yielded the whole genome of GVB. Similarly, we obtained GVL genome from one of the Trinity assembled longest contigs from combined mRNAome of cv. RG through BLAST against nr (NCBI) database. In total, we obtained 19 complete/near complete viral/

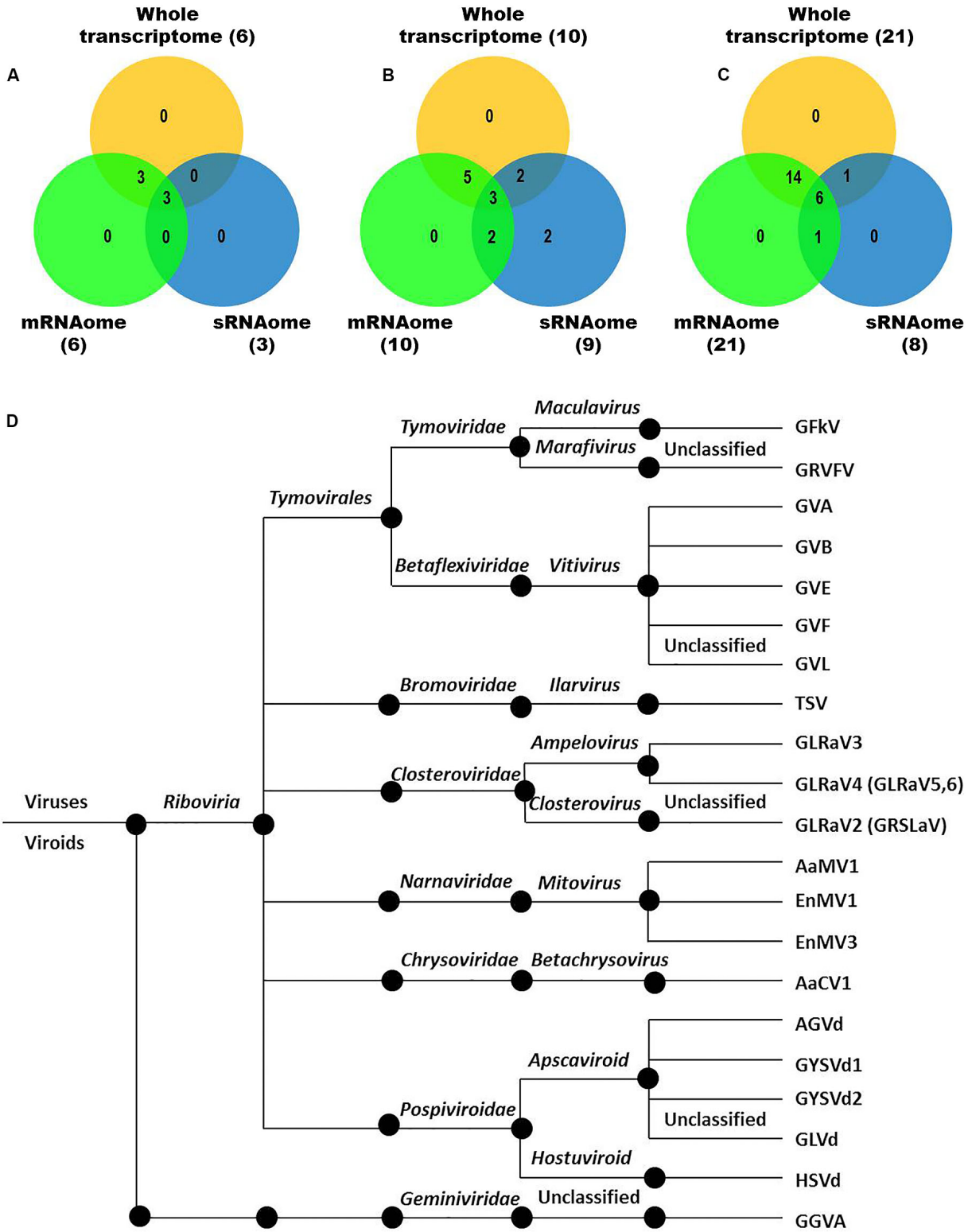


FIGURE 3 | Viruses and viroids identified in three Indian grapevine cultivars – BB, DK, RG. Venn diagram displaying viruses and viroids identified by any of the assemblers from combined mRNAomes, sRNAomes and whole transcriptomes in cultivars BB **(A)**, DK **(B),** and RG **(C)**. Blue, green, and yellow circles in the Venn diagrams represent the sRNAome, mRNAome, and whole transcriptome, respectively. Classification of the identified viruses and viroids is on the basis of taxonomy **(D)**.

TABLE 2 | Summary of complete/near complete viral/viroidal genomes reconstructed from three Indian grapevine cultivars.

Sl. no.	Cultivar	Virus/ viroid	Reference	Size (nt)	Genome recovery (nt)	Assigned isolate name	GenBank accession number	Contigs assembled from	Library[a]	Reference used for mapping
1	Bangalore Blue	GYSVd1	NC_001920.1	366	367	BB GYSVd1	MN662238	Trinity, SPAdes	CmR, WT	NCBI reference genome
2	Bangalore Blue	HSVd	NC_001351.1	302	300	BB HSVd	MN662243	Trinity, SPAdes	CmR, WT	NCBI reference genome
3	Dilkush	GLRaV3	NC_004667.1	17919	17919	DK GLRaV3	MN662228	Trinity	CmR	NCBI reference genome
4	Dilkush	AGVd	NC_003553.1	369	369	DK AGVd	MN662235	Trinity, SPAdes	CmR, WT	NCBI reference genome
5	Dilkush	GYSVd1	NC_001920.1	366	367	DK GYSVd1	MN662239	Trinity, SPAdes	CmR, WT	NCBI reference genome
6	Dilkush	GYSVd2	NC_003612.1	363	363	DK GYSVd2	MN662241	Trinity, SPAdes	CmR, WT	NCBI reference genome
7	Dilkush	HSVd	NC_001351.1	302	301	DK HSVd	MN662244	Trinity, SPAdes	CmR, WT	NCBI reference genome
8	Red Globe	GLRaV3	MH814489.1	18558	18469	RG GLRaV3	MN662229	Trinity	CmR	Closely related genome*
9	Red Globe	GRSLaV	NC_004724.1	16527	16519	RG GRSLaV	MN662231	Trinity	CmR	NCBI reference genome
10	Red Globe	GLRaV4	KY821095.1	13857	13780	RG GLRaV4	MN662230	Trinity, SPAdes	CmR	Closely related genome*
11	Red Globe	GGVA	NC_031340.1	2905	2905	RG GGVA	MN661401	Trinity	CmR	NCBI reference genome
12	Red Globe	GVB	*De novo*	NA	7621	RG GVB	MN662233	Trinity	RGFPR2	NA
13	Red Globe	GVL	*De novo*	NA	7588	RG GVL	MN662234	Trinity	CmR	NA
14	Red Globe	GRVFV	NC_034205.1	6730	6703	RG GRVFV	MN662232	Trinity	CmR	NCBI reference genome
15	Red Globe	AGVd	NC_003553.1	369	370	RG AGVd	MN662236	Trinity	CmR	NCBI reference genome
16	Red Globe	GLVd	NC_028131.1	328	329	RG GLVd	MN662237	Trinity, SPAdes	CmR, WT	NCBI reference genome
17	Red Globe	GYSVd1	NC_001920.1	366	367	RG GYSVd1	MN662240	Trinity, SPAdes	CmR, WT	NCBI reference genome
18	Red Globe	GYSVd2	NC_003612.1	363	363	RG GYSVd2	MN662242	Trinity	CmR	NCBI reference genome
19	Red Globe	HSVd	NC_001351.1	302	301	RG HSVd	MN662245	Trinity, SPAdes	CmR, WT	NCBI reference genome

[a]*CmR- Combined mRNAome, WT- Whole Transcriptome, RGFPR2- Red Globe Fruit Peel Replicate 2 (one of the individual libraries). NA- Not Applicable. *Initially RG GLRaV3, 4 genomes could not be reconstructed by using their corresponding NCBI designated reference genomes during mapping. However, use of their most closely related genome, that were identified by BLASTn analysis of longest Trinity assembled contigs against nr (NCBI) database, yielded their near complete genomes.*

viroid genomes from three cultivars (**Table 2**). Trinity yielded relatively longer contigs for most viruses and viroids as compared to SPAdes in all cultivars with mRNA reads (**Supplementary Figures S8, S9**). On the contrary, SPAdes yielded relatively longer viral/viroid contigs as compared to Velvet in most instances when whole transcriptomes were assembled (**Supplementary Figures S10, S11**). Though Velvet assembled more viral/viroid contigs from combined sRNAomes, CLC yielded longer contigs for most viruses and viroids as compared to Velvet (**Supplementary Figures S12, S13**). However, we could not reconstruct any viral genome using contigs assembled from combined sRNAomes. From the reconstructed complete/near complete viral genomes, we could identify all of the anticipated ORFs for all recovered viruses using NCBI ORF finder (**Supplementary Table S6**). Failure to identify intact ORFs in nearly complete genomes that could be assembled to the tune of >95% (**Supplementary Table S7**) were still deemed incomplete.

Pairwise Distance and Phylogenetic Analyses Using Reconstructed Viral/Viroid Genomes

Each of the complete/near complete genomes obtained were subjected to pairwise distance (**Supplementary Tables S8–S19**) and phylogenetic analyses (**Figures 4A–L**) along with related complete genomes retrieved from NCBI, and the most closely related genomes are indicated here, including their country of origin.

Viruses

a-RG GVB (MN662233) was related to a South African isolate GVB-H1 [79.7% nucleotide (nt) identity; GU733707.1] (**Figure 4A** and **Supplementary Table S8**).

b-RG GLRaV4 (MN662230) was closely related to a Pakistani isolate LH3 (97.9% nt identity; KY821095.1) (**Figure 4B** and **Supplementary Table S9**).

c-RG GRVFV (MN662232) was related to a New Zealand isolate NZ Ch8021 (80.6% nt identity; MF000325.1) (**Figure 4C** and **Supplementary Table S10**).

d-RG GGVA (MN661401) was most closely related to a Japanese isolate Pione (99.6% nt identity; KX570616.1) (**Figure 4D** and **Supplementary Table S11**).

e-RG GVL (MN662234) was closely related to a Croatian isolate VL (94.0% nt identity; MH681991.1) (**Figure 4E** and **Supplementary Table S12**).

f-RG GLVd (MN662237) was closely related to an Italian isolate ITA (98.4% nt identity; MG770884.1) (**Figure 4F** and **Supplementary Table S13**).

g-RG GRSLaV (MN662231) was most closely related to a Californian isolate obtained from the cv. RG (99.8% nt identity; NC_004724.1) (**Figure 4G** and **Supplementary Table S14**).

h-DK GLRaV3 (MN662228) shared 99.0% nt identity with Canadian isolates- 14G463 (MH814490.1), 14G466 (MH814491.1), 14G462 (MH814489.1), 3138-07 (JX559645.1), a Brazilian isolate TRAJ-BR (KX756669.1) and an US isolate WA-MR (GU983863.1).

i-RG GLRaV3 (MN662229) shared 99.4% nt identity with US isolates- Bla223 (MH521090.2), WA-MR 314 (GU983863.1), Canadian isolates- 14G463 (MH814490.1), 14G462 (MH814489.1) and a Brazilian isolate TRAJ-BR (KX756669.1).

j-GLRaV3 isolates from cvs. DK and RG showed only 1.3% divergence (**Figure 4H** and **Supplementary Table S15**).

Viroids

k-BB HSVd (MN662243) shared 99.3% nt identity with Brazilian isolates- VL-TC (MG431974.1), VV-CF (MF774875.1), VV-CG (MF774872.1), VV-IT (MF774871.1), VV-CS (MF774862.1) and a German isolate obtained from cv. Riesling (X06873.1).

l-DK HSVd (MN662244) was closely related to a Nigerian isolate DgHV-6 (99.3% nt identity; MF576419.1) while RG HSVd (MN662245) was closely related to a New Zealand isolate 09-2009-2140hs (96.8% nt identity; HQ447057.1). Interestingly, RG HSVd diverged largely from DK HSVd (4.7%) and BB HSVd (5.5%) while DK HSVd showed 2.9% divergence from BB HSVd (**Figure 4I** and **Supplementary Table S16**).

m-Similarly, while BB GYSVd1 (MN662238) was closely related to a Thailand isolate Wangnamkeay-5 (98.0% nt identity; KP010008.1), DK GYSVd1 (MN662239) shared 99.7% nt identity with a German isolate IXc (X87913.1), two Nigerian isolates DgSV1-8 (MF576407.1), R3SV1-2 (MF576403.1) and two Chinese isolates clone 6s CZZ (KX966267.1), clone 7s CZZ (KX966268.1) and RG GYSVd1 (MN662240) was most closely related to yet another Pakistani isolate Q4-III (99.1% nt identity; KY978404.1).

n-Like HSVd, RG GYSVd1 diverged from DK (3.2%) and BB GYSVd1 (5.0%) while DK GYSVd1 showed 2.9% divergence from BB GYSVd1 (**Figure 4J** and **Supplementary Table S17**).

o-RG GYSVd2 (MN662242) was identical (100% nt identity) to Greek isolates- Sup4 (LR735996.1), Sup3 (LR735995.1), N132 (LR735994.1), a Croatian isolate VB-108 (MF979530.1), a Pakistani isolate SL13- I (KY978405.1) and a Chinese isolate clone 22-8 (FJ490172.1) while the DK GYSVd2 (MN662241) isolate shared 99.7% nt identity with all the six closest relatives of RG GYSVd2 and also with the RG GYSVd2 isolate (**Figure 4K** and **Supplementary Table S18**).

p-DK AGVd (MN662235) was identical to a Chilean isolate 6089_AGVd_Crim (100% nt identity; KF007272.1) while RG AGVd (MN662236) shared 99.4% nt identity with an Indian isolate Ind-2 (KJ019301.1) and two Chinese isolates- clone 22-2 (EU743606.1), clone 22-129 (FJ746822.1). It is noteworthy that DK and RG AGVd isolates fell within two separate clades (1.9% divergence) (**Figure 4L** and **Supplementary Table S19**).

SNP Detection and Recombination Analyses in Reconstructed Genomes

A large number of SNPs was detected for RG GRVFV (168) followed by RG GLRaV3 (117), that were equally distributed throughout the genome, while no SNP was detected in case of RG GGVA, RG GLVd, RG GYSVd2, DK AGVd, and BB GYSVd1. Other viruses that had a good number of SNPs included GLRaV3

from cv. DK (102) and GLRaV4 (100), GVL (64), and GVB (40) from cv. RG (**Figure 5A**).

The reconstructed viral genomes, after alignment, were subjected to detection of recombination events. Among the eight reconstructed viral genomes, recombination events supported by at least five algorithms were detected in only three genomes. In GLRaV3 genomes of DK and RG, a similar recombination event was detected in 5′ region of the genome. An additional recombination event was detected in 3′ region of RG GLRaV3. For RG GLRaV4, we found only one recombinant sequence at 3′ region (**Figure 5B** and **Supplementary Table S20**).

DISCUSSION

In this study, viromes of three Indian grapevine cultivars were determined and some of their whole genomes were reconstructed from publically available mRNAome and sRNAome datasets (Tirumalai et al., 2019). Since the materials used in the present study were obtained from Indian Institute of Horticultural Research (in Bangalore, India) one of the leading grapevine breeding centers in the tropical region (Tirumalai et al., 2019), it is the most appropriate one for performing virome analysis as all the vegetative propagules derived from the breeding stock would be expected to be infected with the same viruses. Interestingly the cv. RG, an introduction from California had the maximum viral load in our study compared to the native cvs., BB and DK.

Uneven distribution of viruses and viroids across tissues of a perennial plant like grapevine (Kominek et al., 2009), suggested that sampling different tissues will reveal a more accurate sanitary status of a plant. We also found pooling samples from different tissues was more reliable than relying on individual tissue for virome analysis. Earlier, Jo et al. (2015), also reported the superiority of tissues-combined assemblies over the individual ones. We further observed that the combined mRNAome and whole transcriptome identified nearly similar acellular pathogens and both these approaches were more sensitive than individual or combined sRNAomes. This might be because of the smaller size and number of reads generated from sRNA libraries. Contrary

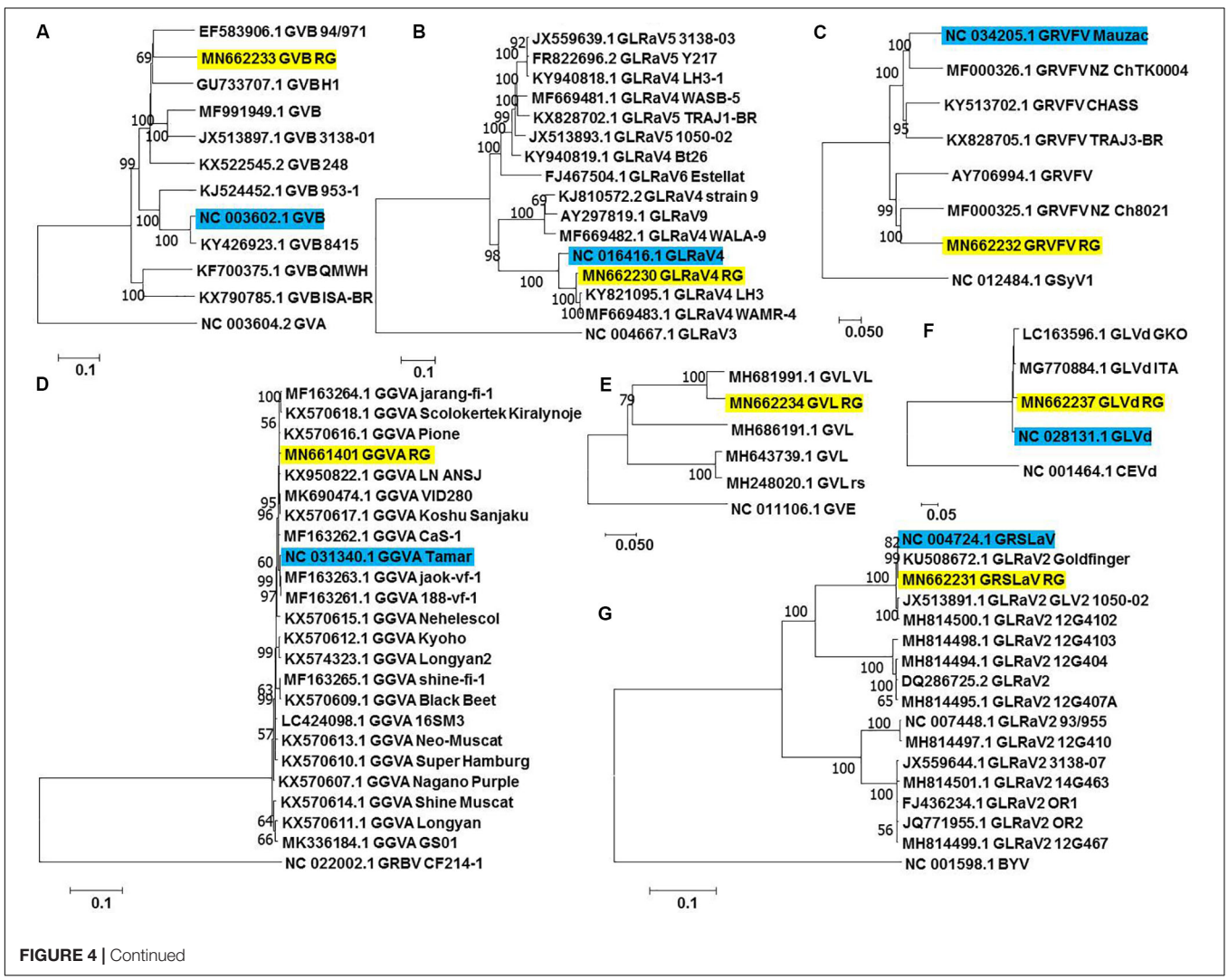


FIGURE 4 | Continued

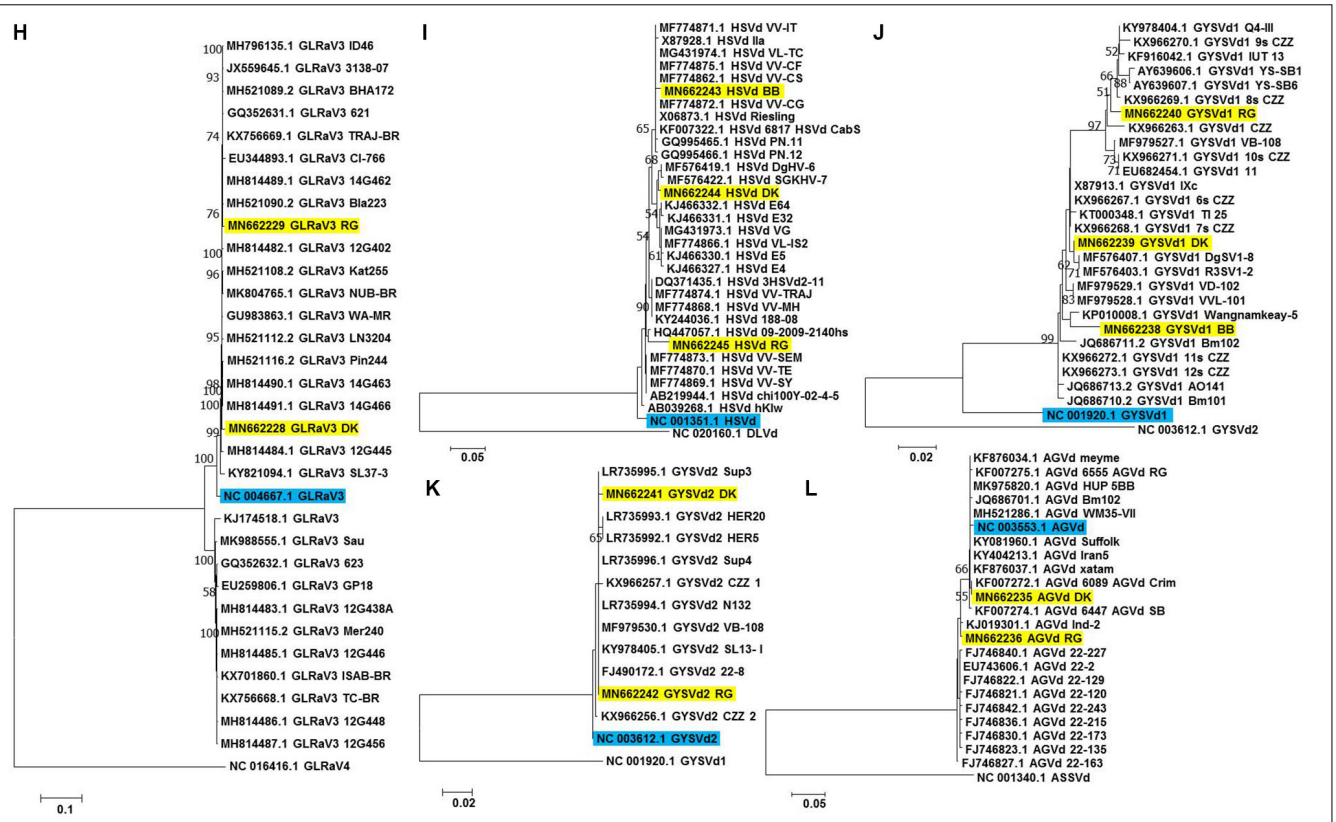


FIGURE 4 | Phylogenetic analyses for genome reconstructed viruses and viroids. Phylogenetic relationships of GVB **(A)**, GLRaV4 **(B)**, GRVFV **(C)**, GGVA **(D)**, GVL **(E)** GLVd **(F)**, GRSLaV **(G)**, GLRaV3 **(H)**, HSVd **(I)**, GYSVd1 **(J)**, GYSVd2 **(K),** and AGVd **(L)** isolates obtained in this study with global isolates. Phylogenetic tree was constructed using Neighborhood joining (NJ) method and Kimura 2-parameter model (K2P) with 1000 bootstrap replicates. Yellow colored boxes indicate the isolates obtained in this study. Blue colored boxes indicate the corresponding NCBI designated reference genome. Only bootstrap values more than 50 are shown. An outgroup was used in the analysis of each virus/viroid.

to the observation of Maliogka et al. (2018), the proportion of viral and viroidal reads in mRNA libraries was higher than sRNA libraries in our study. This might be due to the fact that viral sRNAs are produced only upon activation of host's antiviral defense while mRNAomes can even detect viruses and viroids that are unrecognized by the host (Hily et al., 2018). Further, similar number of viruses and viroids were identified by Trinity and SPAdes assemblers from mRNAomes and CLC and Velvet assemblers from sRNAomes. However, SPAdes outperformed Velvet in case of whole transcriptomes. So, when more than one assembler was used, one or more viruses that escaped detection by one assembler could be detected by the other (Massart et al., 2019). Thus, use of multiple tissues and assemblers enabled better unraveling of grapevine virome.

In the present study, we identified 19 grapevine viruses and viroids (including two variants of GLRaV4) and four mycoviruses associated with the grapevine fungal pathogens- *Erysiphe necator* and *Alternaria* spp. (Kakalikova et al., 2009; Feng et al., 2018). Included among these is GRSLaV, which was earlier reported as a novel virus from California in cv. RG. This indicates the possible introduction of GRSLaV from California along with the RG propagule. However, GRSLaV is now regarded as a strain of grapevine leafroll associated virus 2 (GLRaV2) (as GLRaV-2RG) (Alkowni et al., 2011). Nonetheless, this is the first study that could successfully detect GLRaV-2 or any of its variants in India. Though Kumar et al. (2013) did attempt to detect this virus in India they could not succeed rather they detected GLRaV1 and GLRaV3. Further, GLRaV5 and GLRaV6 are presently regarded as the strains of GLRaV4 (Rai et al., 2017). On this basis, nine grapevine viruses and viroids (GGVA, GLRaV2, GRVFV, GVA, GVE, GVF, GVL, TSV, and GLVd) were detected for the first time in grapevine cultivars grown in Indian soil. Interestingly, we could identify GVL, the reference for which is not yet available in the NCBI, using the GVL genome obtained in this study.

Of the 19 complete/near complete genomes (>99% completion but <100%) obtained in this study, seven viral (including four genomes with > 13 kb) and 1 viroidal genome were recovered for the first time from any Indian grapevine cultivar. None of the viral whole genomes could be recovered from combined sRNAome assembled contigs as reported by Baranwal et al. (2015) and Jo et al. (2016). However, this might be due to the use of lower number of sRNA reads (approximately one-fifth) as compared to the mRNA reads in the current study. Identification of DNA viruses in mRNAome is rare and construction of their whole genome is still scarce (Jo et al., 2017), but we could not only identify GGVA in mRNA of cv.

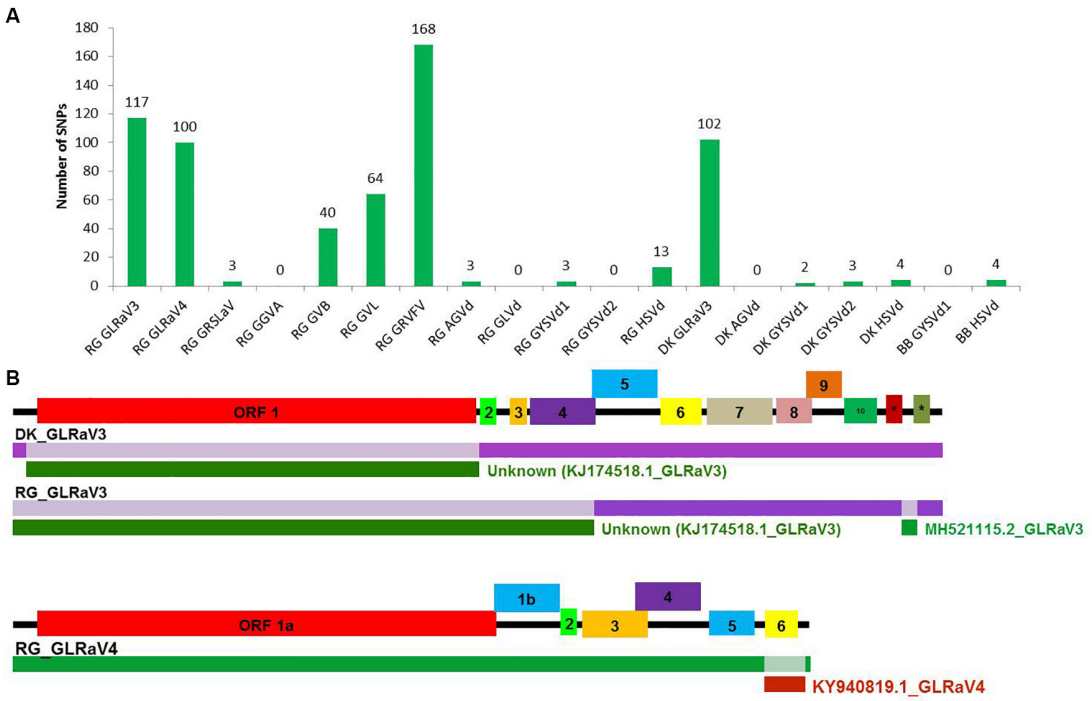


FIGURE 5 | SNP detection and recombination analyses in whole genome reconstructed viruses and viroids. The number of identified SNPs for each virus and viroid **(A)**. Identification of recombination events in reconstructed viral genomes using RDP4 program. **(B)** The positions of identified recombination events are indicated in a different color in the genome. The corresponding minor parental sequence is indicated by a differently colored bar below each event along with the name. #-ORF 11, *-ORF 12 in GLRaV3.

RG but could also reconstruct its genome in entirety with 2905 nucleotides. Initially, the RG GVB and GVL genomes could not be recognized as the former diverged significantly (23%) from the reference genome while there was no reference genome for the latter. However, inspection of Trinity assembled longer contigs of individual and combined mRNA libraries through BLAST analysis against nr (NCBI) database coupled with ORF prediction, identified the whole genomes of these isolates. Complete genomes could not be reconstructed for RG GLRaV3, 4 isolates using reference-based mapping because of their divergence (2.7 and 9.6% divergence of RG GLRaV3, 4 genomes, respectively) from the corresponding reference genome, though their near complete genomes could be reconstructed using the closely related genomes. Thus, examination of larger contigs assembled by various *de novo* assemblers coupled with usage of increased number of reference genomes of a virus during mapping could increase the chances of whole genome recovery. Identification of large number of viral SNPs in this study ascertains the quasispecies nature of plant viruses (Jo et al., 2018). Hence, the complete/near complete genomes reconstructed in this study were the consensus of viral variants present in a given cultivar.

We followed Jo et al. (2015) for copy number estimation except that we considered reads from only the non sgRNA region to reflect the true abundance of viruses that use sgRNA strategy for translation. Phylogenetic and distance matrix analyses revealed the divergence of AGVd, HSVd and GYSVd1 isolates obtained from different cultivars while the GLRaV3 and GYSVd2 isolates from cvs. DK and RG were related. Recombination analyses revealed that the RG GLRaV3, DK GLRaV3, and RG GLRaV4 isolates were recombinants of global isolates. Owing to the vegetative propagation of grapevine and free trade of planting materials, viruses and viroids can easily spread globally. In addition, coinfection of a single plant with numerous isolates of same/different viruses offers scope for recombination among different isolates (Jo et al., 2016).

Plants do not always express symptoms associated with every virus/viroid present, hence symptomology and individual virus/viroid based detection assays are not sufficient to determine the full spectrum of viruses/viroids present in a plant. Rather, use of available or newer transcriptome datasets is a better choice for profiling of viromes that can serve as a reliable base for indexing of planting materials in plant quarantine stations and during certification.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this study are available in the NCBI repository under the Bioprojects PRJNA421907 (sRNA-seq) and PRJNA421908 (mRNA-seq). The whole genomes of 19 viruses and viroids reconstructed in this study have been submitted to GenBank (MN662228 to MN662245 and MN661401).

AUTHOR CONTRIBUTIONS

VS, AS, SJ, and VB conceptualized and formulated the study and read and approved the final manuscript. VS performed the bioinformatics analyses. VS and AS have drafted the manuscript. VB and SJ edited the manuscript.

ACKNOWLEDGMENTS

The authors are thankful to ICAR for the ASHOKA facility at ICAR-IASRI, New Delhi. They also acknowledge the support from Head, Division of Plant Pathology, ICAR-IARI, Dean, ICAR-IARI and Director, ICAR-IARI, New Delhi. VS is grateful to the DST for INSPIRE fellowship. The authors are also thankful to Mr. S. Saravana Kumar (SRF at IISER, Trivandrum), Mr. V. Sureshkumar (Young Professional at ICAR-NIPB, New Delhi) and Dr. Susheel Kumar Sharma (Scientist at ICAR-RC NEH, Manipur centre, Imphal) for providing technical guidance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01232/full#supplementary-material

REFERENCES

Adkar-Purushothama, C. R., Kanchepalli, P. R., Yanjarappa, S. M., Zhang, Z., and Sano, T. (2014). Detection, distribution, and genetic diversity of *Australian grapevine viroid* in grapevines in India. *Virus Genes* 49, 304–311. doi: 10.1007/s11262-014-1085-5

Alkowni, R., Zhang, Y. P., Rowhani, A., Uyemoto, J. K., and Minafra, A. (2011). Biological, molecular, and serological studies of a novel strain of *Grapevine leafroll-associated virus 2*. *Virus Genes* 43, 102–110. doi: 10.1007/s11262-011-0607-7

Baranwal, V. K., Jain, P., Saritha, R. K., Jain, R. K., and Gautam, N. K. (2015). Detection and partial characterization of *Cowpea mild mottle virus* in mungbean and urdbean by deep sequencing and RT-PCR. *Crop. Prot.* 75, 77–79. doi: 10.1016/j.cropro.2015.05.015

Beuve, M., Hily, J. M., Alliaume, A., Reinbold, C., Le Maguet, J., Candresse, T., et al. (2018). A complex virome unveiled by deep sequencing analysis of RNAs from a french pinot noir grapevine exhibiting strong leafroll symptoms. *Arch. Virol.* 163, 2937–2946. doi: 10.1007/s00705-018-3949-9

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. (2019). rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *Gigascience* 8:giz100. doi: 10.1093/gigascience/giz100

Byrd, A. L., and Segre, J. A. (2016). Adapting Koch's postulates. *Science* 351, 224–226. doi: 10.1126/science.aad6753

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.* 9:122. doi: 10.3389/fmicb.2018.00122

FAOSTAT (2017). *Statistical Databases*. Available online at: http://faostat.fao.org.

Feng, X., Nita, M., and Baudoin, A. B. (2018). Evaluation of quinoxyfen resistance of *Erysiphe necator* (Grape Powdery Mildew) in a single virginia vineyard. *Plant Dis.* 102, 2586–2591. doi: 10.1094/PDIS-11-17-1822-RE

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29:644. doi: 10.1038/nbt.1883

Hily, J. M., Candresse, T., Garcia, S., Vigne, E., Tannière, M., Komar, V., et al. (2018). High-throughput sequencing and the viromic study of grapevine leaves: From the detection of grapevine-infecting viruses to the description of a new environmental *Tymovirales* member. *Front. Microbiol.* 9:1782. doi: 10.3389/fmicb.2018.01782

Jo, Y., Choi, H., Cho, J. K., Yoon, J. Y., Choi, S. K., and Cho, W. K. (2015). *In silico* approach to reveal viral populations in grapevine cultivar Tannat using transcriptome data. *Sci. Rep.* 5:15841. doi: 10.1038/srep15841

Jo, Y., Choi, H., Kim, S. M., Kim, S. L., Lee, B. C., and Cho, W. K. (2016). Integrated analyses using RNA-Seq data reveal viral genomes, single nucleotide variations, the phylogenetic relationship, and recombination for *Apple stem grooving virus*. *BMC Genomics* 17:579. doi: 10.1186/s12864-016-2994-6

Jo, Y., Choi, H., Kim, S. M., Kim, S. L., Lee, B. C., and Cho, W. K. (2017). The pepper virome: natural co-infection of diverse viruses and their quasispecies. *BMC Genomics* 18:453. doi: 10.1186/s12864-017-3838-8

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

Jones, S., Baizan-Edge, A., MacFarlane, S., and Torrance, L. (2017). Viral diagnostics in plants using next generation sequencing: computational analysis in practice. *Front. Plant Sci.* 8:1770. doi: 10.3389/fpls.2017.01770

Kakalikova, L', Jankura, E., and Šrobárová, A. (2009). First report of *Alternaria* bunch rot of grapevines in Slovakia. Australas. *Plant Dis. Notes* 4, 68–69. doi: 10.1071/DN09029

Kominek, P., Glasa, M., and Komínková, M. (2009). Analysis of multiple virus-infected grapevine plant reveals persistence but uneven virus distribution. *Acta Virol.* 53:281. doi: 10:4149/av_2009_04_281

Kumar, S., Rai, R., and Baranwal, V. K. (2015). Development of an immunocapture–reverse transcription–polymerase chain reaction (IC-RT-PCR) using modified viral RNA release protocol for the detection of *Grapevine leafroll-associated virus 3* (GLRaV-3). *Phytoparasitica* 43, 311–316. doi: 10.1007/s12600-014-0445-y

Kumar, S., Sawant, S. D., Sawant, I. S., Prabha, K., Jain, R. K., and Baranwal, V. K. (2012). First report of *Grapevine leafroll-associated virus 1* infecting grapevines in India. *Plant Dis.* 96, 1828–1828. doi: 10.1094/PDIS-07-12-0647-PDN

Kumar, S., Singh, L., Ferretti, L., Barba, M., Zaidi, A. A., and Hallan, V. (2013). Evidence of *Grapevine leafroll associated virus-1–3*, *Grapevine fleck virus* and *Grapevine virus B* occurring in Himachal Pradesh, India. *Indian J. Virol.* 24, 66–69. doi: 10.1007/s13337-013-0129-0

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

Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database Collaboration. (2010). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Maliogka, V. I., Minafra, A., Saldarelli, P., Ruiz-García, A. B., Glasa, M., Katis, N., et al. (2018). Recent advances on detection and characterization of fruit tree viruses using high-throughput sequencing technologies. *Viruses* 10:436. doi: 10.3390/v10080436

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

Marwal, A., Kumar, R., Khurana, S. P., and Gaur, R. K. (2019). Complete nucleotide sequence of a new geminivirus isolated from *Vitis vinifera* in India: a symptomless host of *Grapevine red blotch virus*. *Virusdisease* 30, 106–111. doi: 10.1007/s13337-018-0477-x

Massart, S., Chiumenti, M., De Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., et al. (2019). Virus detection by high-throughput sequencing of small RNAs: Large-scale performance testing of sequence analysis strategies. *Phytopathology* 109, 488–497. doi: 10.1094/PHYTO-02-18-0067-R

McGovern, P. E. (2003). *Ancient Wine: The Search for the Origins of Viniculture*. Princeton, NJ: Princeton University Press.

Pantaleo, V., Saldarelli, P., Miozzi, L., Giampetruzzi, A., Gisel, A., Moxon, S., et al. (2010). Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. *Virology* 408, 49–56. doi: 10.1016/j.virol.2010.09.001

Pirovano, W., Miozzi, L., Boetzer, M., and Pantaleo, V. (2015). Bioinformatics approaches for viral metagenomics in plants using short RNAs: model case of study and application to a *Cicer arietinum* population. *Front. Microbiol.* 5:790. doi: 10.3389/fmicb.2014.00790

Pooggin, M. M. (2018). Small RNA-omics for plant virus identification, virome reconstruction and antiviral defense characterization. *Front. Microbiol.* 9:2779. doi: 10.3389/fmicb.2018.02779

Rai, R., Khurana, S. P., Kumar, S., Sharma, S. K., Watpade, S., and Baranwal, V. K. (2017). Characterization of *Grapevine leafroll-associated virus 4* from Indian vineyards. *J. Plant Pathol.* 99, 255–259. doi: 10.4454/jpp.v99i1.3826

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

Sahana, A. B., Adkar-Purushothama, C. R., Chennappa, G., Zhang, Z. X., Sreenivasa, M. Y., and Sano, T. (2013). First report of *Grapevine yellow speckle viroid-1* and *Hop stunt viroid* infecting grapevines (*Vitis vinifera*) in India. *Plant Dis.* 97, 1517–1517. doi: 10.1094/PDIS-05-13-0494-PDN

Seguin, J., Rajeswaran, R., Malpica-Lopez, N., Martin, R. R., Kasschau, K., Dolja, V. V., et al. (2014). *De novo* reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PLoS One* 9:e88513. doi: 10.1371/journal.pone.0088513

Singhal, P., Kumar, S., Rai, R., Singh, S. K., and Baranwal, V. K. (2019). Characterization of viroids infecting grapevine in India. *Indian Phytopathol.* 72, 333–341. doi: 10.1007/s42360-019-00134-9

Tirumalai, V., Swetha, C., Nair, A., Pandit, A., and Shivaprasad, P. V. (2019). miR828 and miR858 regulate VvMYB114 to promote anthocyanin and flavonol accumulation in grapes. *J. Exp. Bot.* 70, 4775–4792. doi: 10.1093/jxb/erz264

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. ıNat. Protoc. 7, 562–578. doi: 10.1038/nprot.2012.016

Xiao, H., Li, C., Al Rwahnih, M., Dolja, V., and Meng, B. (2019). Metagenomic analysis of riesling grapevine reveals a complex virome including two new and divergent variants of *Grapevine leafroll*-associated virus 3. *Plant Dis.* 103, 1275–1285. doi: 10.1094/PDIS-09-18-1503-RE

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using deBruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

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.

Copyright © 2020 Sidharthan, Sevanthi, Jaiswal and Baranwal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 July 2020
doi: 10.3389/fmicb.2020.01235

Insights Into Potato Spindle Tuber Viroid Quasi-Species From Infection to Disease

Charith Raj Adkar-Purushothama, François Bolduc, Pierrick Bru† and Jean-Pierre Perreault**

RNA Group/Groupe ARN, Département de Biochimie, Faculté de Médecine des Sciences de la Santé, Pavillon de Recherche Appliquée au Cancer, Université de Sherbrooke, Sherbrooke, QC, Canada

OPEN ACCESS

Edited by:
John Wesley Randles, The University of Adelaide, Australia

Reviewed by:
José-Antonio Daròs, Polytechnic University of Valencia, Spain
Gerhard Steger, Heinrich Heine University Düsseldorf, Germany

***Correspondence:**
Charith Raj Adkar-Purushothama
charith.adkar@usherbrooke.ca
Jean-Pierre Perreault
jean-pierre.perreault@usherbrooke.ca

†Present address:
Pierrick Bru, Institutionen för Fysiologisk Botanik, Umeå Plant Science Centre, Umeå University, Umeå, Sweden

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

Received: *20 March 2020*
Accepted: *14 May 2020*
Published: *03 July 2020*

Citation:
Adkar-Purushothama CR, Bolduc F, Bru P and Perreault J-P (2020) Insights Into Potato Spindle Tuber Viroid Quasi-Species From Infection to Disease. Front. Microbiol. 11:1235. doi: 10.3389/fmicb.2020.01235

Viroids are non-coding RNA plant pathogens that are characterized by their possession of a high mutation level. Although the sequence heterogeneity in viroid infected plants is well understood, shifts in viroid population dynamics due to mutations over the course of infection remain poorly understood. In this study, the ten most abundant sequence variants of potato spindle tuber viroid RG1 (PSTVd) expressed at different time intervals in PSTVd infected tomato plants were identified by high-throughput sequencing. The sequence variants, forming a quasi-species, were subjected to both the identification of the regions favoring mutations and the effect of the mutations on viroid secondary structure and viroid derived small RNAs (vd-sRNA). At week 1 of PSTVd infection, 25% of the sequence variants were similar to the "master" sequence (i.e., the sequence used for inoculation). The frequency of the master sequence within the population increased to 70% at week 2 after PSTVd infection, and then stabilized for the rest of the disease cycle (i.e., weeks 3 and 4). While some sequence variants were abundant at week 1 after PSTVd infection, they tended to decrease in frequency over time. For example, the variants with insertions at positions 253 or 254, positions that could affect the Loop E as well as the metastable hairpin I structure that has been shown important during replication and viroid infectivity, resulted in decreased frequency. Data obtained by *in silico* analysis of the viroid derived small RNAs (vd-sRNA) was also analyzed. A few mutants had the potential of positively affecting the viroid's accumulation by inducing the RNA silencing of the host's defense related genes. Variants with mutations that could negatively affect viroid abundance were also identified because their derived vd-sRNA were no longer capable of targeting any host mRNA or of changing its target sequence from a host defense gene to some other non-important host gene. Together, these findings open avenues into understanding the biological role of sequence variants, this viroid's interaction with host components, stable and metastable structures generated by mutants during the course of infection, and the influence of sequence variants on stabilizing viroid population dynamics.

Keywords: viroids, quasi-species, high-throughput sequencing, PSTVd, population dynamics, circular RNA, long non-coding RNAs

INTRODUCTION

Viroids are plant pathogenic single-stranded, circular, non-coding RNA molecules composed of 246–401 nt (Ding, 2009). Since viroids are not known to code for any peptides, they rely entirely on their sequence, structure and host factors for their replication and propagation (Flores et al., 2005). Their genome possesses sufficient sequential and structural information to take over both the plant's defense system and its transcriptional machinery to reproduce and spread throughout the host. Upon infection, viroids induce a wide array of symptoms depending on the host plant. However, the degree of viroid induced symptoms depends on the viroid variant and the host cultivar's susceptibility (Owens et al., 2012). To date, 32 viroid species have been classified based on the presence or the absence of the Central Conserved Region (CCR) domain that classifies viroids into two families, the *Pospiviroidae and* the *Avsunviroidae*. The members of the family *Pospiviroidae* (type species: *Potato spindle tuber viroid*; PSTVd) have five structural/function domains such as the Terminal Left (TL), the Pathogenicity (P), the Central (C), the Variable (V), and the Terminal Right (TR) domains. They replicate in the host's nucleus in an asymmetric rolling circle mechanism. Conversely, the members of the family *Avsunviroidae* (type species: *Avocado sunblotch viroid*; ASBVd) lack the CCR, but exhibit self-cleavage via a *cis*-acting hammerhead sequence activity. These members replicate in the chloroplast of the host in a symmetric rolling circle mechanism (Adkar-Purushothama and Perreault, 2020).

Since viroids are non-coding RNA pathogens, they recruit host DNA dependent RNA polymerase during replication. Specifically, members of the family *Avsunviroidae* use nuclear-encoded polymerase (NEP), whereas the members of family *Pospiviroidae* use DNA dependent RNA polymerase II (Flores et al., 2011). Under normal conditions, the NEP and the DNA dependent RNA polymerase II use DNA as template to generate RNA molecules. Viroids redirect the NEP and the DNA dependent RNA polymerases to use the viroid's RNA as template instead of DNA. As a consequence, the replication becomes error prone (Flores et al., 2005). Analysis of chrysanthemum chlorotic mottle viroid (CChMVd) sequences retrieved from the plant infected with a single variant of CChMVd, revealed a mean error rate of 2.5×10^{-3} per nucleotide position. In other words, the 399 nt long CChMVd had one error for every 400 nt, the highest reported mutation rate for any given biological species (Gago et al., 2009). High-throughput sequence analysis of peach latent mosaic viroid (PLMVd) recovered from the peach tree inoculated with a single variant of PLMVd revealed the presence of 3,939 sequence variants (Glouzon et al., 2014). The data showed that most of the sequence variants had an average of 4.6–6.4 mutations as compared to the initially inoculated, or master, sequence. On the other hand, analysis of PSTVd-derived small RNA (PSTVd-sRNA) by deep-sequencing of PSTVd-infected plants revealed that the mean error rate per nucleotide was less than 5×10^{-3}, which lies within the range observed for members of the family *Avsunviroidae* (Brass et al., 2017). Additionally, analysis of the mutation rate for the chloroplast replicating viroid, eggplant latent viroid (ELVd) and nuclear replicating PSTVd in a common host revealed higher mutation frequencies in ELVd than in PSTVd (López-Carrasco et al., 2017). The sequence variants created by the replication of a master sequence during the course of replication are called a "quasi-species" or "viroid cloud" or "mutant swarm" (Eigen, 1971; Domingo, 2002; Glouzon et al., 2014; Domingo and Perales, 2019).

Viroids, as a minimal pathogen consisting of a naked circular single-stranded RNA molecule, are dependent on their thermodynamically stable structure and nucleotide sequence for their survival. Viroids rely on both thermodynamically stable, as well as metastable, secondary structures to interact with host components for their biological functions (Repsilber et al., 1999; Flores et al., 2012). Recently, direct visualization of the native structure of PLMVd at a single-molecule resolution using atomic force microscopy confirmed the stabilizing role of tertiary structures such as kissing-loop interactions (Moreno et al., 2019). Furthermore, it has been demonstrated that the kinetically preferred metastable structure containing hairpin I (HPI) and hairpin II (HPII) of PSTVd is crucial for both its replication and infectivity (Hammond and Owens, 1987; Loss et al., 1991; Gas et al., 2007). However, studies have also shown that certain changes in the nucleotide sequence of the viroid's RNA can be associated with the disease severity (Rodriguez and Randles, 1993; Semancik and Szychowski, 1994; De la Peña et al., 1999; Schnell et al., 2001; De la Peña and Flores, 2002; Malfitano et al., 2003; Maniataki et al., 2003; Tsushima et al., 2016; for a review see Adkar-Purushothama and Perreault, 2020). These findings, along with the detection of viroid-derived small RNAs (vd-sRNAs) in the PSTVd infected plants implicated RNA interference (RNAi, also known as RNA silencing) in viroid pathogenicity (Itaya et al., 2001; Papaefthimiou et al., 2001; Martinez de Alba et al., 2002). Due to their highly base-paired secondary structures, viroids trigger the host's RNA silencing machinery (Pallas et al., 2012). Although the vd-sRNAs are active in guiding the RNA-induced silencing complex (RISC)-mediated cleavage, mature PSTVd is partly resistant to the RISC-mediated cleavage due to its secondary structure (Itaya et al., 2007; Sano et al., 2010; Adkar-Purushothama et al., 2015b; Dalakouras et al., 2015). Several studies using various viroid–host combinations demonstrated the down-regulation of the host's mRNA by its direct interaction with the vd-sRNA (Navarro et al., 2012; Eamens et al., 2014; Adkar-Purushothama et al., 2015a, 2017; Zheng et al., 2017; Adkar-Purushothama and Perreault, 2018). Previously, it has been shown that changes in a few critical nucleotides in the seed region of the vd-sRNA/mRNA complex has substantial effects on the efficiency of RNA silencing of the target mRNA (Adkar-Purushothama et al., 2015a). For example, changing two nucleotides of the seed region of the vd-sRNA of PSTVd-I targeting the *callose synthase 11-like (Cals11-like)* mRNA to that of PSTVd-M negatively affected the down-regulation of *Cals11-like* mRNA and, consequently, the viroid induced symptoms (Adkar-Purushothama et al., 2015a). Recently, analysis of polysome fractions of viroid infected plants revealed the direct interaction of viroid and vd-sRNA with the host's translation machinery, thus inducing ribosomal stress in the host plant (Cottilli et al., 2019). However, given the fact that viroids are quasi-species, it is not clear whether or not the interaction

detected with the host translation machinery is specific to the viroid's structure or it's sequence. Hence, it is crucial to understand the specific structures and sequence variants of a viroid that are involved in the different stages of infection.

Previously it has been noted that PSTVd follows the S-curve in diseased plants. More specifically, after inoculation, PSTVd RNA showed a brief lag phase that was followed by a sudden increase in its titer before the stationary phase and finally a drop in its titer (Adkar-Purushothama and Perreault, 2018). As viroids depend solely on their structure and their sequence for establishing themselves and for conquering the plant's defense mechanism, a shift in viroid sequence dynamics during the course of infection was suspected. To date, different approaches have been used to study viroid's quasi-species nature in viroid infected plants (Gago et al., 2009; Glouzon et al., 2014; Brass et al., 2017; López-Carrasco et al., 2017). However, all of these experiments were restricted to a single point sample collection, and much of the work was focused on estimating the mutation rate. Hence, in this present study, samples of PSTVd infected tomato plants were collected at different time intervals and subjected to deep-sequencing and computational analysis to understand: (i) the evolution of the viroid quasi-species during the course of infection; (ii) the effect of these mutations on stable secondary structures; and, (iii) any changes in their vd-sRNA – host target specificity. The findings provide more insights on how the composition of a viroid quasi-species of sequences changes during the course of infection.

MATERIALS AND METHODS

PSTVd Constructs and Bioassays

The dimeric construct of PSTVd-RG1 (GenBank Acc. No. U23058) was previously inserted in the pBluescript KS + vector (Stratagene) and was then ligated into the binary vector pBIN61 (Adkar-Purushothama et al., 2017). Briefly, the dimeric PSTVd-RG1 construct was purified after being subjected to digestion with the restriction endonucleases *Xba*I and *Bam*HI and was then ligated into the same sites of the binary vector pBIN61. The resulting recombinant binary vector was then transformed into the *Agrobacterium tumefaciens* (*A. tumefaciens*) strain GV3101 as previously described (Adkar-Purushothama et al., 2017).

Tomato seedlings (*Solanum lycopersicum* cv Rutgers; Livingston Seed Co.) were used for the bioassays. All plants were grown in a chamber at 25°C with 16 h light and 8 h darkness. The primary leaves were agro-infiltrated with the *A. tumefaciens* strain GV3101 containing the dimeric construct of PSTVd-RG1 in the binary vector pBIN61 whereas empty vector was used for mock infection. Since, *A. tumefaciens* strain GV3101 and binary vector pBIN61 are not known to move systemically, the agro-infiltrated leaves were excised at 3 days of post-inoculation to avoid continuous generation of master sequence from the vector. Upper non-inoculated, fully opened single whole leaf samples were collected at 1, 2, 3, and 4 weeks post-inoculation (wpi) from the same 3 plants for RNA preparation. In other words, every week the next youngest leaf above the previously sampled leaf was collected for RNA extraction.

RNA Preparation, RT-PCR, and RNA Gel Blots

Total RNA from infected leaf samples was extracted using the mirVana miRNA isolation kit (Ambion) as described previously (Adkar-Purushothama et al., 2018). Briefly, 150 mg of leaf sample was ground with 600 μL of lysis/binding buffer in the presence of sand. Then, 70 μL of microRNA homogenate was added and the sample was mixed by vortexing followed by incubation on ice for 10 min. Total RNA was isolated by acid phenol-chloroform (5:1) extraction and was precipitated by centrifugation after adding 2.5 volumes (vol.) of absolute ethanol. The RNA was further purified by DNase I (Promega) treatment. RNA integrity was examined in a 2100 Bioanalyzer (Agilent Technologies). RNA obtained for each sample was analyzed by RT-PCR assay to detect viroid presence (Adkar-Purushothama et al., 2015a). Equal amounts of RNA obtained for each week were pooled together for the RNA gel blot assay and the cDNA library preparation.

To detect PSTVd, dimeric (−) PSTVd-RG1 riboprobes were prepared as described previously (Adkar-Purushothama and Perreault, 2018). For the RNA gel blot hybridizations, 500 ng of the total RNA samples pooled for each week were denatured at 65°C for 15 min with 3 vol. of sample buffer [50% formamide, 2.2 M formaldehyde (37%), and 1x MOPS], and were then separated by electrophoresis on 1.0% agarose-formaldehyde gels containing 1x MOPS buffer. The RNAs were transferred to Hybond-XL nylon membrane (Amersham, GE Healthcare Life Sciences) and hybridized with radiolabeled probes as described before (Adkar-Purushothama et al., 2018). Radiolabeled DNA probes specific for 5S rRNA used to evaluate the expression level of 5S rRNA, which was used as the loading control for the RNA gel blot assay (Adkar-Purushothama and Perreault, 2018).

Viroid Library Preparation and High-Throughput Sequencing

For the viroid library preparation, circular PSTVd molecules were purified from the pooled total RNA of each week by separating 10 μg of total RNA by 5% polyacrylamide-8 M urea denaturating gel electrophoresis (5%-dPAGE). The circular PSTVd (cPSTVd) molecules were eluted from the gel and precipitated with 2.5 vol. of ethanol. To prepare the two complementary DNA (cDNA) libraries, 1 μg of purified cPSTVd RNA was separately subjected to reverse transcription (RT) using a primer binding at either position 280–269 (L1) or 10–354 (L2) of PSTVd-RG1, respectively, for each week. The resulting cDNA libraries were amplified by polymerase chain reaction (PCR) with Q5 High-fidelity DNA polymerase enzyme (New England Biolabs) in the presence of primers F1/R1 for L1 and F2/R2 for L2 followed by indexing of each library. The 8 libraries (2 libraries per week × 4 weeks) were sequenced using the Illumina MiSeq sequencer at the Laboratoire de Génomique Fonctionnelle de l'Université de Sherbrooke[1]. All the primers used in this study are listed in **Supplementary Table 1**. The deep sequence data generated in this study is deposited in the Gene Expression Omnibus under accession number GSE147577.

[1]http://palace.lgfus.ca

Bioinformatic Analysis

Workflow was divided four main steps, specifically steps A–D (**Figure 1**). Each step was performed independently for both libraries except for step C. All steps were implemented using a combination of Mothur commands and in-house Perl programming (Schloss et al., 2009; Kozich et al., 2013). Step A consisted of creating the library fragments by aligning the read pair in their overlapping parts. Initially, the adapter sequence were trimmed using the Trimmomatic tools (Bolger et al., 2014) and the resulting fragments were filtered against sequence quality metrics, overlapping ambiguities and length criteria. Step B consisted of removing the sequence parts that were targeted by the PCR primers for amplification to mask the PCR primer binding regions before evaluating any redundancy between the two libraries. Step C was the evaluation of the redundancy between the fragments of both libraries. A sequence was retained if it was present in both libraries. As described in step B above, the PCR primer regions were not considered in step C. Finally, step D included three sub-steps: (i) the extraction; (ii) the occurrence filtering; and, (iii) the rotation. Based on the redundancy set obtained in step C, all of the sequences that were present in both the libraries were extracted. Thus, the extracted repetitive sequences were filtered to obtain those sequences that were found at least 10 times in one of the libraries. Then, these sequences were aligned against the master sequence to identify the sequence variation.

RESULTS

Infection Assays and High-Throughput Sequencing of the PSTVd Quasi-Species

To assess the evolution of a viroid during the different stages of infection in the host plant, PSTVd-RG1 and tomato plants were used as a model system as this viroid-host interaction is well studied (Adkar-Purushothama et al., 2017; Adkar-Purushothama and Perreault, 2018). The flow chart outlining the steps used to create the viroid libraries is shown in **Figure 2A**. Briefly, leaf samples were collected from tomato plants that were agro-infiltrated with the plasmid containing a dimeric head-to-tail construct of the PSTVd-RG1 variant. At 1–4 wpi, total RNA was extracted from harvested leaves. The presence of PSTVd in agro-infiltrated plants was verified by RT-PCR, as well as by RNA gel blot assay using radiolabeled PSTVd probes. To develop a full-length PSTVd library, cPSTVd was purified from the total RNA extracted from the PSTVd infected plants by electrophoretic migration under denaturation condition and, was then amplified by RT-PCR amplification using two distinct sets of primers binding at different locations on the PSTVd molecule. Library 1 (L1) was prepared using primers binding at position 260–295 (F1/R1) of PSTVd-RG1, while library 2 (L2) was prepared using primers binding at position 354–30 (F2/R2) of PSTVd-RG1 (**Figure 2B**). Construction of the two libraries provided information on the entire whole viroid genome, including the primer binding sites. Illumina indexed primers were used to identify the origin of each sequence with respect to both the library and the number of weeks post-infection.

The tomato plants were agro-infiltrated with the plasmid containing a dimeric head-to-tail construct of PSTVd-RG1 (hereafter named the master sequence) in *A. tumefaciens*. Disease symptoms such as mild stunting and leaf curling appeared during the 2nd week of infection, followed by severe necrosis in week 3 (**Figure 2C**). In contrast, mock inoculated control tomato plants did not show any disease symptoms. The presence of PSTVd in infected plants was confirmed by RT-PCR reactions performed on RNA extracted from leaf samples at 1, 2, 3, and 4 wpi using the VidR1/VidF1 primers as described previously (Adkar-Purushothama et al., 2015a). Equal amounts of RNA from three plants were pooled for each week, and the amount of PSTVd present was analyzed by RNA gel blot assay (**Figure 2D**). Analysis of the accumulation of PSTVd after normalization against the 5S rRNA revealed that the PSTVd titer increased initially from week 1 to 3, and that it then decreased in week 4 (**Figure 2E**). This variation is in good agreement with previously published data (Adkar-Purushothama and Perreault, 2018).

To evaluate the accumulation of PSTVd-RG1 sequence variants over the time course of the infection, cPSTVd was purified by 5% polyacrylamide – 8 M urea gel electrophoresis followed by the preparation of both libraries as described above. Both libraries were paired end sequenced using an Illumina MiSeq sequencer, and 300 bp reads were filtered using bioinformatics tools (see section "Materials and Methods"). Thus, the sequence of each variant, as well as the number of times it was sequenced, was known for each week. In the analysis of the data, the sequence variants that were detected at least ten times in both libraries for a given week were retained. In the case of L1, 1,559 sequence variants were found at least 10 times out of 639,187 total sequences, while for library L2, 2,299 sequence variants out of 677,758 total sequences found at least 10 times. The total number of sequences retrieved in L1 and L2 for each week is presented in **Figure 2F**. Finally, each sequence variant was aligned to the master sequence (PSTVd-RG1). The most distant sequence variant of the L1 exhibited 95.4% identity to the master sequence, while in the case of the L2 it was 93.9%. The higher mismatch observed for L2 is due to the single sequence variant that had several mutations. Excluding this sequence variant from the analysis resulted in having almost the same sequence similarity in both the libraries.

Mutations That Either Decrease or Increase Accumulation

To identify the different mutations that might explain the adaptation of PSTVd to its host, the sequence variants were analyzed and sorted according to their abundance each week. To reduce the number of variants to be analyzed, the ten most abundant variants for each week were retained. In addition, the proportion of each sequence variant decreases considerably beyond the ten most abundant variants, thus making the biological importance of the infrequent variants relatively insignificant. This observation is particularly true for weeks 2–4. At week 1, a greater dispersion of the variants was

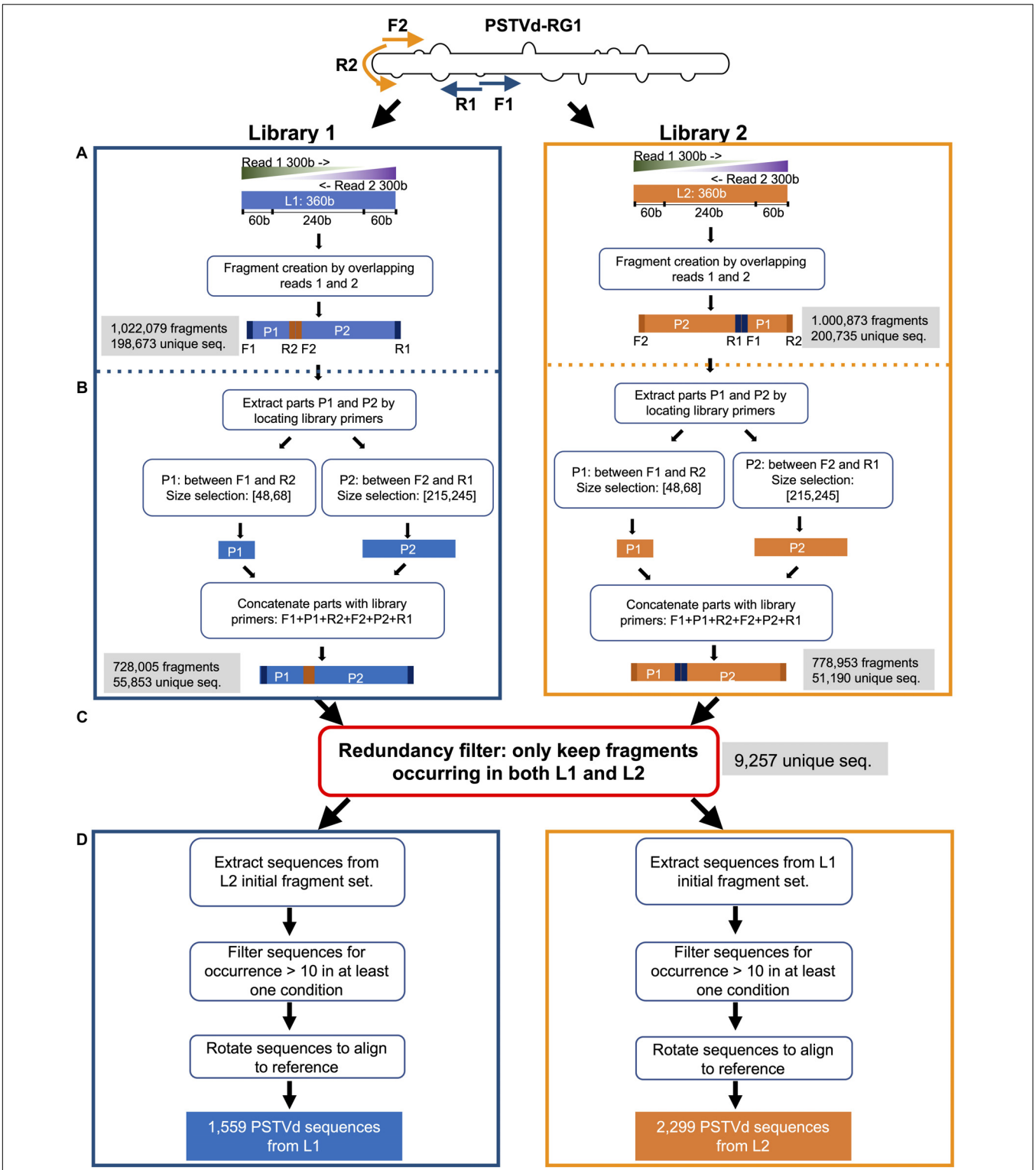


FIGURE 1 | Scheme of the steps followed for the treatment of the high-throughput sequence data. **(A)** Library fragments were created by aligning read pairs by their overlapping sequences after trimming of the sequence adapters and were then filtered using sequence quality metrics. **(B,C)** The resulting sequences were extracted and evaluated for sequence redundancy between libraries 1 (L1) and 2 (L2). **(D)** The redundant sequences were extracted, and those that occurred at least ten times were filtered and aligned against the master sequence (PSTVd-RG1). In the figure, F1 and R1 indicate the forward and reverse primer used for the preparation of L1, as well as their respective binding sites on the PSTVd molecule; F2 and R2 denotes the forward and reverse primer used for the preparation of L2, as well as their respective binding site on the PSTVd molecule; P1 denotes the region located between F1 and R2; P2 denotes the region located between F2 and R1.

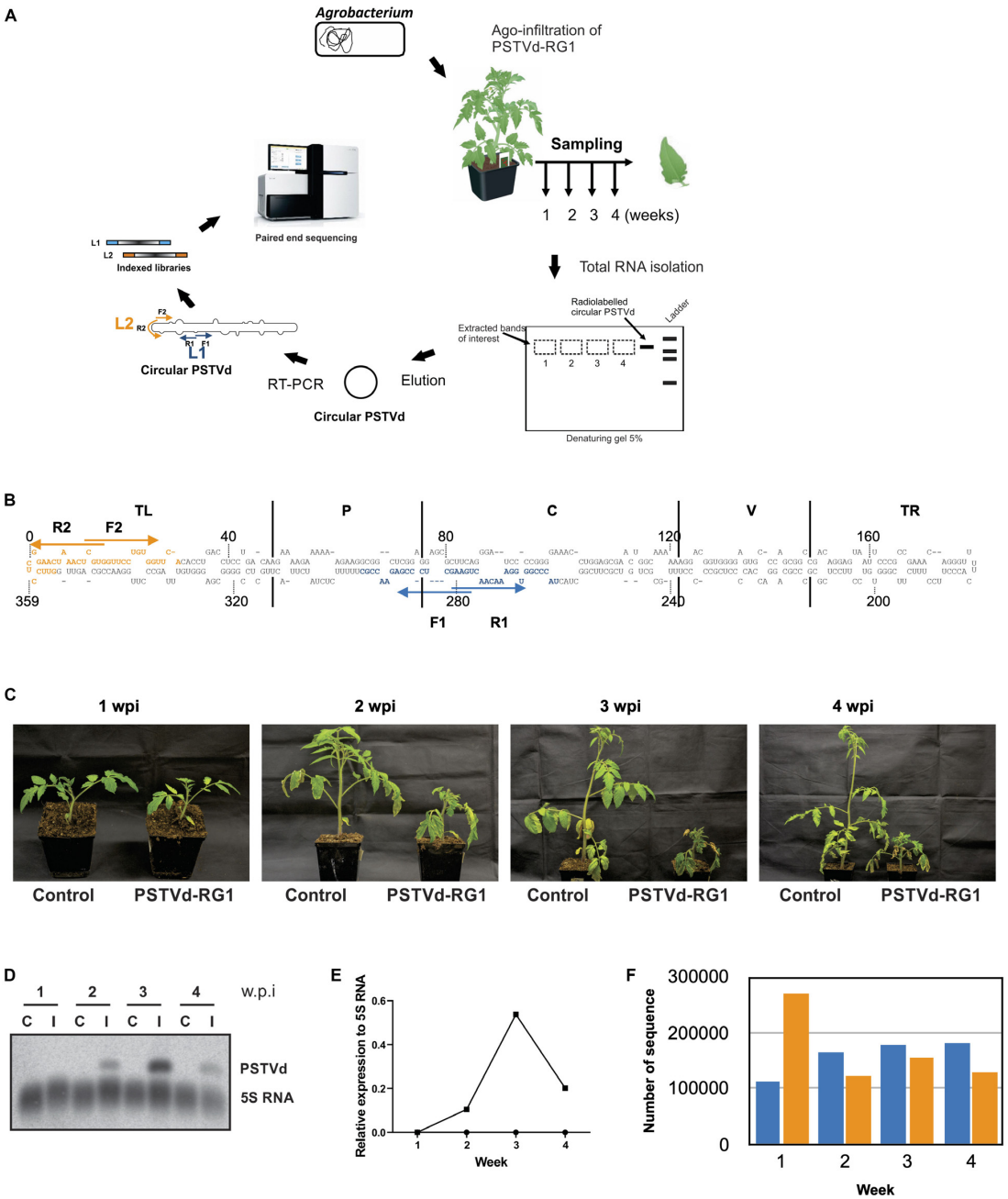


FIGURE 2 | Experimental strategy and analysis of the viroids recovered from the PSTVd-RG1 infected tomato plants. **(A)** Flow diagram outlining the steps involved in the infection assay, sample preparation and viroid high-throughput sequencing from the PSTVd-RG1 infection in tomato plants. *Agrobacterium tumefaciens* (*A. tumefaciens*) strain GV3101 was transformed with recombinant binary vector pBIN61 containing the dimeric construct of PSTVd-RG1. Primary leaves of tomato plants (cv. Rutgers) which were maintained in a chamber were agro-infiltrated with the transformed *A. tumefaciens*. Plants inoculated with empty vector were used for mock infection. Total RNA was extracted from the upper, non-inoculated leaf samples which were harvested at 1–4- weeks post-infection (wpi). After verifying the presence of PSTVd, circular PSTVd (cPSTVd) was isolated through polyacrylamide gel electrophoresis under denaturing conditions. Purified cPSTVd RNA was amplified by RT-PCR using two distinct sets of primers binding at different locations on the PSTVd molecule to construct two libraries namely, Library 1 (L1) and Library 2 (L2). The resulting 8 libraries (2 libraries per week × 4 weeks) were indexed and sequenced using the Illumina MiSeq sequencer. **(B)** Nucleotide sequence and secondary structure of PSTVd-RG1. The region spanning the primers used for amplification are shown in blue and orange, respectively, for libraries 1 and 2 (i.e., L1 and L2). The arrows indicate the primer binding sites in the 5′–3′ orientation. The structural/functional domains of PSTVd: Terminal left (TL), Pathogenicity (P), Central (C), Variable (V), and Terminal right (TR), are delimited by the vertical solid lines and are named accordingly. **(C)** Photos of tomato plants agroinfiltrated at the two leaf stage with *Agrobacterium* harboring, or not (control), a plasmid including a dimeric head-to-tail construct of PSTVd-RG1. **(D)** Autoradiogram of a RNA-gel blot hybridization for the detection of PSTVd and the 5S RNA (as loading control) in PSTVd-infected tomato plants. **(E)** Relative expression level of PSTVd compared to that of 5S rRNA. **(F)** Histogram of the total number of full-length PSTVd reads obtained by high-throughput sequencing for both the L1 (blue color) and the L2 (orange color) libraries.

observed (**Supplementary Figure 1**). Certain sequence variants are found in the top ten for both libraries, while others are present in one. It is important to note that there is some variability in the proportions of the different variants since the primers used for the construction of two libraries do not bind to the same viroid regions. Hence, there may be different binding constraints depending on the context, that result in better or worse PCR amplification (i.e., secondary and tertiary structures, thermodynamic parameters, etc.). To ensure that real true variants that will allow drawing of solid conclusions about the adaptability of PSTVd to its host were identified, it was further decided to keep the top ten most abundant variants found in both libraries. The variants that contain mutations associated with a decrease in accumulation are listed in **Table 1** while those containing mutations associated with an increase in accumulation are listed in **Table 2**. Analyzing the sequence variants based on their abundance through time can reveal the importance of some mutations and could explain adaptation or fitness. When analyzing the evolution of the master sequence, in either library L1 or L2, its abundance increased from 22 to 25% in week 1 and quickly reached 72–77% at week 2. Subsequently, the abundance remained steady at about 70–72% in weeks 3 and 4 (**Supplementary Figure 1** and **Table 1**). The master sequence seems to be well adapted to its host since it represents ∼70% of the quasi-species.

However, if one looks beyond the master sequence and focuses on the other sequence variants, it can be seen that some of them increase in abundance over time while others decrease. Thus, it becomes possible to identify mutations that favor adaptation and others that repress it. By analyzing the behavior of some variants, especially at week 1, it was observed that most of the sequences that are found disappear quite quickly over time, and that this is true for both libraries (**Table 1**). For example, Insertion C253 (Ins:C253), Ins:G254, Ins:U254, and Ins:A254 have a high abundance in week 1 as compared to week 4. The case of Ins:C253 is particularly interesting since it represents about 20–24% of the sequences of the quasi-species at week 1, that is to say almost the same level as master sequence at week 1, and represents about 0.5% at week 4. Although this proportion is still high at week 4 (corresponding to the fifth most represented sequence), it represents a decrease of 26- or 45-fold depending on the library. Even if the abundance of the other three mutants (Ins:G254, Ins:U254, and Ins:A25) is smaller, the observed decreases remain at the same order of magnitude, specifically, 21–28-fold in L1 and 41–48-fold in L2. Other sequence variants detected at week 1 contain insertions or deletions in the region around positions 250–255 of the PSTVd-RG1 genome, and in all cases, result in a loss of adaptation and/or fitness, indicating that this region is quite sensitive to any changes in the nucleotide sequence (**Supplementary Figure 1**).

To identify mutations that favor adaptation, attention was focused on the variants most frequently found at week 4 (**Table 2**). At this stage of the infection, the variants that are found in greater proportion should normally be the best adapted. As previously described, the master sequence occupies the first place with a proportion of 70–72%. In-depth analysis of both the libraries revealed five sequence variants whose frequency increased during the course of infection (**Supplementary Figure 1** and **Table 2**). These are the variants Ins:A55, Deletion A55 (Del:A55), Substitution G168A (Sub:G168A), Del:U296, and Del:A118. For all of these variants, there is an enrichment rate of about 2-fold, regardless of the library. The variant Del:A55 is an exception with a rate of about 95-fold. Unlike the mutations associated with a decreased abundance, the mutations associated with an increased abundance are scattered through the entire genome. All of these mutations of interest are displayed on the genomic structure of PSTVd-RG1 (**Figure 3**).

Structural Analysis of the PSTVd Quasi-Species

To understand the importance of the mutations described above, their effects on the structures of the viroids were evaluated. The different sequence variants were submitted to the structure prediction software RNAstructure 6.1 and the resulting structures were compared with that of the PSTVd-RG1 (**Figure 4**). Variation of the nucleotide sequence by insertion, deletion or substitution resulted in changes in the structures. For example, a change at position 55 resulted in either an increased or a decreased size of the loop located in the upper P domain of PSTVd (**Figure 4A**). Similarly, the insertion or deletion of adenosine at position 118 resulted in either an increased or a reduced size of the loop located in the central region of PSTVd molecule (**Figure 4B**). Substitution at position 168 (Sub:G168A) resulted in an increased loop size for the loop located in TR region of the PSTVd (**Figure 4C**). Insertions at position 253 and 254 affected structure of both the Loop E and HPI of PSTVd, the regions that are known to be involved in viroid infectivity and replication (**Figures 4D,E**) (Hammond and Owens, 1987; Gas et al., 2007). The deletion of uracil at position 296 introduced a loop in the lower P domain of PSTVd and increased the size of loop A (**Figure 4F**), while the insertion of uracil resulted in a longer helix and a smaller loop A as compared to PSTVd-RG1. Some mutations altering the nucleotide base-pairing on PSTVd secondary were also observed. For example, the mutation at position 353 from G to A resulted in a more stable helical structure by changing the G:U to A-U base pairs. The deletion of the guanosine residue at position 353 resulted in an increased size of a loop located in the upper TL domain of the PSTVd molecule. These data illustrate the minor structural changes in PSTVd-RG1 that result from the mutations that occur during the course of the infection.

Potential Impact on the Host Genes Targeted by Viroid Derived Small RNAs

RNA silencing is a potent antiviral defense mechanism in eukaryotes directed against invading double-stranded and/or highly structured RNA pathogens such as viruses and viroids. Upon infection, PSTVd-RG1 produces vd-sRNAs of 21–24 nt that are located throughout its genome (Adkar-Purushothama et al., 2017; Adkar-Purushothama and Perreault, 2018). Previously, it has been shown that the vd-sRNAs derived from both the (+) and (−) strands of PSTVd are capable of down-regulating the expression of host mRNAs that are involved in various functions such as defense, development and

TABLE 1 | Abundance (%) of sequence variants common to both libraries for each week that bear mutations associated with a reduced accumulation.

Variants		Week 1	Week 2	Week 3	Week 4	Rate of decrease* (times)
Ins:C253	L1	20.1103	0.1045	0.0408	0.7745	26.0
	L2	24.7860	0.1622	0.0733	0.5159	48.0
Ins:G254	L1	5.3295	0.0224	0.0056	0.2507	21.3
	L2	8.1040	0.0468	0.0193	0.1944	41.7
Ins:U254	L1	1.3632	0.0073	0.0039	0.0487	28.0
	L2	1.4970	0.0121	0.0051	0.0333	45.0
Ins:A254	L1	1.3605	0.0091	0.0022	0.0602	27.6
	L2	2.8546	0.0202	0.0064	0.0643	44.4

**Rate of decrease: abundance at week 1/abundance at week 4.*

TABLE 2 | Abundance (%) of sequence variants common to both libraries for each week that bear mutations associated with an increased accumulation.

Variants		Week 1	Week 2	Week 3	Week 4	Rate of increase* (times)
Ins:A55	L1	0.5989	0.9771	1.0361	1.1314	1.9
	L2	0.5156	0.7862	0.8788	0.6088	1.2
Del:A55	L1	0.0447	0.5511	0.9814	4.2924	96.0
	L2	0.0438	0.5085	0.8994	4.1564	94.9
Sub:G168A	L1	0.1189	0.3843	0.2864	0.2567	2.2
	L2	0.1148	0.2978	0.2957	0.2603	2.3
Del:U296	L1	0.2128	0.6236	0.6481	0.5994	2.8
	L2	1.9319	6.1942	6.4775	4.9930	2.6
Del:A118	L1	0.1833	0.5856	0.4376	0.3842	2.1
	L2	0.2006	0.5037	0.5246	0.4926	2.5

**Rate of increase: abundance at week 4/abundance at week 1.*

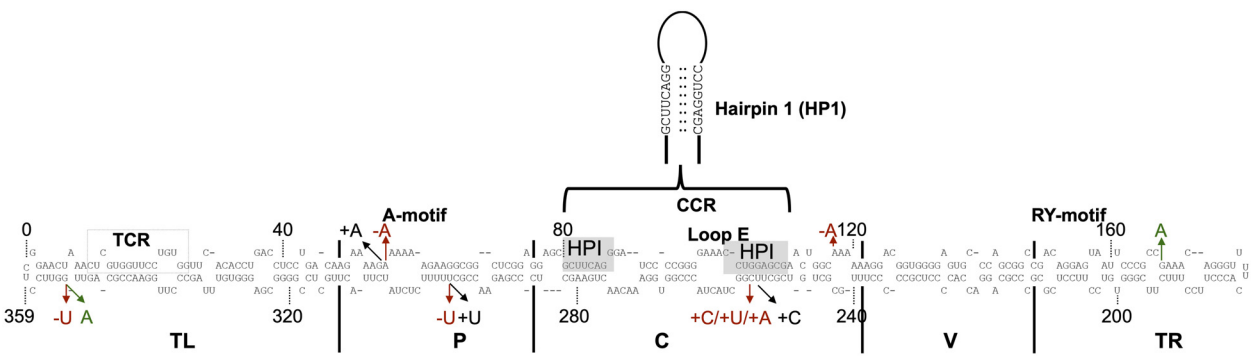


FIGURE 3 | Sequence mutations of the ten most retrieved PLMVd variants. The nucleotide sequence and secondary structure of PSTVd-RG1. The structural/functional domains of PSTVd: TL, P, C, V, and TR are delimited by the vertical solid lines and are named accordingly. The A-motif, Central Conserved Region (CCR), loop E, and RY-motif are indicated. Additionally, the sequence that form the metastable structure hairpin I (HPI) via pairing of distant complementary nucleotides sequences is enlarged. The arrows indicate the positions where mutations occurred. Nucleotide insertion, deletion and substitution are shown in black, red, and green, respectively.

reproduction through RNA silencing (Eamens et al., 2014; Adkar-Purushothama et al., 2015a, 2017, 2018; Adkar-Purushothama and Perreault, 2018). To examine the presence of mutations in previously verified vd-sRNA that are known to target host mRNAs, all of the PSTVd-sRNAs that are known to target host genes were mapped on the secondary structure of PSTVd-RG1. In the highly expressed PSTVd-RG1 variants, the vd-sRNA sequences that target the *FRIGIDA-like protein 3* (*FRL3*) mRNA, the *serine/threonine-protein kinase At1g01540-like* (*STPK At1g01540-like*), the *chloride channel protein CLC-b-like,* the *putative receptor-like serine/threonine-protein kinase At5g57670* (*RSTPK At5g57670-like*), the *vacuole membrane protein 1-like* (*VMP1-like*) mRNA and the *pentatricopeptide repeat-containing protein At2g17033* (*PPR*) mRNAs were not altered. However, either a deletion or insertion of an adenosine at position 55, which has been observed in some variants, alters the sequence

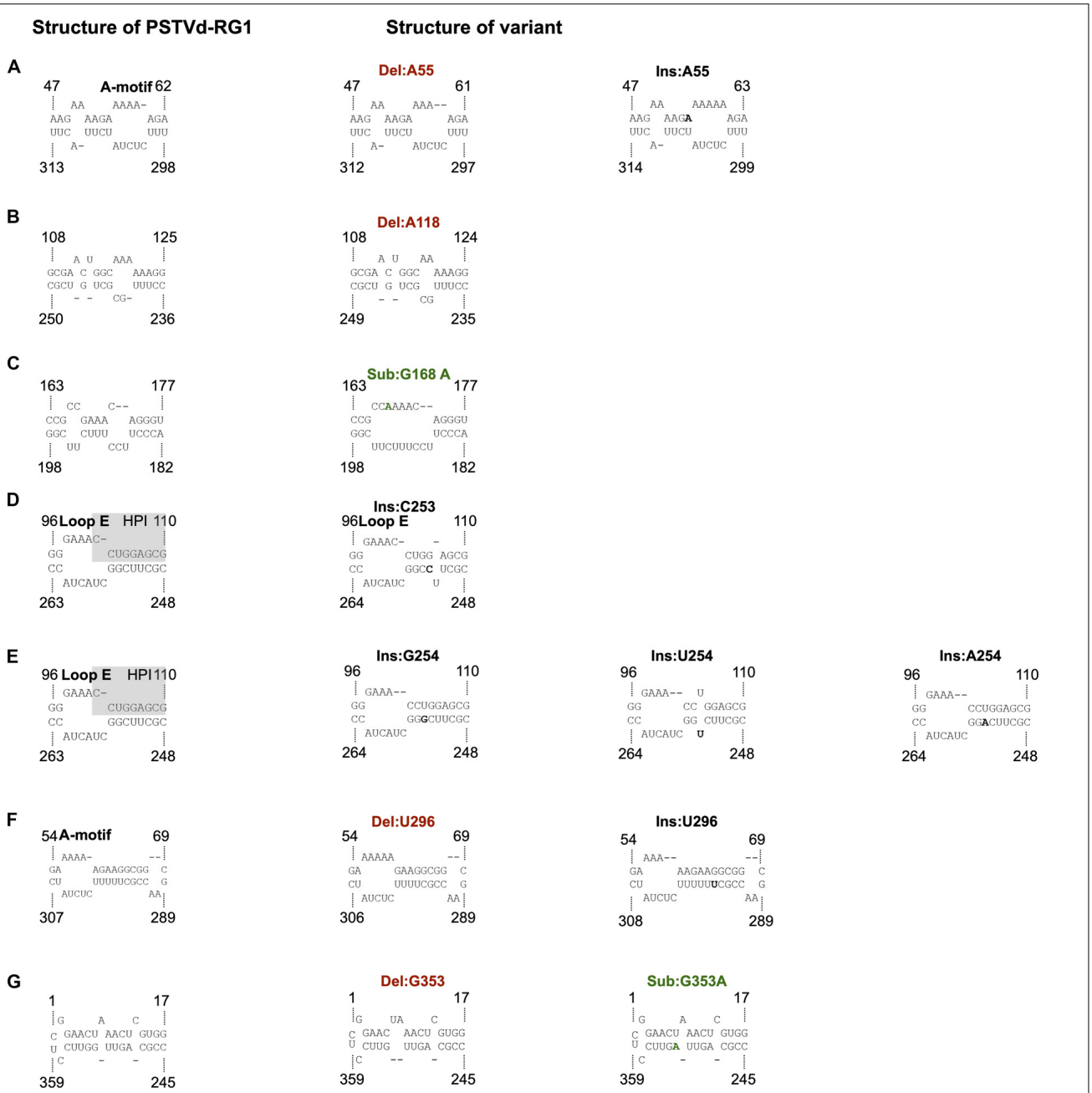


FIGURE 4 | Impact of the mutations on the structure of PSTVd-RG1. **(A)** The deletion or insertion of an adenosine at position 55 results in a decrease and an increase in the size of the A-motif, respectively. A nucleotide in bold indicates an insertion. **(B)** The deletion of the adenosine at position 118 reduces the size of the loop. **(C)** The substitution of adenosine for the guanosine located at position 168 (in green) makes a single larger loop in the TR domain instead of the 2 smaller ones that are present in the wild type. **(D)** The structure of the hairpin I (HPI) is affected by an insertion of one nucleotide (in bold) at position 253. **(E)** The structure of both the hairpin I (HPI) and Loop E are affected by an insertion of one nucleotide (in bold) at position 254. **(F)** The deletion or insertion of a uracil at position 296 resulted in either an increase or a decrease in the size of the A-motif. **(G)** The deletion of the guanine located at position 353 increases the size of the loop in the TL domain, whereas the substitution G353A is a covariation (i.e., G:U base pair becomes an A:U base pair). Nucleotide insertion, deletion and substitution are shown in black, red, and green, respectively.

of the vd-sRNA that was shown to induce the RNA silencing of the *CalS11-like* mRNA, the *CalS12-like* mRNA, the *40S ribosomal protein S3a-like (40S RPS3a-like)* mRNA, *the putative leucine-rich repeat receptor-like serine/threonine-protein kinase At2g24230-like (LRR-RSTPK At2g24230-like)* mRNA *and* the *phosphatidylinositol 4-kinase alpha* (*PI4KA*) mRNA of tomato (**Figure 5A**). Since the vd-sRNA derived from this region of PSTVd is involved in targeting different host defense genes,

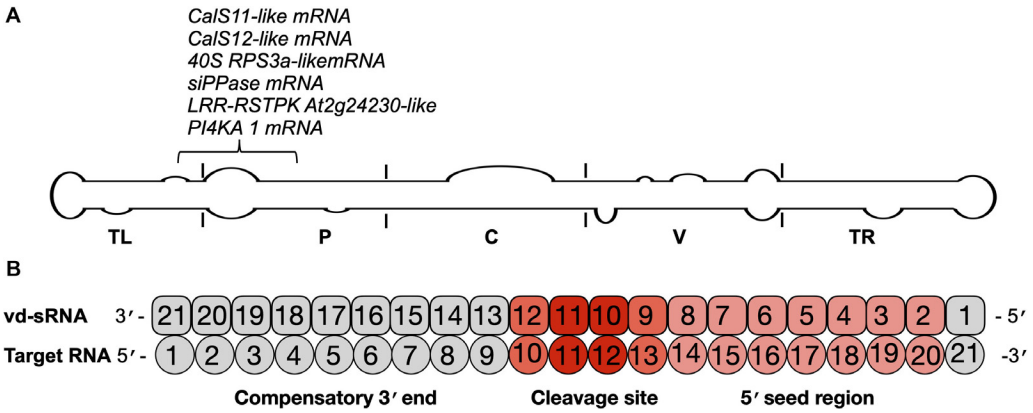


FIGURE 5 | Identification of the mutations on the vd-sRNA known to target host genes. **(A)** The mutations observed in this study were analyzed according to the regions of previously examined vd-sRNAs that are known to downregulate host genes through RNA silencing. These mutations are shown on the potential secondary structure of PSTVd-RG1, and their target genes are indicated. The structural/functional domains of PSTVd are delimited by the vertical solid lines and are named accordingly. Out of all the mutations, the mutation located at position 55 of PSTVd-RG1 showed variation in the vd-sRNA that is known to target the host genes which are involved in defense, development and translation. **(B)** Schematic representation showing the predicted vd-sRNA:target duplex.

changes in this vd-sRNA could severely affect the fitness and adaptability of the viroid.

For the efficient downregulation of a target mRNA through RNAi-mediated silencing, a strong sequence complementarity is required in both the 5′ seed region of the targeting miRNA and across the cleavage site (Ossowski et al., 2008). To understand the impact of mutants on the host genome, the 21 nt long vd-sRNAs derived from the mutant PSTVd-RG1 sequences were dissected in such a way that the mutant nucleotide within the cleavage site of the vd-sRNA:target duplex (i.e., in either the 10th or 11th position of the vd-sRNA:target duplex; **Figure 5B**). The resulting vd-sRNAs were then used to interrogate publicly available tomato transcriptome data sets using the WMD3 Web-based tool[2]. The total number of detected targets for each vd-sRNA is given in **Table 3**. The Gibbs free energy (ΔG), represented as ΔG, for each vd-sRNA:target duplex was calculated using the PairFold software (Andronescu et al., 2003). Although previous *in planta* studies revealed an RNA silencing at ΔG -16.21 (Adkar-Purushothama et al., 2015a), in the present *in silico* analysis, the vd-sRNA:target having a ΔG lower than -20 kcal/mol were grouped and kept for further analysis. The resulting putative target sequences were analyzed by BLAST, and the mRNAs coding for proteins with known functions were sorted (**Table 4**). The results revealed that the majority of these mutations on PSTVd-RG1 substantially changed the targets of the vd-sRNAs. For instance, the PSTVd-RG1-sRNA derived from the region spanning positions 44–64 (5′-ACAAGAAAAGAAAAAAGAAGG-3′) is known to target the *Pto kinase interactor* 1 (*Pti* 1) which is involved in the *serine-threonine kinase* involved in the hypersensitive response (HR)-mediated signaling cascade, but the mutant is unable to produce vd-sRNA against *Pti* 1 mRNA. Conversely, vd-sRNAs derived from position 108–128 of PSTVd-RG1 had no known target in the tomato genome. That said, the insertion of an adenosine residue at position 118, as was observed in the deep-sequencing data, resulted in the targeting of a probable disease resistance protein. In addition, some mutations favored viroid adaptability by targeting host defense genes, while others showed negative effects on viroid survival by changing the vd-sRNA sequence that was supposed to target the host's defense genes. Details of the important targets observed in this study, and of the vd-sRNA:target duplexes formed, are shown in **Table 4**. These data indicate that certain mutations improve viroid accumulation, while others negatively affect its survival.

DISCUSSION

Upon infection, viroids form a population of sequence variants in their host plant, called quasi-species. This includes the master sequence as well as a certain number of sequence harboring slight differences (Codoñer et al., 2006). Two different forces play crucial roles in defining the viroid quasi-species. Firstly, mutations occur during replication due to the fact that viroids use either the host NEP or the DNA dependent RNA polymerase II for their replication (depending on viroid family), neither of which possess proofreading ability (Gago et al., 2009). The relevant polymerases are indeed not able to perform proofreading, but they have to act even on RNA instead of their natural DNA template, which might increase their error rate while replicating viroid. The second factor that influences viroid quasi-species is host-imposed selection pressure as discussed elsewhere (Tessitori et al., 2013). To date, all of the experiments studying viroid quasi-species have concentrated on analyzing the mutation at a specific point in the infection. While this kind of approach does provide details of viroid quasi-species present at that specific point, it fails to explain the shift in viroid population dynamics during course of the infection as viroid infection follows a S-curve (**Figure 2E**). This proposes that during the course of infection the viroid titer, as well as the

[2]http://wmd3.weigelworld.org/cgi-bin/webapp.cgi

TABLE 3 | Vd-sRNAs derived from the master sequence as well as its mutants and the total number of vd-sRNA targets detected in the tomato genome.

Position on PSTVd-RG1	Master sequence vd-sRNA	Number targets	Mutant	Mutant vd-sRNS	Number targets
MS:55	ACAAGAAAAG**A**AAAAAGAAGG	17	**Del:A55**	ACAAGAAAAGAAAAAGAAGGC	10
	CAAGAAAAG**A**AAAAAGAAGGC	4	**Ins:-A55**	ACAAGAAAAG**A**AAAAAAGAAG	26
			Ins:+A55	CAAGAAAAG**A**AAAAAAGAAGG	10
MS:118	GCGAACTGGC**A**AAAAAGGACG	0	**Del:A118**	GCGAACTGGCAAAAAGGACGG	0
	CGAACTGGC**A**AAAAAGGACGG	0	**Ins:-A118**	GCGAACTGGC**A**AAAAAAGGAC	4
			Ins:+A118	CGAACTGGC**A**AAAAAAGGACG	2
MS:168	AATTCCCGCC**G**AAACAGGGTT	0	**Sub:-G168A**	AATTCCCGCC**A**AAACAGGGTT	0
	ATTCCCGCC**G**AAACAGGGTTT	0	**Sub:+G168A**	ATTCCCGCC**A**AAACAGGGTTT	2
MS:253	GCTGTCGCTTCGGCTACTACC	1	**Ins:-C253**	GCTGTCGCTT**C**CGGCTACTAC	0
			Ins:+C253	CTGTCGCTT**C**CGGCTACTACC	0
MS:254	CTGTCGCTTCGGCTACTACCC	0	**Ins:-A254**	CTGTCGCTTC**A**GGCTACTACC	0
			Ins:+A254	TGTCGCTTC**A**GGCTACTACCC	0
			Ins:-G254	CTGTCGCTTC**G**GGCTACTACC	0
			Ins:+G254	TGTCGCTTC**G**GGCTACTACCC	0
			Ins:-U254	CTGTCGCTTC**T**GGCTACTACC	0
			Ins:+U254	TGTCGCTTC**T**GGCTACTACCC	0
MS:296	CGAGAACCGC**T**TTTTCTCTAT	1	**Del:U296**	CGAGAACCGCTTTTCTCTATC	1
	GAGAACCGC**T**TTTTCTCTATC	5	**Ins:U296**	CGAGAACCGC**T**TTTTTCTCTA	6
			Ins:U296	GAGAACCGC**T**TTTTTCTCTAT	4
MS:353	AACCGCAGTT**G**GTTCCTCGGA	0	**Del:G353**	AACCGCAGTTGTTCCTCGGAA	1
	ACCGCAGTT**G**GTTCCTCGGAA	0	**Sub:-G353A**	AACCGCAGTT**A**GTTCCTCGGA	0
			Sub:+G353A	ACCGCAGTT**A**GTTCCTCGGAA	0

MS denotes Master sequence; Ins denotes Insertion; Del indicates Deletion; and, Sub represents Substitution. Underlined letters denote the 10th and 11th nucleotides of the vd-sRNA:target RNA duplex. Bold black colored letters indicate the nucleotides on the master sequence that undergoes mutation. Bold red colored letters indicate the nucleotides on the viroid variant that showed mutation when compared to the master sequence. "+" indicates that a mutant nucleotide is located at position 10 of the vd-sRNA:target duplex, while "−" denotes that a mutant nucleotide is located at position 11 of vd-sRNA:target duplex.

host's morphological and physiological condition, are constantly changing. In the present study samples were collected from 1st to 4th wpi, which covers the different phases of viroid infection. After verifying the systemic infection of PSTVd in inoculated plants, leaf samples obtained from different plants for each week were subjected to library preparation using two different sets of primers and then subjected to paired end sequencing using the Illumina platform. Detailed analysis of each sequence that was detected at least ten times in a given pool per week showed a maximum of 6.1% divergence from the master sequence at 1 wpi. Except for one particular sequence, both libraries had about a 4.6% difference from master sequence. These data are very similar to the previous mutagenic studies performed on PLMVd, a member of the family, *Avsunviroidae* (Glouzon et al., 2014).

To act as an efficient pathogen, it is crucial for a viroid to adapt itself to the host environment. To understand how quasi-species could help the master sequence to adapt, the ten most abundant sequence variants were further considered, ensuring that the analysis was not biased by technical sequencing errors. As shown in **Supplementary Figure 1**, of the ten sequence variants, a few were found in both libraries. To increase the confidence in the data analysis, variants recovered from both the L1 and the L2 libraries were analyzed for their adaptability and fitness. It is interesting to note that at 1 wpi the master sequence accounted for 22% of the quasi-species. It then increased in week 2 to as high as 77% before declining slightly. It was also observed that the variants Ins:G254 and Ins:C253, which accounted for more than 20% at week 1 in both libraries, fell below 0.2% at week 2 and then recovered in the later stages of infection (i.e., up to 0.8 at week 4). It should be noted that the mutations at positions 253 and 254 affected sequences that are involved in the formation of HPI and Loop E structures of PSTVd, that are known to involved in viroid infectivity and replication (**Figure 3**; Hammond and Owens, 1987; Schrader et al., 2003; Gas et al., 2007). This may explain why these sequence variants were not detected after week 1. While this region is quite sensitive to any changes in the nucleotide sequence, deep-sequence data reveals that this is the most favored mutagenic point on PSTVd-RG1. Similar observations were made for other sequence variants; however, the abundance was less than 10% at any given point of infection. For example, sequence variants such as Ins:A55 and Del:U296 showed a gradual increase during the course of infection, while Del:A55 suddenly increased to 4%, from less than 1% during initial stages of infection, at week 4. Although a constant change in sequence variation was observed throughout the infection, the master sequence is the most abundant at any given point of infection, indicating that this is the fittest sequence of the PSTVd quasi-species.

The master sequence is not the only one that consistently increases with time. Some sequence variants showed better adaptability than others by slowly increasing their proportion during the course of infection, consequently they were represented in higher proportions (**Table 2**). More precisely, the top five variants showed at least a 2-fold enrichment in

TABLE 4 | Selected vd-sRNA:mRNA duplexes and the functions of the target host's genes.

Sl. No	Viroid position	vd-sRNA (3′ > 5′):mRNA (5′ > 3′) duplex	dG in kcal/mol	Target	Function of host gene
1	MS:55	vd-sRNA GGAAGAAAAAAGAAAAGAACA X:::X:::::::::::::X: mRNA TCTTTTTTTTTCTTTTCTTTT	−24.76	CDP-diacylglycerol–serine O-phosphatidyltransferase 1-like	Microspore maturation[a]
2	MS:55	vd-sRNA GGAAGAAAAAAGAAAAGAACA X:::::::::::::::X:::X mRNA TCTTCTTTTTTCTTTTATTGA	−24.21	Autophagy-related protein 2	Required for cytoplasm to vacuole transport (Cvt) vesicle and autophagosome completion[a]
3	MS:55	vd-sRNA GGAAGAAAAAAGAAAAGAACA XX::::::::::X:::::::: mRNA TGTTCTTTTTTCCTTTCTTGT	−23.69	Probable methyltransferase PMT20	CATALYSIS of the transfer of a methyl group to an acceptor molecule[a]
4	MS:55	vd-sRNA GGAAGAAAAAAGAAAAGAACA ::::X::::::X:X::::::: mRNA. CCTTTTTTTTTTTCTTCTTGT	−22.30	Pto kinase interactor 1 (Pti1)	A serine-threonine kinase involved in the hypersensitive response (HR)-mediated signaling cascade[a]
5	MS:55	vd-sRNA CGGAAGAAAAAAGAAAAGAAC XX:::::::::::::::X::: mRNA. CTCTTCTTTTTTCTTTTATTG	−22.93	Autophagy-related protein 2	Required for cytoplasm to vacuole transport (Cvt) vesicle and autophagosome completion[a]
6	Del:A55	vd-sRNA CGGAAGAAAAAGAAAAGAACA X::::::::::X:::X::::: mRNA ACCTTCTTTTTTTTTCCTTGT	−26.66	RNA-binding family protein	RNA-binding proteins: bind to the double or single stranded RNA in cells and participate in forming ribonucleoprotein complexes[b]
7	Del:A55	vd-sRNA CGGAAGAAAAAGAAAAGAACA X:X:::::::::::::::X: mRNA TCTTTCTTTTTCTTTTCTTTT	−26.24	Serine/threonine-protein kinase BLUS1-like	Ser/Thr protein kinase mediating a primary step for phototropin signaling in guard cells. Essential for stomatal opening[a]
8	Ins:A55	vd-sRNA GAAGAAAAAAAGAAAAGAACA ::X:::::::::::::::X: mRNA. CTACTTTTTTTCTTTTCTTTT	−21.91	Vacuolar protein sorting-associated protein 32 homolog 2	Component of the ESCRT-III complex, which is required for multivesicular bodies (MVBs) formation and the sorting of endosomal cargo proteIns:into MVBs[a]
9	Ins:A55	vd-sRNA GAAGAAAAAAAGAAAAGAACA ::::::::::::::::X::: mRNA CTTGTTTTTTTCTTTTCATGT	−20.51	Oligouridylate-binding protein 1B	Heterogeneous nuclear ribonucleoprotein (hnRNP)-like protein that acts as a component of the pre-mRNA processing machinery[a]
10	Ins:A55	vd-sRNA GAAGAAAAAAAGAAAAGAACA :::::::::::::::XXX:: mRNA CTTCTTTTTTTCTTTTTCCGT	−20.42	Thionin-like protein 2	May be involved in plant defense[a]
11	Ins:A55	vd-sRNA GAAGAAAAAAAGAAAAGAACA X:::X:::::::::::::X: mRNA TTTCCTTTTTTCTTTTCTTTT	−20.34	Pentatricopeptide repeat-containing protein At2g17033	PPR proteIns:function in organelles to target specific transcripts and are involved in post-transcriptional regulation[c]
12	Ins:A55	vd-sRNA GGAAGAAAAAAAGAAAAGAAC X::X:::::::::::::::X mRNA. ACTACTTTTTTTCTTTTCTTT	−24.05	Vacuolar protein sorting-associated protein 32 homolog 2	Same as Sl No. 8
13	Ins:A55	vd-sRNA GGAAGAAAAAAAGAAAAGAAC ::::::::::::X:X::X::: mRNA CCTTCTTTTTTTTTGTTTTTG	−21.49	Heat stress transcription factor A-4a	Transcriptional activator that specifically binds the DNA sequence 5′-AGAAnnTTCT-3′ known as heat shock promoter elements (HSE)[a]
14	Ins:A118	vd-sRNA CAGGAAAAAAACGGTCAAGCG X:::X:::::::::::::::XX mRNA CTCCAGTTTTTGCCAGTTCAG	−29.31	F-box/LRR-repeat MAX2 homolog A-like	Component of the SCF(ASK-cullin-F-box) E3 ubiquitin ligase complexes, which may mediate the ubiquitination and subsequent proteasomal degradation of target proteins. Is necessary for response to strigolactones and may be involved in the ubiquitin-mediated degradation of specific proteIns:that activate axillary growth[a]

(Continued)

TABLE 4 | Continued

Sl. No	Viroid position	vd-sRNA (3′ > 5′):mRNA (5′ > 3′) duplex	dG in kcal/mol	Target	Function of host gene
15	Ins:A118	vd-sRNA CAGGAAAAAAACGGTCAAGCG ::::::X::::::::X::X:: mRNA GTCCTTATTTTGCCATTTTGC	−27.15	Probable disease resistance protein At4g33300	Defense[a]
16	Ins:A118	vd-sRNA GCAGGAAAAAAACGGTCAAGC :X:::XX::::::::::::X mRNA CCTCCAGTTTTTGCCAGTTCA	−29.31	F-box/LRR-repeat MAX2 homolog A-like	Component of SCF (ASK-cullin-F-box) E3 ubiquitin ligase complexes, which may mediate the ubiquitination and subsequent proteasomal degradation of target proteins. Is necessary for responses to strigolactones and may be involved in the ubiquitin-mediated degradation of specific proteIns:that activate axillary growth[a]
17	MS:296	vd-sRNA TATCTCTTTTTCGCCAAGAGC ::::::X:X:::::::::XX mRNA AAAGAGGAGAAGCGGTTCTAC	−34.30	Ankyrin-3	Defense
19	MS:296	vd-sRNA CTATCTCTTTTTCGCCAAGAG ::X::::X:X:::::::::X mRNA GAAAGAGGAGAAGCGGTTCTA	−34.30	Ankyrin-3	Defense (no reference)
20	MS:296	vd-sRNA CTATCTCTTTTTCGCCAAGAG :X::::::X:::::::::X:X mRNA GGAAGAGAGAAAGCGGTTTTA	−31.90	Protein WHAT'S THIS FACTOR 1 homolog	RNA-binding protein involved in group II intron splicing[a]
21	Del:U296	vd-sRNA CTATCTCTTTTCGCCAAGAGC ::X::::XX::::X::::::: mRNA GACAGAGGGAAGCTGTTCTCG	−31.06	Threonine dehydratase biosynthetic, chloroplastic-like.	Involved in L-isoleucine biosynthesis pathway[a]
22	Ins:U296	vd-sRNA ATCTCTTTTTTCGCCAAGAGC ::::::::::X::::::::X: mRNA TAGAGAAAAATGCGGTTCTTG	−34.16	Protein NETWORKED 1A.	Plant-specific actin binding protein. Associates with F-actin at the plasma membrane and plasmodesmata. May be part of a membrane-cytoskeletal adapter complex[a]
23	Ins:U296	vd-sRNA ATCTCTTTTTTCGCCAAGAGC X::X::::X::::X::::::X mRNA CAGGGAAAGAAGCCGTTCTCC	−31.52	RNA-binding protein EWS.	RNA-binding proteIns:(RBPs) are central regulatory factors controlling posttranscriptional RNA metabolism during plant growth, development, and stress responses[d]
24	Ins:U296	vd-sRNA ATCTCTTTTTTCGCCAAGAGC :::::::X::::XX::::::: mRNA TAGAGAAGAAAGGCGTTCTCG	−30.23	Protein TONNEAU 1a-like	Involved in the control of the dynamic organization of the cortical cytoskeleton. May play a role in the organization of microtubule arrays at the centrosome through interaction with centrin 1 (CEN1)[a].
25	Ins:U296	vd-sRNA TATCTCTTTTTTCGCCAAGAG X::::::::::X::::::::X mRNA CTAGAGAAAAATGCGGTTCTT	−33.03	Protein NETWORKED 1A	Same as Sl. No. 25

MS denotes the Master Sequence; Ins denotes Insertion; Del indicates Deletion; and, Sub represents Substitution. [a]https://www.uniprot.org. [b]Dassi (2017). [c]Colcombet et al. (2013). [d]Lee and Kang (2016).

both libraries. Del:A55 showed a huge enrichment of 95-fold. The adenosine residue located at position 55 is part of a motif-A within the P domain of the PSTVd-RG1 secondary structure. Previous studies have shown that the deletion of this loop resulted in both decreased pathogenicity and trafficking of the viroid molecule as compared to the wild type PSTVd (Zhong et al., 2008).

Previous research demonstrated the involvement of vd-sRNA mediated RNA silencing in viroid pathogenesis irrespective of the viroid family (Adkar-Purushothama and Perreault, 2020). Analysis of the RNA silencing targets of vd-sRNA derived from both the master sequence and the variants provides a potential explanation as to why certain variants are more adaptable and fit. As mentioned earlier, the Del:A55 variant showed highest enrichment over the period of infection. vd-sRNAs derived from this region are probably targeting, in a more efficient way, genes preventing the establishment of infection than those derived from the other variants as well as from the master sequence. At this point, the obvious question is why Del:A55 did not outnumber the master sequence if it targets key genes more

efficiently than the latter. There might be several explanations: (i) the master sequence was the one that was initially inoculated, and chances are that its replication is favored as compared to that of the variant Del:A55; (ii) as the deletion of a nucleotide results in a change in the secondary structure, it might also have an effect on the stability of the mutants as well as on their interaction with the host's components; and, (iii) as infection progresses the host factors are continuously changing and the effect of this on the master sequence as well as on Del:A55 variant is not known. That said, a multi-dimensional approach including, but not limited to, structural analysis, study of the interaction of host components with each variant, biochemical, bioinformatic and biotechnological techniques are required to shed light on this problem.

The data presented here shows that the viroid quasi-species is constantly changing during the course of disease as the disease symptoms progress in the host plant. At the initial stage of infection, a quasi-species of viroid sequences is formed with the master sequence representing no more than 25% of the quasi-species. However, it adapted well, stabilized itself and constituted more than 70% during next stages of disease. In other words, the complexity of the quasi-species decreased from 75 to 30% during the course of infection, indicating the survival of the fittest variants during the course of disease. The data presented here raise many questions, such as what will the quasi-species be if another variant is used for the infection (example, Del:A55), what are the driving forces that stabilize the quasi-species of sequences throughout infection, what features of the viroid make it the most adaptive and fittest variant and what are the host factors involved. These studies require further research in this direction to understand the selection pressures that govern the viroid quasi-species stabilization and their interaction with host components.

Note: When this manuscript was in the final stage of review, deep-sequence analysis of columnea latent viroid, a member of the family *Pospiviroidae,* revealed 79–561 viroid sequence variants, depending on the plant species (Tangkanchanapas et al., 2020).

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/geo/, GSE147577.

AUTHOR CONTRIBUTIONS

CA-P, PB, and J-PP conceived and designed the experiments. CA-P and PB performed the experiments. CA-P and FB analyzed the data. J-PP contributed reagents, materials, and analytical tools. CA-P, FB, and J-PP contributed to the writing of the manuscript.

FUNDING

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC, Grant Nos. 155219–12 and -17) to J-PP. The RNA group was supported by grants from the Université de Sherbrooke. J-PP holds the Research Chair of Université de Sherbrooke in RNA Structure and Genomics and, was a member of the Centre de Recherche du CHUS. The funders had no role in study design, data collection and analysis, decision to publish or in the preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01235/full#supplementary-material

REFERENCES

Adkar-Purushothama, C. R., Brosseau, C., Giguère, T., Sano, T., Moffett, P., and Perreault, J.-P. (2015a). Small RNA derived from the virulence modulating region of the potato spindle tuber viroid silences callose synthase genes of tomato plants. *Plant Cell* 27, 2178–2194. doi: 10.1105/tpc.15.00523

Adkar-Purushothama, C. R., Iyer, P. S., and Perreault, J.-P. (2017). Potato spindle tuber viroid infection triggers degradation of chloride channel protein CLC-b-like and Ribosomal protein S3a-like mRNAs in tomato plants. *Sci. Rep.* 7:8341. doi: 10.1038/s41598-017-08823-z

Adkar-Purushothama, C. R., Kasai, A., Sugawara, K., Yamamoto, H., Yamazaki, Y., He, Y.-H., et al. (2015b). RNAi mediated inhibition of viroid infection in transgenic plants expressing viroid-specific small RNAs derived from various functional domains. *Sci. Rep.* 5:17949. doi: 10.1038/srep17949

Adkar-Purushothama, C. R., and Perreault, J. (2020). Current overview on viroid-host interactions. *Wiley Interdiscip. Rev. RNA* 11:e1570. doi: 10.1002/wrna.1570

Adkar-Purushothama, C. R., and Perreault, J.-P. (2018). Alterations of the viroid regions that interact with the host defense genes attenuate viroid infection in host plant. *RNA Biol.* 15, 955–966. doi: 10.1080/15476286.2018.1462653

Adkar-Purushothama, C. R., Sano, T., and Perreault, J.-P. (2018). Viroid-derived small RNA induces early flowering in tomato plants by RNA silencing. *Mol. Plant Pathol.* 19, 2446–2458. doi: 10.1111/mpp.12721

Andronescu, M., Aguirre-Hernández, R., Condon, A., and Hoos, H. H. (2003). RNAsoft: a suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Res.* 31, 3416–3422. doi: 10.1093/nar/gkg612

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

Brass, J. R. J., Owens, R. A., Matoušek, J., and Steger, G. (2017). Viroid quasispecies revealed by deep sequencing. *RNA Biol.* 14, 317–325. doi: 10.1080/15476286.2016.1272745

Codoñer, F. M., Darós, J.-A., Solé, R. V., and Elena, S. F. (2006). The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* 2:e136. doi: 10.1371/journal.ppat.0020136

Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthomé, R., et al. (2013). Systematic study of subcellular localization of *Arabidopsis* PPR proteins confirms a massive targeting to organelles. *RNA Biol.* 10, 1557–1575. doi: 10.4161/rna.26128

Cottilli, P., Belda-Palazón, B., Adkar-Purushothama, C. R., Perreault, J.-P., Schleiff, E., Rodrigo, I., et al. (2019). Citrus exocortis viroid causes ribosomal stress in tomato plants. *Nucleic Acids Res.* 47, 8649–8661. doi: 10.1093/nar/gkz679

Dalakouras, A., Dadami, E., and Wassenegger, M. (2015). Engineering viroid resistance. *Viruses* 7, 634–646. doi: 10.3390/v7020634

Dassi, E. (2017). Handshakes and fights: the regulatory interplay of RNA-binding proteins. *Front. Mol. Biosci.* 4:67. doi: 10.3389/fmolb.2017.00067

De la Peña, M., and Flores, R. (2002). Chrysanthemum chlorotic mottle viroid RNA: dissection of the pathogenicity determinant and comparative fitness of symptomatic and non-symptomatic variants. *J. Mol. Biol.* 321, 411–421. doi: 10.1016/s0022-2836(02)00629-0

De la Peña, M., Navarro, B., and Flores, R. (1999). Mapping the molecular determinant of pathogenicity in a hammerhead viroid: a tetraloop within the in vivo branched RNA conformation. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9960–9965. doi: 10.1073/pnas.96.17.9960

Ding, B. (2009). The biology of viroid-host interactions. *Annu. Rev. Phytopathol.* 47, 105–131. doi: 10.1146/annurev-phyto-080508-081927

Domingo, E. (2002). Quasispecies theory in virology. *J. Virol.* 76, 463–465. doi: 10.1128/JVI.76.1.463-465.2002

Domingo, E., and Perales, C. (2019). Viral quasispecies. *PLoS Genet.* 15:e1008271. doi: 10.1371/journal.pgen.1008271

Eamens, A. L., Smith, N. A., Dennis, E. S., Wassenegger, M., and Wang, M. B. (2014). In Nicotiana species, an artificial microRNA corresponding to the virulence modulating region of Potato spindle tuber viroid directs RNA silencing of a soluble inorganic pyrophosphatase gene and the development of abnormal phenotypes. *Virology* 45, 266–277. doi: 10.1016/j.virol.2013.12.019

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523. doi: 10.1007/BF00623322

Flores, R., Grubb, D., Elleuch, A., Nohales, M. -Á, Delgado, S., and Gago, S. (2011). Rolling-circle replication of viroids, viroid-like satellite RNAs and hepatitis delta virus: variations on a theme. *RNA Biol.* 8, 200–206. doi: 10.4161/rna.8.2.14238

Flores, R., Hernández, C., Martínez de Alba, A. E., Daròs, J.-A., and Di Serio, F. (2005). Viroids and viroid-host interactions. *Annu. Rev. Phytopathol.* 43, 117–139. doi: 10.1146/annurev.phyto.43.040204.140243

Flores, R., Serra, P., Minoia, S., Di Serio, F., and Navarro, B. (2012). Viroids: from genotype to phenotype just relying on RNA sequence and structural motifs. *Front. Microbiol.* 3:217. doi: 10.3389/fmicb.2012.00217

Gago, S., Elena, S. F., Flores, R., and Sanjuan, R. (2009). Extremely high mutation rate of a hammerhead viroid. *Science* 323, 1308–1308. doi: 10.1126/science.1169202

Gas, M.-E., Hernández, C., Flores, R., and Daròs, J.-A. (2007). Processing of nuclear viroids in vivo: an interplay between RNA conformations. *PLoS Pathog.* 3:e182. doi: 10.1371/journal.ppat.0030182

Glouzon, J.-P. S., Bolduc, F., Wang, S., Najmanovich, R. J., and Perreault, J.-P. (2014). Deep-sequencing of the peach latent mosaic viroid reveals new aspects of population heterogeneity. *PLoS One* 9:e87297. doi: 10.1371/journal.pone.0087297

Hammond, R. W., and Owens, R. A. (1987). Mutational analysis of potato spindle tuber viroid reveals complex relationships between structure and infectivity. *Proc. Natl. Acad. Sci. U.S.A.* 84, 3967–3971. doi: 10.1073/pnas.84.12.3967

Itaya, A., Folimonov, A., Matsuda, Y., Nelson, R. S., and Ding, B. (2001). Potato spindle tuber viroid as inducer of RNA silencing in infected tomato. *Mol. Plant. Microbe. Interact.* 14, 1332–1334. doi: 10.1094/MPMI.2001.14.11.1332

Itaya, A., Zhong, X., Bundschuh, R., Qi, Y., Wang, Y., Takeda, R., et al. (2007). A structured viroid RNA serves as a substrate for dicer-like cleavage to produce biologically active small RNAs but is resistant to RNA-induced silencing complex-mediated degradation. *J. Virol.* 81, 2980–2994. doi: 10.1128/JVI.02339-06

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13

Lee, K., and Kang, H. (2016). Emerging roles of RNA-binding proteins in plant growth, development, and stress responses. *Mol. Cells* 39, 179–185. doi: 10.14348/molcells.2016.2359

López-Carrasco, A., Ballesteros, C., Sentandreu, V., Delgado, S., Gago-Zachert, S., Flores, R., et al. (2017). Different rates of spontaneous mutation of chloroplastic and nuclear viroids as determined by high-fidelity ultra-deep sequencing. *PLoS Pathog.* 13:e1006547. doi: 10.1371/journal.ppat.1006547

Loss, P., Schmitz, M., Steger, G., and Riesner, D. (1991). Formation of a thermodynamically metastable structure containing hairpin II is critical for infectivity of potato spindle tuber viroid RNA. *EMBO J.* 10, 719–727. doi: 10.1002/j.1460-2075.1991.tb08002.x

Malfitano, M., Di Serio, F., Covelli, L., Ragozzino, A., Hernández, C., and Flores, R. (2003). Peach latent mosaic viroid variants inducing peach calico (extreme chlorosis) contain a characteristic insertion that is responsible for this symptomatology. *Virology* 313, 492–501. doi: 10.1016/S0042-6822(03)00315-5

Maniataki, E., Martinez de Alba, A. E., Sägesser, R., Tabler, M., and Tsagris, M. (2003). Viroid RNA systemic spread may depend on the interaction of a 71-nucleotide bulged hairpin with the host protein VirP1. *RNA* 9, 346–354. doi: 10.1261/rna.2162203

Martinez de Alba, A. E., Flores, R., and Hernandez, C. (2002). Two chloroplastic viroids induce the accumulation of small RNAs associated with posttranscriptional GENE silencing. *J. Virol.* 76, 13094–13096. doi: 10.1128/jvi.76.24.13094-13096.2002

Moreno, M., Vázquez, L., López-Carrasco, A., Martín-Gago, J. A., Flores, R., and Briones, C. (2019). Direct visualization of the native structure of viroid RNAs at single-molecule resolution by atomic force microscopy. *RNA Biol.* 16, 295–308. doi: 10.1080/15476286.2019.1572436

Navarro, B., Gisel, A., Rodio, M. E., Delgado, S., Flores, R., and Di Serio, F. (2012). Small RNAs containing the pathogenic determinant of a chloroplast-replicating viroid guide the degradation of a host mRNA as predicted by RNA silencing. *Plant J.* 70, 991–1003. doi: 10.1111/j.1365-313X.2012.04940.x

Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* 53, 674–690. doi: 10.1111/j.1365-313X.2007.03328.x

Owens, R. A., Tech, K. B., Shao, J. Y., Sano, T., and Baker, C. J. (2012). Global analysis of tomato gene expression during potato spindle tuber viroid infection reveals a complex array of changes affecting hormone signaling. *Mol. Plant Microbe Interact.* 25, 582–598. doi: 10.1094/MPMI-09-11-0258

Pallas, V., Martinez, G., and Gomez, G. (2012). The interaction between plant viroid-induced symptoms and RNA silencing. *Methods Mol. Biol.* 894, 323–343. doi: 10.1007/978-1-61779-882-5_22

Papaefthimiou, I., Hamilton, A., Denti, M., Baulcombe, D., Tsagris, M., and Tabler, M. (2001). Replicating potato spindle tuber viroid RNA is accompanied by short RNA fragments that are characteristic of post-transcriptional gene silencing. *Nucleic Acids Res.* 29, 2395–2400. doi: 10.1093/nar/29.11.2395

Repsilber, D., Wiese, S., Rachen, M., Schröder, A. W., Riesner, D., and Steger, G. (1999). Formation of metastable RNA structures by sequential folding during transcription: time-resolved structural analysis of potato spindle tuber viroid (-)-stranded RNA by temperature-gradient gel electrophoresis. *RNA* 5, 574–584. doi: 10.1017/s1355838299982018

Rodriguez, M. J., and Randles, J. W. (1993). Coconut cadang-cadang viroid (CCCVd) mutants associated with severe disease vary in both the pathogenicity domain and the central conserved region. *Nucleic Acids Res.* 21:2771. doi: 10.1093/nar/21.11.2771

Sano, T., Barba, M., Li, S.-F., and Hadidi, A. (2010). Viroids and RNA silencing: mechanism, role in viroid pathogenicity and development of viroid-resistant plants. *GM Crops* 1, 23–29. doi: 10.4161/gmcr.1.2.11871

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Schnell, R. J., Kuhn, D. N., Olano, C. T., and Quintanilla, W. E. (2001). Sequence diversity among avocado sunblotch viroids isolated from single avocado trees. *Phytoparasitica* 29, 451–460. doi: 10.1007/BF02981864

Schrader, O., Baumstark, T., and Riesner, D. (2003). A mini-RNA containing the tetraloop, wobble-pair and loop E motifs of the central conserved region of

potato spindle tuber viroid is processed into a minicircle. *Nucleic Acids Res.* 31, 988–998. doi: 10.1093/nar/gkg193

Semancik, J. S., and Szychowski, J. A. (1994). Avocado sunblotch disease: a persistent viroid infection in which variants are associated with differential symptoms. *J. Gen. Virol.* 75, 1543–1549. doi: 10.1099/0022-1317-75-7-1543

Tangkanchanapas, P., Haegeman, A., Ruttink, T., Höfte, M., and De Jonghe, K. (2020). Whole-genome deep sequencing reveals host-driven in-planta evolution of columnea latent viroid (CLVd) quasi-species populations. *Int. J. Mol. Sci.* 21:3262. doi: 10.3390/ijms21093262

Tessitori, M., Rizza, S., Reina, A., Causarano, G., and Di Serio, F. (2013). The genetic diversity of Citrus dwarfing viroid populations is mainly dependent on the infected host species. *J. Gen. Virol.* 94, 687–693. doi: 10.1099/vir.0.048025-0

Tsushima, D., Tsushima, T., and Sano, T. (2016). Molecular dissection of a dahlia isolate of potato spindle tuber viroid inciting a mild symptoms in tomato. *Virus Res.* 214, 11–18. doi: 10.1016/j.virusres.2015.12.018

Zheng, Y., Wang, Y., Ding, B., and Fei, Z. (2017). Comprehensive transcriptome analyses reveal that Potato spindle tuber viroid triggers genome-wide changes in alternative splicing, inducible trans -acting activity of phased secondary small interfering RNAs, and immune responses. *J. Virol.* 91:JVI.00247-17. doi: 10.1128/JVI.00247-17

Zhong, X., Archual, A. J., Amin, A. A., and Ding, B. (2008). A genomic map of viroid RNA motifs critical for replication and systemic trafficking. *Plant Cell* 20, 35–47. doi: 10.1105/tpc.107.056606

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.

Copyright © 2020 Adkar-Purushothama, Bolduc, Bru and Perreault. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 July 2020
doi: 10.3389/fmicb.2020.01627

Check for updates

Apple Russet Ring and Apple Green Crinkle Diseases: Fulfillment of Koch's Postulates by Virome Analysis, Amplification of Full-Length cDNA of Viral Genomes, *in vitro* Transcription of Infectious Viral RNAs, and Reproduction of Symptoms on Fruits of Apple Trees Inoculated With Viral RNAs

*Chunjiang Li[1], Hajime Yaegashi[2], Ryusuke Kishigami[1], Ayaka Kawakubo[1], Noriko Yamagishi[3], Tsutae Ito[2] and Nobuyuki Yoshikawa[1,3]**

[1] Faculty of Agriculture, Iwate University, Morioka, Japan, [2] Division of Apple Research, Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization, Morioka, Japan, [3] Agri-Innovation Center, Iwate University, Morioka, Japan

OPEN ACCESS

Edited by:
Ahmed Hadidi, Agricultural Research Service (USDA), United States

Reviewed by:
Francesco Di Serio, Istituto per la Protezione Sostenibile delle Piante, Italy
Jeremy Thompson, Cornell University, United States
Wilhelm Jelkmann, Julius Kühn-Institut, Germany

***Correspondence:**
Nobuyuki Yoshikawa
yoshikawa@iwate-u.ac.jp

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

Received: *26 March 2020*
Accepted: *22 June 2020*
Published: *10 July 2020*

Citation:
Li C, Yaegashi H, Kishigami R, Kawakubo A, Yamagishi N, Ito T and Yoshikawa N (2020) Apple Russet Ring and Apple Green Crinkle Diseases: Fulfillment of Koch's Postulates by Virome Analysis, Amplification of Full-Length cDNA of Viral Genomes, in vitro Transcription of Infectious Viral RNAs, and Reproduction of Symptoms on Fruits of Apple Trees Inoculated With Viral RNAs. Front. Microbiol. 11:1627. doi: 10.3389/fmicb.2020.01627

Apple russet ring and apple green crinkle are graft-transmitted diseases first reported more than 60 years ago, but at present, no association between a specific virus (variant) and the disease has been clearly demonstrated. In this study, we conducted the following series of experiments to identify the causal viruses (variants) of these apple diseases; (1) comprehensive analysis by next-generation sequencing of all viruses in each apple tree affected with russet ring or green crinkle disease, (2) amplification of full-length genomic cDNA of viruses using primers containing the T3 promoter and the *in vitro* transcription of infectious viral RNAs, (3) inoculation of viral RNA transcripts to both herbaceous and apple plants, (4) analysis of sequence variants of viruses present in infected plants, (5) back-inoculation of sequence variants of candidate viruses to apple seedlings combined with the virus-induced flowering technology using the apple latent spherical virus vector to reproduce the symptom on the fruit as soon as possible, and (6) reproduction of symptoms on the fruits of apple trees inoculated with sequence variants and the re-isolation of each virus variant from apples showing fruit symptoms. The results showed that one of the sequence variants of the apple chlorotic leaf spot virus causes a characteristic ring-shaped rust on the fruits of infected apple trees and that a sequence variant of the apple stem pitting virus probably causes green crinkle

symptoms on an infected apple fruit. Thus, we were able to fulfill Koch's postulates to prove the viral etiology of both the apple russet ring and green crinkle diseases. We also propose an experimental system that can prove whether a virus found in diseased tissues is the pathogen responsible for the diseases when the etiology is undetermined.

Keywords: apple russet ring disease, apple green crinkle disease, next generation sequencing, apple chlorotic leaf spot virus, apple stem pitting virus, Koch' postulates

INTRODUCTION

There are many virus-like diseases affecting deciduous fruit trees whose causal agents are not yet identified (Németh, 1986; Howell et al., 2011). Even if a certain virus is presumed to be the cause of the disease, the causal relationship between the presence of a virus and specific symptoms has not been conclusively proven in many fruit tree viral diseases because back-inoculation of the virus to the original fruit trees has not been achieved; that is, Koch's postulate has not yet been fulfilled. There are several reasons why it is difficult to identify the causative virus of fruit tree diseases. Firstly, many viruses are not so easy to isolate into herbaceous test plants and/or to back-inoculate the isolated virus into their original fruit trees. Secondly, some viruses of fruit trees have not only difficulties to be transmitted to herbaceous host plant, but they simply have a limited host range. This makes it difficult to isolate the causative agent as a single virus. The third reason is that the infection pattern of viruses is often very complex in fruit trees. For example, reports show that a single cherry tree affected by a bud blight disease was infected with 7 – 8 different virus species at the same time, and that most of these viruses could not be transmitted and isolated in herbaceous test plants (Isogai et al., 2004; Yaegashi et al., 2020). In Japan, most apple trees are infected with viruses such as apple chlorotic leaf spot virus (ACLSV), apple stem pitting virus (ASPV), and apple stem grooving virus (ASGV), etc. without showing obvious symptoms (Yaegashi et al., 2011). In addition, each virus (ACLSV, ASGV, or ASPV) in single apple trees consists of a mixture of sequence variants that differ considerably from one another in terms of nucleotide sequences (Candresse et al., 1995; Yoshikawa et al., 1996, 2001; Magome et al., 1997b, 1999; Yaegashi et al., 2007). The complexity of the virus infection patterns in fruit trees makes it difficult to identify the virus or virus variant responsible for a specific disease. Examples of graft-transmitted diseases of apples with undetermined etiology include the apple russet ring and apple green crinkle diseases (Howell et al., 2011).

The apple russet ring is a graft-transmitted disease of apple in which a characteristic ring-shaped rust appears on the fruits. The disease was firstly reported on a 'McIntosh' apple that showed leaf pucker symptoms in 1954 in British Columbia, Canada (Wood, 1971). The russet ring symptom of the fruit was shown to be associated with leaf pucker (Wood, 1971). The graft transmission of the russet ring disease was reported on the 'Golden Delicious' apple from Washington state, United States (Wood, 1971); in Japan, the russet ring disease was found in the 'Mutsu' apple from Iwate Prefecture in 1972 and was also shown to be grafted-transmitted (Yanase et al., 1988). The disease is now reported to occur worldwide and implicated to be caused by a strain (or by strains) of ACLSV (van der Meer, 1986; Desvignes and Boye, 1988; Wood, 2001; Howell et al., 2011). However, many strains or isolates of ACLSV are naturally in apple trees that are both unaffected and affected by russet ring disease (Yaegashi et al., 2007, 2011). In order to demonstrate that ACLSV may be responsible for apple russet ring disease, it is necessary to isolate and characterize the strain and its sequence variants at the molecular level, and to back-inoculate these into apple plants for symptom reproduction.

Apple green crinkle disease was first observed in 1929 in Aomori Prefecture, Japan and the graft-transmission of this disease was demonstrated in Japan in 1934 (Sawamura, 1965), thus this disease is the world's first reported virus-like disease affecting apple plants. The disease was also reported in Canada and New Zealand in the 1930s (Chamberlain et al., 1974). Symptoms occur only in the fruit and appear most clearly during the young fruit stage, wherein some of the fruit are not enlarged and are irregularly dented, resulting in abnormal fruit morphology. The pericarp of the indented part will later cork and become rust-like, and in severe cases, it will tear (Chamberlain et al., 1974). To date, multiple viruses such as ASPV, ACLSV, and ASGV etc. have been detected in trees affected by green crinkle disease (Yaegashi et al., 2007; James et al., 2013). Among them, ASPV has been implicated as the causal virus, but the relationship between ASPV isolates and crinkle symptoms of fruits has not been demonstrated (Howell et al., 2011). James et al. (2013) reported that a putative new foveavirus closely related to ASPV, namely the apple green crinkle associated virus (AGCaV), had been isolated from apples showing severe crinkle disease.

In this study, we conducted the following series of experiments to determine the causal viruses (variants) of the apple russet ring and green crinkle diseases: (1) comprehensive analysis by next-generation sequencing (NGS) of all viruses in each apple tree affected by russet ring or green crinkle diseases, (2) amplification of the full-length genomic DNA of viruses using primers containing the T3 promoter and the *in vitro* transcription of infectious viral RNAs, (3) inoculation of virus RNA transcripts to herbaceous and apple plants, (4) analysis of sequence variants of viruses in infected plants, (5) back-inoculation of sequence variants of candidate viruses to apple seedlings combined with virus-induced flowering technology using the apple latent spherical virus (ALSV) vector (Yamagishi et al., 2014, 2016; Yamagishi and Yoshikawa, 2016), and (6) reproduction of symptoms on the fruits of apple trees inoculated with sequence variants and the re-isolation of each virus variant from apples showing fruit symptoms. From the results of the above studies, we were

able to prove the specific sequence variants of the causal viruses of the two apple diseases. Our study establishes an experimental system to identify the causal virus of fruit tree diseases with undetermined etiology by isolation of the viruses using the amplification of full-length cDNA of viral genomes followed by *in vitro* transcription of infectious viral RNAs and reproduction of symptoms on original hosts by back inoculation.

MATERIALS AND METHODS

Plants and dsRNA Extraction

An apple tree (PK-51, 'Royal Gala') affected with the apple russet ring disease (ARRD) and an apple tree (P-190, 'Golden Delicious') with the apple green crinkle disease (AGCD) were maintained in the field of the Division of Apple Research, Institute of Fruit Tree and Tea Science, NARO in Morioka, Iwate prefecture, Japan. These apple trees show typical symptoms of russet-ring and green crinkle on fruits, respectively (**Supplementary Figure S1**). dsRNAs were extracted from the newly developed leaves of an apple tree PK-51 infected with ARRD and an apple tree P-190 infected with AGCD. Leaf samples (20 g) were collected from diseased apple trees and stored at −80°C until use. Extraction of dsRNA from the leaves of diseased trees were performed using cellulose powder as described previously (Morris and Dodds, 1979; Yoshikawa and Converse, 1990), and the dsRNA was dissolved in 20 μL TAE buffer (40 mM Tris, 20 mM sodium acetate, 1 mM EDTA, pH 7.0). The dsRNA was electrophoresed on a 5% polyacrylamide gel and stained with silver nitrate to determine the presence of dsRNAs (Yoshikawa and Converse, 1990).

Next-Generation Sequencing and Computational Analysis

dsRNAs from the ARRD or AGCD-infected samples were then subjected to library preparation as reported previously (Noda et al., 2017; Yaegashi et al., 2020). The prepared library was analyzed using 6% polyacrylamide gel electrophoresis and the products (200–300 bp) were purified and subjected to paired-end sequencing (read length 101 bases) using the HiSeq2000 (Illumina Inc., San Diego, CA, United States). Resulting reads were assembled *de novo* into contiguous sequences (contigs) using Velvet (ver. 1.2.08) (Zerbino and Birney, 2008) with $k = 97$ in RR and $k = 95$ in GC samples. The obtained contigs (>500 bp) were then used to search for virus sequences using the blast program of NCBI BLAST+.[1]

RT-PCR

Leaves and petals were collected from apple trees affected by ARRD (PK-51) and AGCD (P-190) and were stored frozen at −80°C. Frozen samples (1 g) of each affected tree were put into a 2.0 mL microtube together with stainless steel beads SUB-50 (TOMY SEIKO Co., Ltd., Tokyo, Japan) and added 800 μL of RNA extraction buffer [2% CTAB, 2% PVP K-30, 100 mM Tris-HCl (pH 8.0), 25 mM EDTA (pH 8.0), 2 M NaCl, 2% mercaptoethanol]. Samples in microtubes were crushed using a Micro Smash MS-100R (TOMY SEIKO) and incubated at 65°C for 20 min with occasional stirring. After the addition of 800 μL of chloroform, the extracts were sufficiently stirred and centrifuged at 10,000 × *g* for 10 min (4°C). Then, 700 μL of the supernatants were added to 233 μL of 7.5 M LiCl, sufficiently mixed, and left at 4°C overnight. After centrifugation at 14,000 × *g* for 25 min, the pellets were added to 1 mL of 80% ethanol and centrifuged at 14,000 × *g* for 5 min. The pellets were then suspended in sterilized distilled water (SDW) to obtain the total RNA solutions. Total RNA concentrations were measured by NanoDrop (Thermo Fisher Scientific, Tokyo, Japan).

The coat protein (CP) region of each virus was amplified through RT-PCR as follows: 1 μL of total RNA solution (1 μg/μL), 0.5 μL of Oligo (dT)12 primer (10 μM), 4 μL of 2.5 mM dNTP mixture (TaKaRa Bio Inc., Kusatsu, Japan), 2 μL of 5 × RT Buffer (TOYOBO Co., Ltd., Tokyo, Japan), 0.5 μL of ReverTra Ace (100 unit/μl) (TOYOBO), and 2 μL of SDW were mixed and incubated at 42°C for 60 min using the TaKaRa PCR Thermal Cycler Dice® Version II (TaKaRa) for reverse transcription reaction. PCR was then conducted using TaKaRa Ex Taq and the following primer pairs: SP-CP8148-8167 (+) [5′-ttcgaccctaaccttcatgg-3′] and SP-CP8870-8890 (–) [5′-ctttgagtttgcagcatgagg-3′], with the expected product corresponding to base numbers 8148 to 8890 of ASPV-IF38 (accession no. AB045371) (Yoshikawa et al., 2001) for the ASPV-CP region, AC-CP 6821-6840 (+) [5′-agatctgaaagcgttcctg-3′] and AC-CP7342-7365 (–), with the expected product is corresponding to bases 6821–7365 of ACLSV-P205 (accession no. D14996) (Sato et al., 1993) for the ACLSV-CP region, and SG-CP5590-5611(+) (5′-aaagagaaagtttaggtccctc-3′) and SG-CP6395-6413 (–) (5′-taaaggcaggcatgtcaac-3′) corresponding to bases 5590–6413 of ASGV-P209 (accession no. D14995) (Yoshikawa et al., 1992). The PCR conditions were: 35 cycles of 94°C for 30 s, 58°C for ASPV, 54°C for ACLSV, and 55°C for ASGV for 30 s, 72°C for 1 min. PCR products were electrophoresed on a 1% agarose gel, purified using MonoFas® DNA purification kit (GL sciences), and then TA cloning was performed using the TaKaRa Mighty TA-cloning Kit (TaKaRa). Plasmids were extracted from *Escherichia coli* and used for sequence analysis as descried below.

In vitro Transcription of Virus Genomic RNA

Synthesis of full-length DNA from the genomic RNAs of ACLSV, ASPV, and ASGV was done by subjecting total RNA samples to a reverse transcription reaction at 42°C for 60 min using the PrimeScript II RT enzyme and oligo dT primer from the PrimeScript® High Fidelity RT-PCR Kit (TaKaRa) according to manufacturer protocols. Full-length DNAs of the three viruses were amplified by PrimeSTAR GXL DNA polymerase (TaKaRa) using primer pairs containing the T3 promoter sequence: ACLSV5'endT3(+) (5′-aattaaccctcactaaagtgatactgatacagtgtacact-3′) and ACLSV3'end(–)

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

(5′-ttttttttagtagtaaaatatttaaaagt-3′) for ACLSV, ASPV5′endT3(+) (5′-aattaaccctcactaaaggatacgcaaacaaactctg-3′ and ASPV3′end(−) (5′-ttttttgaaaatctagttaaaacaaaaataag-3′) for ASPV, and ASGV5′ endT3(+) (5′-aattaaccctcactaaaggaaatttaacaggcttaattt-3′) and ASGV3′end(−) (5′-ttttttttttttttttttttttagagtggacaaactctagac-3′) for ASGV. PCR was performed as follows: 98°C for 30 s, 54°C for ACLSV, 59°C for ASPV, and 52°C for ASGV for 30 s, and 68°C for 10 min for 40 cycles. The PCR product was then purified using the MonoFas® DNA Purification Kit II (GL Sciences Inc., Tokyo, Japan). Purified full-length genomic DNAs were then used as templates for the *in vitro* transcription of full-length genomic RNA of each virus using the MEGAscript® Kit (Ambion, Thermo Fisher scientific, Tokyo, Japan). The reaction mixtures contained the full-length genomic cDNA of each virus, ATP, CTP, UTP (75 mM each), GTP (15 mM), Cap Analog (40 mM) Enzyme Mix, and RNasin® Plus RNase Inhibitor (Promega, Madison, United States), and were incubated for 3.5 h at 37°C. The reaction mixtures were then treated with TURBO DNase (Thermo Fisher scientific) (2 U/μl) at 37°C for 15 min to decompose the template DNAs. RNA samples were then extracted using water-saturated phenol and chloroform, precipitated using ethanol, and suspended in SDW. RNA transcript concentrations were measured using NanoDrop, while RNA transcript quality was also checked using 1% agarose gel electrophoresis.

Inoculation of RNA Transcripts to Plants and Sequence Analysis of Virus Variants in Infected Plants

Inoculation of the RNA transcripts of ACLSV, ASPV, and ASGV to *Nicotiana occidentalis* was conducted as follows: two leaves of *N. occidentalis* plants (6th to 7th leaf stage) were inoculated with 10 μL of viral RNA transcripts (0.5–1 μg/μl) per leaf, along with carborundum. After inoculation, plants were grown in a growth chamber at 25°C with the day length set to 8 h. *Chenopodium quinoa* was also used for inoculation of the ACLSV-RNA transcript.

Total RNA was extracted from the leaves of infected plants and the CP region of each virus was amplified as described above. For sequence analysis of virus variants in each infected plant, the PCR products were purified using MonoFas® DNA purification kit (GL sciences) and TA cloning was performed using the TaKaRa Mighty TA-cloning Kit (TaKaRa). DNA plasmids were purified using a Plasmid DNA Extraction Mini Kit (FAVORGEN Biotech, Ping-Tung, Taiwan) according to the manufacturer's protocol. The sequences of the cDNA clones were analyzed using a DNA sequencing service from Sigma-Aldrich Japan. The homology of the CP region sequence between DNA clones was analyzed using the DNASIS software (Hitachi, Tokyo, Japan).

Back-Inoculation of Apple Seedlings With ACLSV and ASPV

Both ARRD and AGCD cause characteristic symptoms on apple fruits as shown in **Supplementary Figure S1**. However, it generally takes at least 6–7 years for apple seedlings to bloom. For this reason, we used virus-induced flowering (VIF) system using the ALSV vector (Yamagishi et al., 2014, 2016; Yamagishi and Yoshikawa, 2016) for symptom reproduction on the apple fruits. VIF using the ALSV vector can reduce flowering 2 months after germination and lead to fruit formation within 1 year in apple seedlings (Yamagishi et al., 2014).

Total RNA samples extracted from infected *N. occidentalis* leaves were used as templates for generating full-length cDNA of each sequence variant of both ACLSV and ASPV, followed by *in vitro* transcription of infectious viral RNA, as described above. As inoculum sources for inoculation to apple seedlings, we selected five *N. occidentalis* plants infected with RRACV1, RRACV2, RRACV3, RRACV5, or RRACV1 + RRACV4. In the case of ASPV variants, the RNA transcripts of ASPV from the AGCD-sample were also inoculated to apple seedlings with the ALSV-AtFT/MdTFL1 vector through particle bombardment, as described above.

Immediately after germination, cotyledons of apple seedlings from seeds of 'Ourin' were biolistically inoculated with the RNA transcript of each viral RNA, along with the RNA of ALSV-AtFT/MdTFL1 as described previously (Yamagishi et al., 2010, 2016). Inoculation was performed in a helium pressure of 1379 kilopascal (kPa) using the Helios Gene Gun system (Bio-Rad Laboratories, München, Germany), with two shots used per cotyledon (4 shots in total). Inoculated apple seedlings were incubated in a growth chamber at 25°C for ACLSV and at 20°C for ASPV, with the day length set to 16 h for approximately 2 months, and then grown in a non-contaminant greenhouse.

Infection assay was conducted by RT-PCR as described before (Yamagishi et al., 2014) using primer pairs for ACLSV [ACCP6821-6840(+) (5′-agatctgaaagcattcctg-3′) and ACCP7342-7365(−) (5′-ctaaatgcaaagatcagttgtaac-3′)], ASPV [SPCP8148-8167(+) (5′-ttcgaccctaaccttcatgg-3′) and SPCP8 870-8890(−) (5′-ctttgagtttgcagcatgagg-3′)], ALSV-RNA1 [ALSV RNA16598(+) (5′-gtacattcctcccaatcaaag-3′) and ALSVRNA 16691(–) (5′-ggatcaggagaacaaactag-3′)], and ALSV-RNA2 [ALS VRNA21418(+) (5′-cccaaatctgctagaaggtc-3′) and ALSVR NA21511(–) (5′-gcaaggtggtcgtga-3′)]. When the infected apple seedlings flowered, they were pollinated to produce fruits.

Grafting

To confirm that RRACV2 induced ARRD in apple seedlings as described above, graft-transmission tests were conducted using apple seedlings infected with each RRACV and ALSV-AtFT/MdTFL1. Approximately 1 year after inoculation by particle bombardment inoculation, the shoots of infected apple seedlings infected with each sequence variant (SV) of ACLSV were cut and grafted to the stem of 3-year-old virus-free apple trees ('Golden Delicious'/JM7) using the side-grafting method for the reproduction of russet ring symptoms on the fruits. Grafted apple trees were named APT1 (an apple tree grafted with RRACV1), APT2 (with RRACV2), APT3 (with RRACV3), APT1 + 4 (with RRACV1 + RRACV4), and APT5 (with RRACV5). Inoculated apple plants were grown in a greenhouse for 4 years at the Apple Research Center, NARO in Morioka, Japan.

Nucleotide Sequencing of the Complete Genome of ACLSV and ASPV Variants

DNA fragments covering entire genomes of the sequence variants of both ACLSV and ASPV were amplified from samples infected with each variant using the primer pairs shown in **Supplementary Tables S1**, **S2**. TA cloning of the amplified sequences was conducted as described above. Nucleotide sequencing of the cDNA clones was performed using a DNA sequencing service (Sigma-Aldrich, Japan).

RESULTS

Virome Analysis of Apple Trees Affected by ARRD and AGCD

Using NGS analysis of dsRNAs (**Supplementary Figure S2**) from the ARRD and AGCD samples in this study, we obtained 51,514,584 reads from the ARRD sample and 70,424,624 reads from the AGCD sample (**Table 1**). The conventional *de novo* assembly of the reads from ARRD and AGCD samples generated 411 contigs with an average of 2,095 base pairs (bp) and 853 contigs with an average of 771 bp, respectively (**Table 1**). In a subsequent blastn analysis, 63 contigs from the ARRD-sample and 244 contigs from the ACGD-sample corresponded to virus sequences (**Table 1**). In the ARRD sample, ACLSV had the largest number of contigs (42.9% of the total), followed by ASPV (13%), ASGV (6.4%), AGCaV (3.2%), and apricot latent virus (ApLV) (1.6%) (**Table 1**). On the other hand, in the GC sample, ASPV was the largest, accounting for 34.4% of the total, followed by ASGV (8.2%), ACLSV (7.8%), AGCaV (2.9%), and ApLV (2.9%) (**Table 1**).

Isolation of Viruses Into Separate Plants by Inoculation of *in vitro* Transcripts of Viral RNAs

Virome analysis by NGS revealed that both PK-51 and P-190 trees were infected with multiple viruses as described above. In order to isolate each virus from a diseased apple trees into separate herbaceous plants, the full-length DNA of the genomic RNA of each virus was amplified from the total RNA of an apple sample with ARRD (PK-51) using RT-PCR (**Supplementary Figure S3**), and the obtained full-length viral RNAs were transcribed *in vitro* using the full-length DNAs as templates. It can be seen that RNAs corresponding to the size of each virus genome [ASPV, 9.3 kb: ACLSV, 7.4 kb; ASGV, 6.4 kb] were synthesized (**Supplementary Figure S3**). Each RNA transcript was inoculated into *N. occidentalis* plants, and the infection was assayed using RT-PCR, from which it was found that infection was present in 17 plants of the 24 plants inoculated with the ACLSV RNA transcript (infection rate, 71%), 5 out of 10 plants inoculated with ASPV (50%), and 6 out of 6 plants inoculated with ASGV (100%). Infected *N. occidentalis* plants showed symptoms consisting of mosaic and malformation of leaves (ACLSV) and of interveinal chlorosis (ASPV) (**Supplementary Figure S4**). *C. quinoa* plants infected with ACLSV showed chlorotic spots and mosaic in the upper leaves, which is typical of ACLSV infection (**Supplementary Figure S4**). *N. occidentalis* plants infected with ASGV did not show any obvious symptoms. RT-PCR using a primer pair for each virus showed that only ASPV was detected in plants inoculated with the ASPV RNA transcript, and the other two viruses (ACLSV and ASGV) were not detected (**Supplementary Figure S3**). Similarly, only ACLSV was detected in plants inoculated with the ACLSV RNA transcript (**Supplementary Figure S3**), and only ASGV was detected in plants inoculated with the ASGV-RNA transcript (**Supplementary Figure S3**). Thus, by the inoculation of *in vitro* transcripts of viral RNAs, each of the three viruses co-infecting an apple tree could be isolated into separate plants.

Complexity of Sequence Variants of ACLSV in an Apple Tree (PK-51) Affected by ARRD

We analyzed the nucleotide sequence of the selected 18 clones of the ACLSV-CP region amplified from the ARRD-infected apple tree (PK-51). Here, those with nucleotide sequence identity of 99% or more in the CP region (582 nt) were regarded as deriving from the same SV. The results indicate that at least six sequence variants (SV1∼SV6) were present in samples from an apple with ARRD (**Supplementary Figure S5**). The detection frequency of each SV was different and the detected numbers of each SV in the total 18 clones analyzed were 5 clones for SV1, 2 clones for SV2, 1 clone for SV3, 4 clones for SV4, 6 clones for SV5, and 1 clone for SV6. Sequence identities of the CP region between SVs of ACLSV were 83.3–95.1% in terms of nucleotides and 91.7–97.9% in terms of amino acids.

In subsequent experiments, we analyzed the SVs in 17 herbaceous plants (*N. occidentalis* and *C quinoa*) infected with the RNA transcript of the ACLSV genome. Nucleotide sequence analysis of DNA clones from 17 infected plants (10 clones/plant) indicated that five SVs (named RRACVs1 ∼ 5) were detected in infected plants (**Figure 1**). Among the 17 infected plants, most *N. occidentalis* plants were singly infected with either RRACV1 (11 plants), RRACV2 (2 plants), RRACV3 (1 plant), or RRACV5 (2 plants), except for a plant infected with RRACV4 that was found to be co-infected with RRACV1. Comparison of the RRACV1–RRACV5 nucleotide sequences with those of SV1-SV6 from an RR-apple in **Supplementary Figure S5**, RRACV1 was the same as SV1, RRACV2 was SV2, RRACV3 was SV3, and RRACV4 was SV4. On the other hand, RRACV5 was different from both SV5 and SV6 in an apple with ARRD, suggesting that at least seven SVs were co-infecting the RR-apple (PK-51). We could not find *N. occidentalis* plants infected with SV5 and/or SV6 in this experiment. Sequence identities of the CP region among RRACVs were 83.8–92.4% in terms of nucleotides and 91.7–97.9% in terms of amino acids (**Table 2**). The sequence identities between RRACVs 1-5, SVs 5 and 6, and isolates reported previously (P205 and B6) are shown in **Table 2**. It has been reported that ACLSV can be divided into two types (P205 and B6) based on the CP amino acid sequence (Yaegashi et al., 2007). Among the seven SVs found in an ARRD-apple and inoculated *N. occidentalis*,

TABLE 1 | Summary of the conventional *de novo* assembly strategy of the reads from russet ring (RR)- and green crinkle (GC)-diseased apple samples.

Apple trees	No. of reads	Total no. of contigs			No. of contigs with hit to virus**					
		Total (mean, bp)	With virus hit*	With no hit	ACLSV	ASGV	ASPV	AGCaV	ApLV	others
PK-51 with RR	51,514,584	441 (2,095)	63	378	27	8	11	2	1	14
P-190 with GC	70,424,624	853 (771)	244	609	19	20	84	7	7	107

**No. of contigs with hit to viruses with significant E-value (<0.001). **ACLSV, apple chlorotic leaf spot virus; ASGV, apple stem grooving virus; ASPV, apple stem pitting virus; AGCaV, apple green crinkle-associated virus; ApLV, apricot latent virus; others, stealth virus 1 etc.*

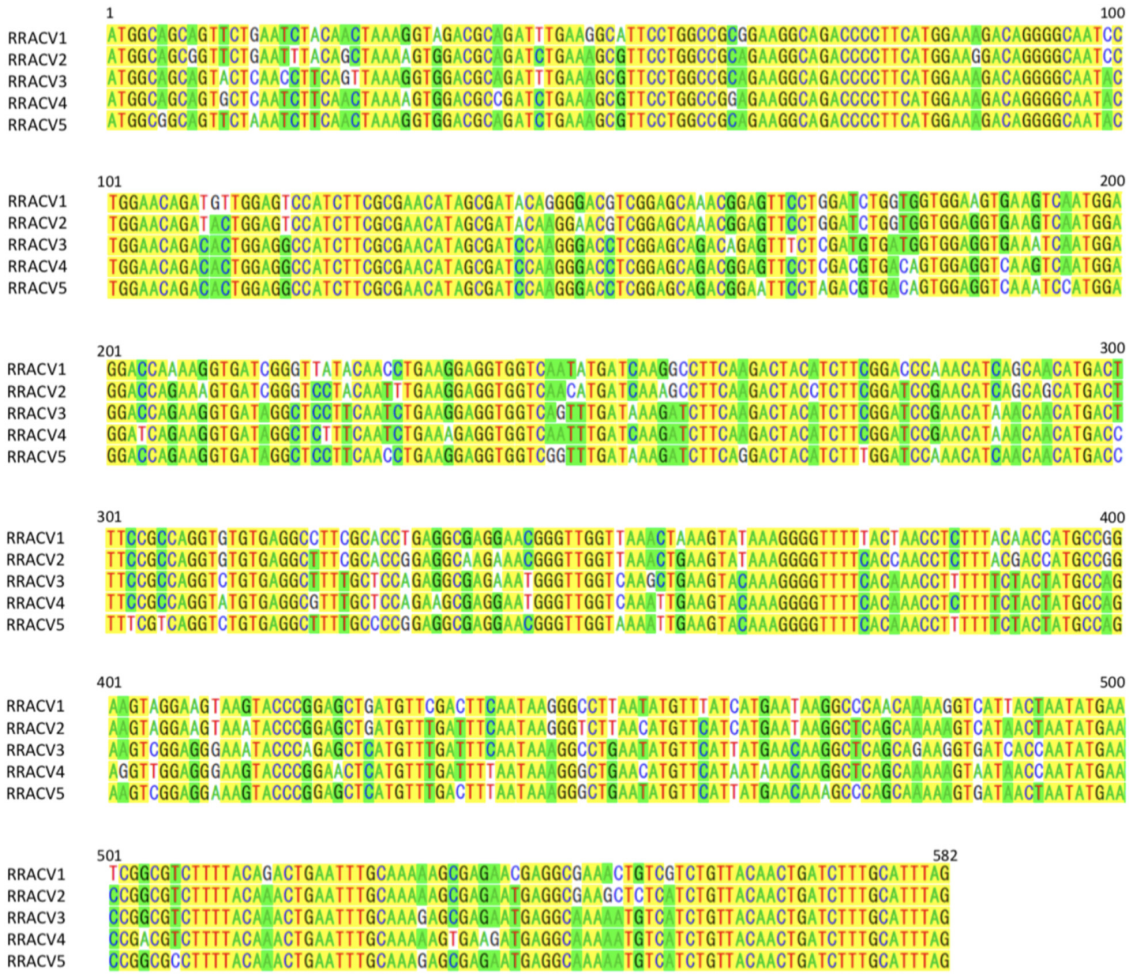


FIGURE 1 | Alignment of the nucleotide sequences of CP region of ACLSV-RRACVs detected in leaf samples of *N. occidentalis* plants inoculated with the RNA transcript of ACLSV from ARRD. Identical bases among the five SVs are filled in yellow, and if there are substitutions in the bases, they are filled in green (the number is dominant) or unfilled (the number is inferior).

three SVs (RRACV3, RRACV4, and RRACV5) belong to the P205 type and four (RRACV1, RRACV2, SV5, and SV6) belong to the B6 type.

ACLSV-RRACV2 Induces Russet-Ring Symptoms on Fruits of Infected Apple Seedlings

Approximately 1 month after inoculation with viral RNA transcripts of RRACV1, RRACV2, RRACV3, RRACV5, and RRACV1 + 4 (10–20 seedlings per each RRACV), the upper leaves of apple seedlings were sampled, and virus infection was examined by RT-PCR. Both ACLSV and ALSV were detected in most of the inoculated apple seedlings (example shown in **Supplementary Figure S6**). However, some seedlings in which the *AtFT* gene was deleted from the ALSV vector were observed as shown in **Supplementary Figure S6**. As a result, 10 seedlings infected with RRACV1, 6 seedlings with RRACV2, 4 seedlings with RRACV3, and 4 seedlings with RRACV5 were used for further experiments. Infected

TABLE 2 | Identities of the nucleotide and amino acid sequences of the coat protein gene between the sequence variants (SV) and the isolates of ACLSV found in infected *N. occidentalis* and an apple tree (PK-51) with ARRD.

SVs and isolates	SVs and isolates								
	RRACV1	**RRACV2**	**RRACV3**	**RRACV4**	**RRACV5**	**SV5**	**SV6**	**P205**	**B6**
RRACV1		91.0	85.4	83.8	85.0	95.1	93.4	84.5	91.2
RRACV2	97.9		87.1	86.0	85.5	91.7	89.1	85.2	96.0
RRACV3	93.2	92.7		91.7	92.4	86.4	85.7	89.7	86.6
RRACV4	92.2	91.7	97.4		91.7	84.8	83.3	87.1	86.4
RRACV5	97.9	91.7	97.9	92.2		84.7	82.8	88.4	86.2
SV5	97.9	96.8	92.7	91.7	91.7		92.7	83.7	90.5
SV6	92.2	97.4	91.7	97.9	90.6	96.3		83.6	88.3
P205	91.7	91.7	96.3	94.3	96.4	90.7	90.1		85.0
B6	97.9	97.9	94.3	92.7	92.7	97.4	96.4	90.2	

The percentages of nucleotide sequence (582 bases) identities are above the diagonal and the percentages of amino acid sequence (193 aa) identities are below the diagonal. RRACV1∼5 was found in infected N. occidentalis, and both SV5 and SV6 was found in an ARRD-apple. P205 is an isolate from apple (Sato et al., 1993) and B6 is an isolate from an ARRD-apple (Yaegashi et al., 2007).

seedlings showed early flowering from approximately 2–3 months after inoculation, and the flowered plants were artificially pollinated. Finally, the number of apple seedlings with fruits was found to be one with RRACV1, two with RRACV2, two with RRACV3s, and one with RRACV5. We could not obtain apple seedlings with fruits infected with RRACV1 + 4.

Among the infected apple seedlings, a plant inoculated with RRACV2 (No. 1) showed a ring and line pattern of chlorosis on the leaf about 2 months after inoculation (**Figure 2A**, left). The fruits of the pollinated seedlings grew to 5–6 cm in diameter 5 months after pollination. In plant No. 1 infected with RRACV2, a dark green ring was observed on the fruits as shown in **Figure 2A**, center. Ring-shaped rust was observed on the fruit of another seedling infected with RRACV2 (no. 2) (**Figure 2A**, right). Though these apple seedlings were co-infected with ACLSV-RRACV2 and ALSV-AtFT/MdTFL1, ALSV-AtFT/MdTFL1 has never caused any symptoms in both leaves and fruits of infected apple seedlings (Yamagishi and Yoshikawa, 2016), and it is most likely that the symptoms observed were induced by RRACV2 infection. No symptoms were observed on either leaves or fruits of apple seedlings infected with RRACV1, RRACV3, or RRACV5. RT-PCR assays indicated that ACLSV was detected in fruit peel and flesh tissues of apple seedlings infected with RRACV1, RRACV2, RRACV3, or RRACV5, regardless of whether symptoms were observed.

ACLSV-RRACV2 Causes ARRD in Apple 'Golden Delicious'

All grafted apple trees (APT1 with RRACV1, APT2 with RRACV2, APT3 with RRACV3, APT1 + 4 with RRACV1 + RRACV4, and APT5 with RRACV5) produced fruits from a year after grafting, but no symptoms were observed on either leaves or fruits. Two years after grafting, the apple tree inoculated with RRACV2 (APT2) showed virus-like symptoms consisting of ring and line patterns of chlorosis on leaves as shown in **Figure 2B**; however, no symptoms were found on the fruits obtained this year. Three years after grafting, russet-ring symptoms typically of ARRD were observed on the fruits of APT2, in addition to ring- and line-patterns on leaves (**Figure 2B**). Other apple trees (APT1, APT3, APT1 + 4, APT5, and non-grafted control apple plants) did not show any symptoms on either leaves or fruits even 4 years after grafting. Thus, only apple trees inoculated with RRACV2 showed russet ring symptoms typical of ARRD. Amplification of the CP regions by RT-PCR and sequence analysis of the DNA products (6 clones per apple plant) from APT1, APT2, APT3, APT1 + 4, and APT5 indicated that only RRACV1 was detected from APT1, only RRACV2 from APT2, only RRACV3 from APT3, and only RRACV5 in APT5 (**Supplementary Table S3**). On the other hand, only RRACV1 was amplified from APT1 + 4, suggesting that RRACV4 might be lost from APT1 + 4 (**Supplementary Table S3**). ALSV-AtFT/MdTFL1 was not detected in any of the leaves of all apple trees (APT1–APT5) because we found that it is very difficult to transmit ALSV from infected to uninfected apple trees by grafting (unpublished result). From these results, we concluded that ACLSV-RRACV2 is a causal agent of ARRD.

We have determined the complete nucleotide sequences of the RRACV2 and RRACV1 genomes. The genome of RRACV2 consists of 7557 nt excluding the 3′-polyA tail (GenBank/EMBL/DDBJ accession LC533838) and encodes three proteins from the 5′-terminus: an RNA replication-associated protein (Rep) [216 kilo dalton (kDa), 1883 aa], a movement protein (MP) (50 kDa, 460 aa), and a CP (21 kDa, 193 aa). The genome of RRACV1 consists of 7562 nt (accession LC533837) and contains three ORFs encoding a Rep (216 kDa, 1885 aa), an MP (50 kDa, 460 aa), and a CP (21 kDa, 193 aa). Phylogenetic analysis based on the complete nucleotide sequence of the ACLSV isolates showed that RRACV2 is grouped into the B6 type and RRACV1 into the P205 type (**Figure 3**). The RRACV2 genome sequence was closest to that of B4 isolated from

FIGURE 2 | Symptoms of the apple plants inoculated with ACLSV-RRACV 2. **(A)** Ring pattern of chlorosis on a leaf (left) and ring pattern on the fruits (center and right) of apple seedlings co-inoculated with RRACV2 and ALSV-AtFT/MdTFL1. **(B)** An apple tree ('Golden delicious') 3 years after grafting with cut branches infected with ACLSV-SV. Arrows in a left photo indicated the positions of grafting. Ring and line patterns of chlorosis on leaves of an apple tree (AP2) infected with ACLSV-RRACV2 (upper right) and young fruits showing russet-ring symptom on an apple tree (AP2) infected with RRACV2 three years after grafting (lower right). Scale bar; ca. 1 cm.

another apple tree affected by ARRD, and the identities of nucleotide and amino acid sequences between RRACV2 and B6 were 89.4% (complete genome), 94.3% (Rep), 93.9% (MP), and 98.0% (CP).

Eight ASPV-SVs Were Detected in an Apple Tree (P-190) Affected by AGCD

RT-PCR analysis of inoculated *N. occidentalis* plants showed that all plants (10 out of 10 inoculated plants) were infected with ASPV (**Supplementary Figure S7**). The PCR products (745 bp of the CP region of the ASPV genome) were purified from 10 infected plants and analyzed by direct sequencing using the SPCP8148-8167(+) and SPCP8870-8890(−) primer pair. We expected that some plants were infected with a single different SV, as was the case with RRACV; however, all *N. occidentalis* plants were found to be infected with multiple SVs, judging from the fluorescence peaks in the sequences of the amplified DNA (data not shown). Next, the amplified DNAs from 5 infected *N. occidentalis* plants were cloned as described above, and the resulting DNA clones were sequenced (7–10 DNA clones per infected plant). Among apple seedlings infected with both the ASPV-SV and ALSV-AtFT/MdTFL1 vector, we selected five seedlings for the amplification of their CP region, followed by cloning of the DNA products; DNA clones (13–18 clones per infected apple seedling) were also sequenced. Sequence analysis of DNA clones from 10 infected plants (5 *N. occidentalis* and 5 apple seedlings each) indicated that a total of eight SVs were detected in plants inoculated with RNA transcripts from apple (P-190) affected by AGCD. The SVs were named GCSPV1 to GCSPV8 (**Supplementary Figure S8**). Sequence identities of the amplified CP region of the ASPV genome (705 bp excluding primer sequences) between GCSPV1 and GCSPV8

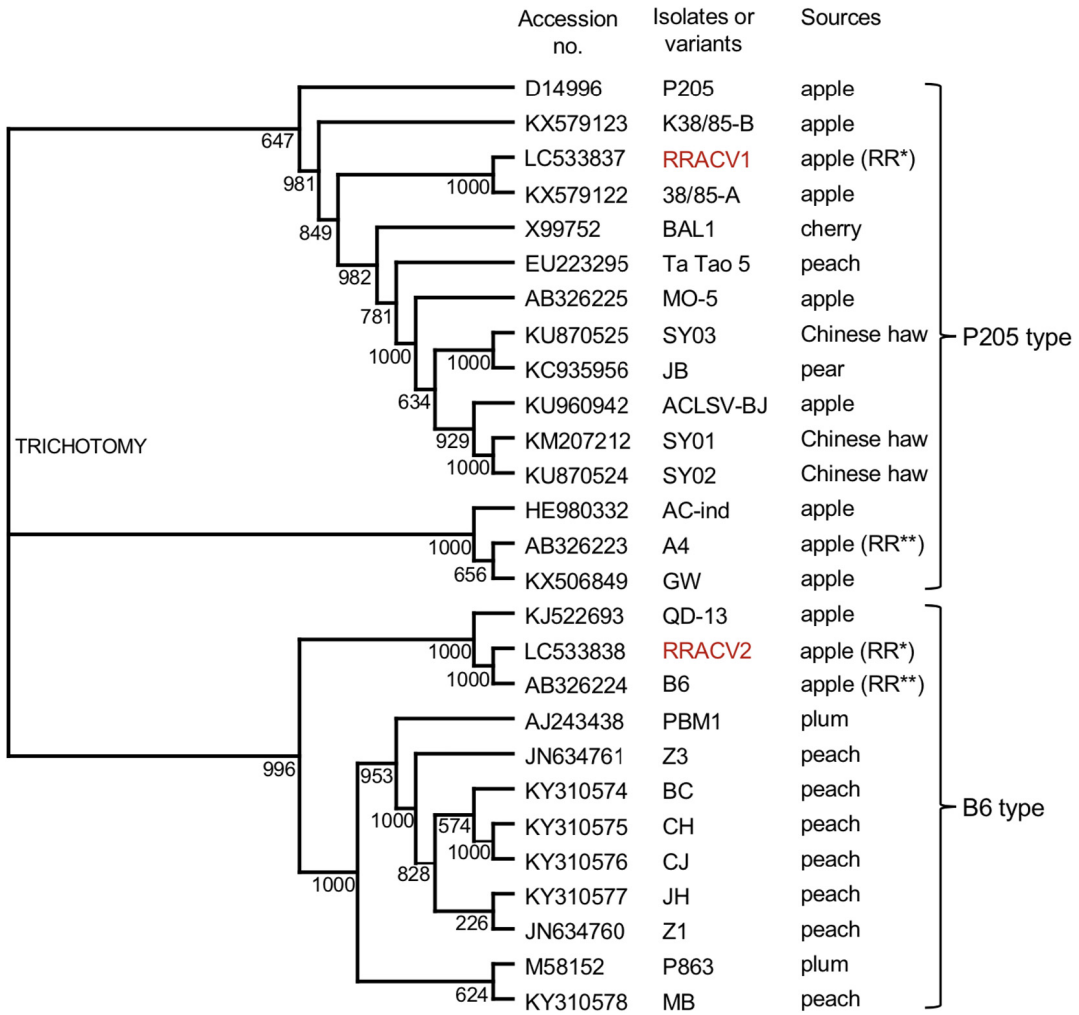


FIGURE 3 | Phylogenetic tree based on the complete nucleotide sequences of ACLSV genomes. The tree was generated by CLUSTALW program (DNA databank of JAPAN [DDBJ]; URL: http://clustalw.ddbj.nig.ac.jp/). Bootstrap values for each node are shown (1000 replicates). Sequence variants analyzed in this study are shown in red. Apple (RR*) indicated an apple tree affected with ARRD used in this study and apple (RR**) is an another apple tree affected with ARRD described before (Yaegashi et al., 2007).

were 75.4–89.7% in terms of nucleotides and 83.8–91.8% in terms of amino acid identities (**Table 3**). In total, 119 DNA clones were sequenced, and the number of each SV clone was observed to be as follows: 2 (1.7% of the whole) for GCSPV1, 31 (26.1%) for GCSPV2, 10 (8.4%) for GCSPV3, 2 (1.7%) for GCSPV4, 45 (37.8%) for GCSPV5, 12 (10.1%) for GCSPV6, and 14 (11.8%) for GCSPV7, and 3 (2.5%) for GCSPV8. The types and clone numbers of the SVs were different in each infected plant. For example, the type (the number of clones) of GCSPV detected in apple seedlings No. 1–No. 5 were GCSPV2 (13 clones) and GCSPV6 (1 clone) in the No. 1 plant, GCAPV6 (9) and GCAPV7 (9) in the No. 2 plant, GCAPV5 (6), GCAPV6 (1), GCAPV7 (3), and GCAPV8 (3) in the No. 3 plant, GCAPV5 (14) in the No. 4 plant, and GCAPV5 (12), GCAPV6 (1), and GCAPV7 (2) in the No. 5 plant (**Table 4**). Three SVs (GCSPV1, GCSPV3, and GCSPV4) were detected only from infected *N. occidentalis* plants and not from the infected apple seedlings.

ASPV-GCSPV2 Probably Induces Green Crinkle Symptoms in the Fruit of Inoculated Apple Seedling

Apple seedlings (No. 1 and No. 5) inoculated with RNA transcripts of ASPV and ALSV-AtFT/MdTFL1 vector flowered 2–3 months after inoculation and produced fruits, as shown in **Figure 4**. The infected apple seedlings did not show any symptoms on leaves, but a fruit on the No. 1 apple seedling infected with GCSPV2 + GCSPV6 showed cracks on the fruit surface in a manner similar to the symptoms of AGCD (**Supplementary Figure S1**). Other apple seedlings (Nos. 2, 3, 4, and 5) infected with ASPV-SVs and control seedlings with ALSV-AtFT/MdTFL1 only produced normal shaped apple fruits

TABLE 3 | Identities of the nucleotide and amino acid sequences of the coat protein region (705 bp) between the sequence variants (SV) of ASPV found in *N. occidentalis* and apple seedlings inoculated with the RNA transcript of ASPV from AGCD-apple.

	GCSPV1	GCSPV2	GCSPV3	GCSPV4	GCSPV5	GCSPV6	GCSPV7	GCSPV8
GCSPV1		79.2	89.7	78.2	79.6	83.1	80.9	79.3
GCSPV2	87.2		77.9	82.8	85.8	75.4	78.6	84.1
GCSPV3	91.8	85.4		79.6	80.2	85.5	80.2	80.2
GCSPV4	86.6	89.3	85.5		83.3	76.9	79.6	88.4
GCSPV5	89.7	91.8	88.4	89.3		78.2	80.8	83.5
GCSPV6	88.8	85.5	89.7	84.6	88.0		82.1	78.5
GCSPV7	87.1	86.7	84.1	83.8	89.3	86.7		80.5
GCSPV8	89.3	90.1	87.6	90.6	91.0	86.8	86.7	

TABLE 4 | Analysis of the sequence variants (GCSPVs) of ASPV found in leaves of apple seedlings (No. 1∼ No. 5) inoculated with ASPV-RNA transcripts and ALSV-AtFT/MdTFL1.

Plant no. apple seedlings	GCSPV (number of clone)								No. of clones sequenced
	1	2	3	4	5	6	7	8	
1	–	+ (13)	–	–	–	+(1)	–	–	14
2	–	–	–	–	–	+ (9)	+ (4)	–	13
3	–	–	–	–	+ (6)	+ (1)	+ (3)	+ (3)	13
4	–	–	–	–	+ (14)	–	–	–	14
5					+ (12)	+ (1)	+ (2)	–	15
Total no. of clones	0	13	0	0	32	12	9	3	

The CP region (705 nt) of ASPV genome was amplified by RT-PCR and the DNA products were cloned and sequenced.

(**Figure 4**). We then extracted RNA from the peel and flesh tissues of fruit showing cracks on the apple seedling No. 1, and subjected the RNA to RT-PCR, in which we found that ASPV was distributed in the fruit showing cracks. Cloning and sequence analysis of the amplified DNA products from the fruit indicated that twenty DNA clones from the fruit tissues had the same sequence as GCSPV2, whereas GCSPV6 was not detected in the fruit tissues. Furthermore, apple seedlings Nos. 2, 3, and 5 infected with GCSPV6 showed no symptoms on the fruits (**Table 4** and **Figure 4**). These results strongly suggest that one of the ASPV sequence variants, namely GCSPV2, probably serves as one of the causal agents of AGCD.

We have determined the complete nucleotide sequences of the GCSPV2 genomes using DNA clones amplified from a fruit showing symptoms. The genome of GCSPV2 consists of 9294 nt (accession LC533839) excluding the 3′-poly A tail and encodes five proteins from the 5′-terminus; an RNA replication-associated protein (Rep) (248 kDa, 2018 aa), triple gene block (TGB) proteins [TGB1 (25 kDa, 223 aa), TGB2 (13 kDa, 120 aa), TGB3 (8 kDa, 75 aa)] and a CP (42 kDa, 396 aa). Phylogenetic analysis based on the complete nucleotide sequence of 28 isolates of ASPV, 2 of AGCaV, and 5 species of other foveaviruses showed that these ASPV isolates (including AGCaV) could be grouped into 4 clades (I ∼ IV), and that GCSPV2 was most closely related to IF38 from an apple in clade III (**Figure 5**). The identity of the complete nucleotide sequence between GCSPV2 and IF38 was 93.7% and the identities of the amino acid sequences between GCSPV2 and IF38 were 97.8% (Rep), 97.8% (TGB1), 97.5% (TGB2), 100% (TGB3), and 96.7% (CP). ASPV-IF38 was originally isolated from 'Fuji' apple in Iwate University. Morioka, Japan which is not affected with AGCD (Yoshikawa et al., 2001). The nucleotide sequence identities of the genomes between ASPV-GCSPV2 and AGCaV-aurora were 76.6%, and their amino acid identities between them were 88.8% (Rep), 90.6 (TGB1), 84.2% (TGB2), 84.5% (TGB3), and 76.2% (CP).

DISCUSSION

To investigate viruses in fruit trees, interest in using comprehensive NGS analysis has been increasing in recent years (Coetzee et al., 2010; Ho and Tzanetakis, 2014; Rott et al., 2017; Maliogka et al., 2018). Virome analysis using NGS is expected to be a technique used to search for unknown viral pathogens (Ho and Tzanetakis, 2014; Pecman et al., 2017). For this reason, we first conducted a virome analysis of diseased apple trees in order to determine the types of viruses present in apple trees affected by ARRD and AGCD. Polyacrylamide gel electrophoresis of the dsRNAs extracted from ARRD- and AGCD-infected samples showed several high molecular weight bands as shown in **Supplementary Figure S2**, which is consistent with the dsRNA patterns of plants infected with ACLSV, ASGV, or ASPV, as reported previously (Yoshikawa and Takahashi, 1988; Magome et al., 1997a; Yoshikawa, 2008). In virome analysis of dsRNAs from diseased samples, we found that five known viruses (ACLSV, ASGV, ASPV, AGCaV, and ApLV) have been infecting both ARRD and AGCD apple trees, and that no new viruses were detected in either type of diseased apples. It was reported that all apple trees with AGCD are infected with ACLSV, ASGV, and/or ASPV (Pacific Northwest Plant Disease Management

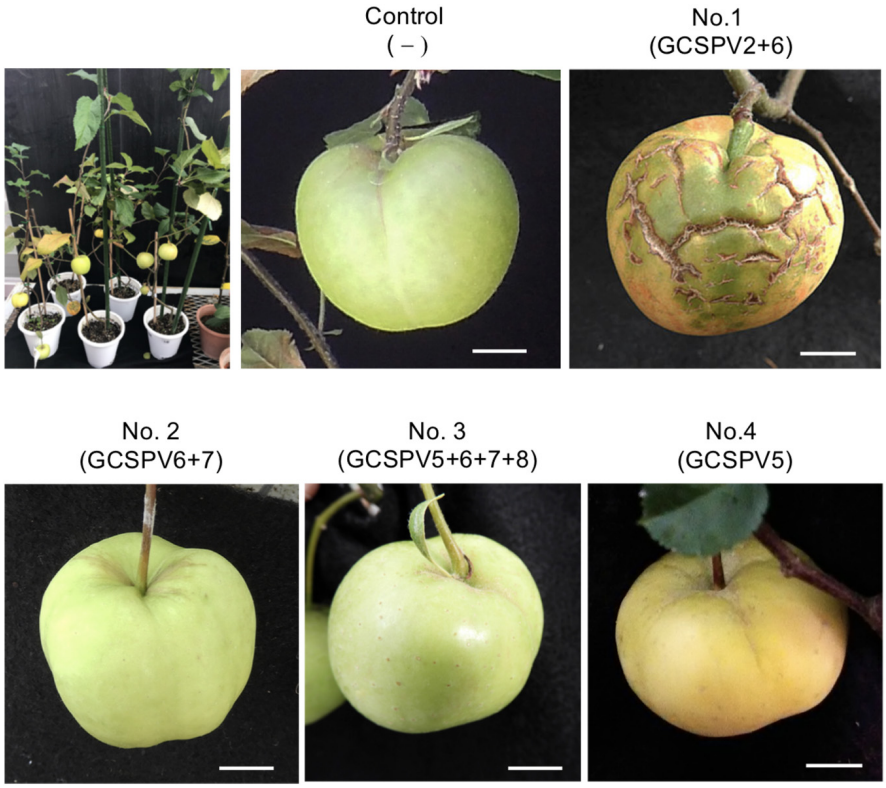


FIGURE 4 | Apple seedlings inoculated with RNA transcripts of ASPV and ALSV-AtFT/MdTFL1 about 10 months after inoculation (upper left) and fruits produced on the inoculated apple seedlings. Control; infected with ALSV-AtFT/MdTFL1 only, No. 1; with GCSPV2 and GCSPV6, No. 2; with GCSPV6 and GCSPV7, No. 3; with GCSPV5, GCSPV6, GCSPV7, and GCSPV8, No. 4 with GCSPV5.

Handbook[2]; Viral diseases, WUS Tree Fruit[3]). Deep sequencing analysis of an apple sample from Korea showed that these five viruses were identified from eight apple samples showing growth retardation (Cho et al., 2016). In Japan, ACLSV, ASGV, and ASPV have been reported as causative viruses of the top working diseases in apple fruits growing on crab apples (*Malus prunifolia* var. ringo, *M. sieboldii*, and *M. sieboldii arborescens*) as rootstocks (Yanase, 1974). On the other hand, most domesticated orchard apple plants (*M. domestica*) were latently infected with ACLSV, ASGV, and/or ASPV and these are widely distributed as latent viruses in cultivated apples in Japan.

To investigate the association between a specific disease and a particular virus, it is necessary to isolate each virus from the diseased apple trees, and to infect each virus into separate herbaceous plants. In general, leaf extracts of diseased samples were sap-inoculated to different test plants for the separation of target viruses. However, it is difficult to inoculate each apple virus into separate plants because the available hosts (*C quinoa* and *N. occidentalis* etc.) of apple viruses are common for two or three apple viruses (Yanase, 1974; Koganezawa and Yanase, 1990). For that reason, we amplified the full-length DNAs of viral genomic RNAs by RT-PCR using primers containing the T3 promoter sequence, and the full-length viral RNAs were transcribed *in vitro* from the amplified DNAs as templates (**Supplementary Figure S3**). Inoculation of RNA transcripts into herbaceous plants resulted in a high infection rate (50–100% depending on virus), and we were able to infect each virus coexisting in an ARRD-infected apple tree into separate herbaceous plants (**Supplementary Figure S3**). This method of using *in vitro* viral RNA transcripts is a very effective method for isolating and separating each virus from fruit trees with multiple viruses, if the entire genome sequences of the target viruses are available. Because the apple russet ring disease is implicated to be caused by a strain (or by strains) of ACLSV (van der Meer, 1986; Desvignes and Boye, 1988; Wood, 2001; Howell et al., 2011), in our subsequent experiments, we considered ACLSV as a candidate for the ARRD causal virus.

For the identification of the pathogenic viruses of diseases, it is important to demonstrate not only the viral species but also their SVs in the diseased trees. The present study indicated that at least seven SVs of ACLSV were found in leaf samples of ARRD-apple (PK-51) (**Table 2**, **Figure 1**, and **Supplementary Figure S5**). Though the data are not shown, multiple SVs were also found for ASGV and ASPV in the ARRD-apple. Similarly, at least eight SVs of ASPV were detected in the leaf samples of AGCD-apple (P-190) (**Table 3** and **Supplementary Figure S8**). Just as multiple virus species gathered in a single tree, each

[2]https://pnwhandbooks.org/node/2133

[3]http://treefruit.wsu.edu/web-article/viral-diseases/

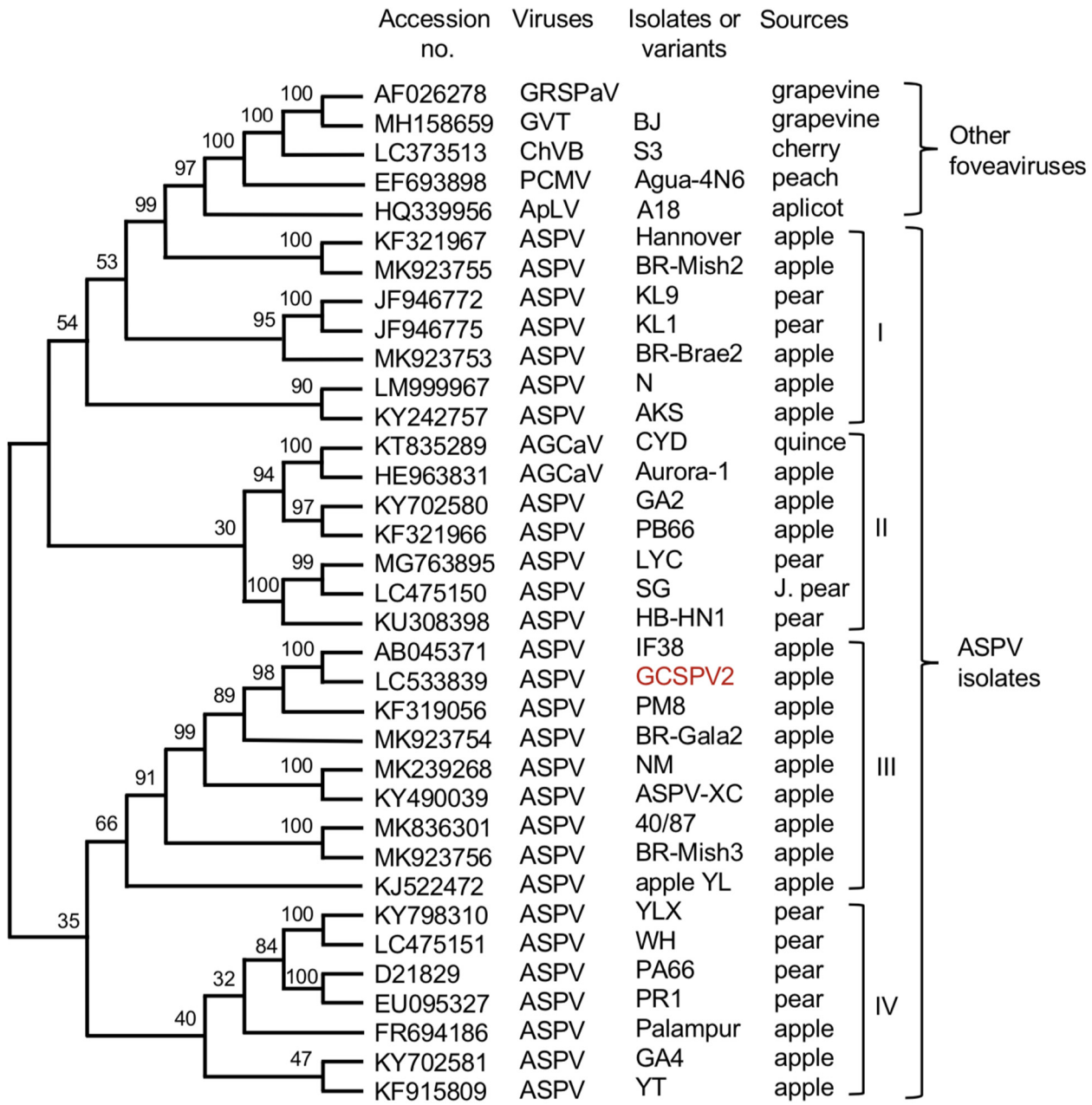


FIGURE 5 | Phylogenetic tree based on the complete nucleotide sequences of ASPV isolates, AGCaV isolates, and five other foveaviruses including grapevine rupestris stem pitting associated virus (GRSPaV), apricot latent virus (ApLV), grapevine virus T (GVT), peach chlorotic mottle virus (PCMV), and cherry virus B (ChVB). A sequence variant analyzed in this study is shown in red. The tree was generated by CLUSTALW program (DNA databank of JAPAN [DDBJ]; URL: http://clustalw.ddbj.nig.ac.jp/). Bootstrap values for each node are shown (1000 replicates).

SV in the same virus is likely to have been brought into a single tree by grafting widely used in cultivation of fruit trees in Japan (Magome et al., 1997b). We previously reported the use of identical ALSV labeled with yellow and cyan fluorescence proteins (YFP and GFP) replicated predominantly in discrete areas and separately distributed in leaves coinfected with ALSV-YFP and ALSV-CFP (Takahashi et al., 2007). Spatial separation was also found in plants co-infected with bean yellow mosaic virus (BYMV)-YFP and BYMV-CFP (Takahashi et al., 2007) and plum pox virus (Dietrich and Maiss, 2003), indicating that the separate distribution of virus population may be a universal phenomenon among the same virus species. Considering the distribution of SVs of apple viruses in apple plants, it is assumed that many SVs of each virus are spatially separated from each other in either mosaic or patch form within the tissues and leaves of infected trees. In PK-51 affected by ARRD, there are

obvious year-to-year fluctuations in which symptoms appear only for a few fruits in a shoot, or with a relatively large number of symptoms appearing in the same tree. Although temperatures during spring to early summer are reported to have a significant effect on the expression of fruit symptoms (Wood, 1971), another reason for this fluctuation may be due to the interaction between sequence variants coexisted in the infected tree. That is, if the pathogenic variant predominates, symptoms will appear; otherwise, it will not appear at all.

In this study, we reported that one of the SVs of ACLSV (RRACV2), but not other SVs (RRACV1, RRACV3, and RRACV5), caused leaf and fruit symptoms characteristic of ARRD in apple trees. van der Meer (1986) reported that ACLSV isolates from ARRD-apples were graft-transmitted from *C. amaranticolor* to the apple seedlings and some isolates induced russet rings on apple cultivars. He also mentioned that clearly different ACLSV strains can be obtained from the same tree. Desvignes and Boye (1988) also reported that russet rings are very likely to be caused by particular isolates of ACLSV, because back inoculation from *C. quinoa* to 'Golden Delicious' was positive. Although these two reports did not characterize ACLSV as a single isolate, the results obtained in this study are consistent with those reports. Wood (2001) reported that at least six distinct strains or forms appear to be present, assuming that all of the russet ring disorders found in New Zealand apple trees are related. Further study is needed to determine whether ACLSV-RRACV2 alone causes ARRD, or if other isolates/strains also cause ARRD. An isolate B6 of ACLSV was isolated from an ARRD-apple different from PK-51 (Yaegashi et al., 2007), and B6 is very closely related to ACLSV-RRACV2 in the phylogenetic tree (**Figure 3**). It would be interesting to test if B6 induces ARRD in apple trees. As most ACLSV strains or isolates do not cause ARRD in apple (Yanase, 1974; Yaegashi et al., 2011), it is important to identify which genes, or nucleotides, and amino acid sequences in the ACLSV-RRACV2 genome are involved in the induction of leaf and fruit symptoms characteristic to ARRD.

For AGCD, it is assumed that ASPV or AGCaV may be a causal virus (Howell et al., 2011; James et al., 2013); however, there are no reports of back-inoculation of the isolated virus to apple plants. In AGCD, the symptoms appear only on the fruits and not on the leaves, so we must observe the fruits of the back-inoculated apple plants. In this study, we used VIF technology using the ALSV vector to accelerate apple fruit formation, because it takes at least 6–7 years for apple seedlings to blossom. ALSV does not induce any symptoms on leaves and fruits (Yamagishi et al., 2014).

The apple tree (P-190) affected by AGCD tested in this study showed symptoms in almost all fruits every year. Our results indicated that the apple tree (P-190) with AGCD has a complex viral infection because the plant contained at least eight SVs (GCSPV1- GCSPV8) (**Table 3**), even for ASPV alone. The type and proportion of SVs were different among the inoculated plants, even when using the same RNA transcript as inoculum (**Table 4**). We showed that one of the sequence variants (GCSPV2) probably induces severe cracks on the fruit surface (apple seedling No. 1 in **Figure 4**), which is very similar to the AGCD symptoms shown in **Supplementary Figure S1**. Though the apple seedling No. 1 was infected with both GCSPV2 and GCSPV6 (**Table 4**), most of the variants detected were GCSPV2 (13/14 clones) (**Table 4**) and only GCSPV2 was detected from fruit showing cracks (20/20 clones). Furthermore, apple plants infected with other SVs including GCSPV6 produced normal-shaped apple fruits (**Table 4** and **Figure 4**). ASPV-GCSPV2 was distributed in the fruit and showed cracks, strongly suggesting that ASPV-GCSPV2 might be one of the AGCD causal agents. We plan to inoculate the virus-free 'Golden Delicious' trees with branches of infected trees (No. 1 to No. 5) by grafting in ARRD to further confirm the ASPV-SVs pathogenicity this year.

Since Jelkmann (1994) first reported the nucleotide sequence of the pear ASPV genome, the complete nucleotide sequences of many isolates from apple, pear, and quince have been reported, as shown in **Figure 5**. There is a great genetic diversity among ASPV isolates, and some isolates had less than about 72% nt identity (or 80% aa identity) between their CP or polymerase genes, which corresponds to the species demarcation criteria in the *Foveavirus* genus (Youssef et al., 2011; Adams et al., 2012). James et al. (2013) reported that a putative new foveavirus (AGCaV), or a variant or strain of ASPV was isolated from an Aurora Golden Gala apple showing severe symptoms of AGCD. A phylogenetic analysis based on the complete nucleotide sequences showed that ASPV isolates might be grouped into 4 clades, and that GCSPV2 was in clade III with IF38 from apple (**Figure 3**). On the other hand, AGCaV (aurora-1) was grouped into clade II (**Figure 3**), suggesting that multiple ASPV variants including AGCaV may cause AGCD. It will be interesting to determine which gene(s) in the ASPV genome are involved in the induction of fruit symptoms of AGCD.

As previously mentioned, there are virus-like diseases of deciduous fruit trees whose causal agents are still not identified. The strategy for fulfillment of the Koch's Postulates presented here (**Supplementary Figure S9**) can provide a system to prove if the virus found in diseased tissues is the same as the pathogen causing diseases in fruit trees.

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

CL amplified the full-length DNA, transcribed infectious RNAs, and analyzed the sequence variants of ASPV. HY conducted a phylogenetic analysis of ACLSV and ASPV. RK amplified the full-length DNA, transcribed infectious RNAs, and analyzed the sequence variants of ACLSV. AK analyzed the viruses of apple plants back-inoculated with ACLSV and ASPV and determined the complete nucleotide sequence of the ASPV-GCSPV2 genome. NYa carried out the inoculation of virus RNA to apple plants by particle bombardment and determined the complete nucleotide sequences of the ACLSV-RRACV1 and RRACV2 genome. TI

conducted the graft-transmission tests. NYo is the principal investigator, and he supervised the experiments and wrote the manuscript. All the authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by JSPS KAKENHI Grant Number 24380027.

ACKNOWLEDGMENTS

We would like to thank Mr. Yudai Kikuchi for technical assistant.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01627/full#supplementary-material

REFERENCES

Adams, M. J., Candresse, T., Hammond, J., Kreuze, J. F., Martelli, G. P., Namba, S., et al. (2012). *Virus Taxonomy: Classification and Nomenclature of Viruses*. San Diego, CA: Elsevier, 920–941.

Candresse, T., Lanneau, M., Revers, F., Macquaire, G., German, S., Dunez, J., et al. (1995). An immunocapture PCR Assay adapted to the detection and the analysis of apple chlorotic leaf spot virus. *Acta Hort.* 386, 136–147. doi: 10.17660/actahortic.1995.386.17

Chamberlain, E. E., Atkinson, J. D., Hunter, J. A., and Wood, G. A. (1974). Green crinkle disease of apple. *N. Z. J. Agric. Res.* 17, 137–146. doi: 10.1080/00288233.1974.10420992

Cho, I.-S., Igori, D., Lim, S., Choi, G.-S., Hammond, J., Lim, H.-S., et al. (2016). Deep sequencing analysis of apple infecting viruses in Korea. *Plant Pathol. J.* 32, 441–451. doi: 10.5423/ppj.oa.04.2016.0104

Coetzee, B., Freeborough, M. J., Maree, H. J., Celton, J. M., Rees, D. J., 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

Desvignes, J. C., and Boye, R. (1988). Different diseases caused by the chlorotic leaf spot virus on the fruit trees. *Acta Hort.* 235, 31–36.

Dietrich, C., and Maiss, E. (2003). Fluorescent labeling reveals spatial separation of potyvirus populations in mixed infected *Nicotiana benthamiana* plants. *J. Gen. Virol.* 84, 2871–2876. doi: 10.1099/vir.0.19245-0

Ho, T., and Tzanetakis, I. E. (2014). Development of a virus detection and discovery pipeline using next generation sequencing. *Virology* 47, 54–60. doi: 10.1016/j.virol.2014.09.019

Howell, W. E., Thompson, D., and Scott, S. (2011). "Virus-like disorders of fruit trees with undetermined etiology," in *Virus and Virus-like Diseases of Pome and Stone Fruits*, eds A. Hadidi, M. Barba, T. Candresse, and W. Jelkmann (St. Paul, MN: APS press), 259–269.

Isogai, M., Aoyagi, J., Nakagawa, M., Kubodera, Y., Satoh, K., Katoh, T., et al. (2004). Molecular detection of five cherry viruses from sweet cherry trees in Japan. *J. Gen. Plant. Pathol.* 70, 288–291. doi: 10.1007/s10327-004-0129-4

James, D., Varga, A., Jesperson, G. D., Navratil, M., Safarova, D., Constable, F., et al. (2013). Identification and complete genome analysis of a virus variant or putative new foveavirus associated with apple green crinkle disease. *Arch. Virol.* 158, 1877–1887. doi: 10.1007/s00705-013-1678-7

Jelkmann, W. (1994). Nucleotide sequences of apple stem pitting virus and of the coat protein gene of a similar virus from pear associated with vein yellows disease and their relationship with potex- and carlavirus. *J. Gen. Virol.* 75, 1535–1542. doi: 10.1099/0022-1317-75-7-1535

Koganezawa, H., and Yanase, H. (1990). A new type of elongated virus isolated from apple trees containing the stem pitting agent. *Phytopathology* 91, 1085–1091.

Magome, H., Terauchi, H., Yoshikawa, N., and Takahashi, T. (1997a). Analysis of double-stranded RNA in tissues infected with apple stem grooving capillovirus. *Ann. Phytopathol. Soc. Jpn.* 63, 450–454. doi: 10.3186/jjphytopath.63.450

Magome, H., Yoshikawa, N., and Takahashi, T. (1999). Single-strand conformation polymorphism analysis of apple stem grooving capillovirus sequence variants. *Phytopathology* 89, 136–140. doi: 10.1094/phyto.1999.89.2.136

Magome, H., Yoshikawa, N., Takahashi, T., Ito, T., and Miyakawa, T. (1997b). Molecular variability of the genomes of capilloviruses from apple, Japanese pear, European pear, and citrus trees. *Phytopathology* 87, 389–396. doi: 10.1094/phyto.1997.87.4.389

Maliogka, V., Minafra, A., Saldarelli, P., Ruiz-Garcia, A. B., Glasa, M., Katis, N., et al. (2018). Recent advances on detection and characterization of fruit tree viruses using high-throughput sequencing technologies. *Viruses* 10:436. doi: 10.3390/v10080436

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

Németh, M. (1986). *Virus, Mycoplasma and Rickettsia Diseases of Fruit Trees*. Netherlands: Springer, 750.

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

Pecman, A., Kutnjak, D., Gutierrez-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

Rott, M., Xiang, Y., Belton, M., Saeed, P., Kesanakurti, P., Hayes, S., et al. (2017). Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroid. *Plant Dis.* 101, 1489–1499. doi: 10.1094/pdis-03-17-0306-re

Sato, K., Yoshikawa, N., and Takahashi, T. (1993). Complete nucleotide sequence of the genome of an apple isolate of apple chlorotic leaf spot virus. *J. Gen. Virol.* 74, 1927–1931. doi: 10.1099/0022-1317-74-9-1927

Sawamura, K. (1965). Studies on apple virus diseases I. On mosaic, Kikei-ka, and Sabi-ka diseases. *Bull. Fruit Tree Res. Stn. C* 3, 25–33.

Takahashi, T., Sugawara, T., Yamatsuta, T., Isogai, M., Natsuaki, T., and Yoshikawa, N. (2007). Analysis of the special distribution of identical and two distinct virus populations differently labeled with cyan and yellow fluorescent proteins in coinfected plants. *Phytopathology* 97, 1200–1206. doi: 10.1094/phyto-97-10-1200

van der Meer, F. A. (1986). Observation on the etiology of some virus diseases of apple and pear. *Acta Hort.* 193, 73–74. doi: 10.17660/actahortic.1986.193.11

Wood, G. A. (1971). Russet ring and some associated virus disorders of apple (*Malus sylvestris* (L.) Mill.) in New Zealand. *N. Z. J. Agric. Res.* 15, 405–412. doi: 10.1080/00288233.1972.10421269

Wood, G. A. (2001). Sensitivity of apple (*Malus domestica*) indicator cultivars to russet ring disease, and the results of graft-transmission trials of other fruit-affecting disorders of apple. *N. Z. J. Crop Hort. Sci.* 29, 255–265. doi: 10.1080/01140671.2001.9514187

Yaegashi, H., Isogai, M., Tajima, H., Sano, T., and Yoshikawa, N. (2007). Combinations of two amino acids (Ala^{40} and Phe^{75} or Ser^{40} and Tyr^{75}) in the coat protein of apple chlorotic leaf spot virus are crucial for infectivity. *J. Gen. Virol.* 88, 2611–2618.

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 Path.* 86, 134–142. doi: 10.1007/s10327-019-00896-0

Yaegashi, H., Yoshikawa, N., and Candresse, T. (2011). "Apple chlorotic leaf spot virus in pome fruits," in *Virus and Virus-like Diseases of Pome and Stone Fruits*, eds A. Hadidi, M. Barba, T. Candresse, and W. Jelkmann (St. Paul, MN: APS press), 17–21. doi: 10.1094/9780890545010.004

Yamagishi, N., Kishigami, R., and Yoshikawa, N. (2014). Reduced generation time of apple seedlings to within a year by means of a plant virus vector: a new plant-breeding technique with no transmission of genetic modification to the next generation. *Plant Biotech. J.* 12, 60–68. doi: 10.1111/pbi.12116

Yamagishi, N., Li, C., and Yoshikawa, N. (2016). Promotion of flowering by *Apple latent spherical virus* vector and virus elimination at high temperature allow accelerated breeding of apple and pear. *Front. Plant Sci.* 7:171. doi: 10.3389/fpls.2016.00171

Yamagishi, N., Sasaki, S., and Yoshikawa, N. (2010). Highly efficient method for inoculation of apple viruses to apple seedlings. *Julius Kuhn Archiv.* 427, 226–229.

Yamagishi, N., and Yoshikawa, N. (2016). "A new plant breeding technique using ALSV vectors to shorten the breeding periods of fruit trees," in *Genetic Engineering- an Insight into the Strategies and Applications*, ed. F. James (London: InTech), 47–61.

Yanase, H. (1974). Studies on apple latent viruses in Japan-The association of apple topworking disease with apple latent viruses. *Bull. Fruit Tree Res. Stn. C* 1, 47–110.

Yanase, H., Koganezawa, H., and Yamaguchi, A. (1988). Occurrence of apple russet ring in Japan and its graft transmission. *Bull. Fruit Tree Res. Stn. C* 14, 53–60.

Yoshikawa, N. (2008). *Cappilovirus, Foveavirus, Trichovirus, Vitivirus. Encyclopedia of Virology*, *Third Edition*, Vol. 1. Cambridge, MA: Academic press, 419–427.

Yoshikawa, N., and Converse, R. H. (1990). Strawberry pallidosis disease: distinctive dsRNA species associated with latent infections in indicators and in diseased strawberry cultivars. *Phytopathology* 80, 543–548.

Yoshikawa, N., Matsuda, H., Oda, Y., Isogai, M., Takahashi, T., Ito, T., et al. (2001). Genome heterogeneity of apple stem pitting virus in apple trees. *Acta Hort.* 550, 285–290. doi: 10.17660/actahortic.2001.550.41

Yoshikawa, N., Sasaki, E., Kato, M., and Takahashi, T. (1992). The nucleotide sequence of apple stem grooving capillovirus genome. *Virology* 191, 98–105. doi: 10.1016/0042-6822(92)90170-t

Yoshikawa, N., Sasamoto, K., Sakurada, M., Takahashi, T., and Yanase, H. (1996). Apple stem grooving and citrus tatter leaf capilloviruses obtained from a single shoot of Japanese pear (*Pyrus serotica*). *Ann. Phytopathol. Soc. Jpn.* 62, 119–124. doi: 10.3186/jjphytopath.62.119

Yoshikawa, N., and Takahashi, T. (1988). Properties of RNAs and proteins of apple stem grooving and apple chlorotic leaf spot viruses. *J. Gen. Virol.* 69, 241–245. doi: 10.1099/0022-1317-69-1-241

Youssef, F., Marias, A., Faure, C., Barone, M., Gentit, P., and Candresse, T. (2011). Characterization of Prunus-infecting Apricot latent virus-like foveaviruses: evolutionary and taxonomic implications. *Virus Res.* 155, 440–445. doi: 10.1016/j.virusres.2010.11.013

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

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.

Copyright © 2020 Li, Yaegashi, Kishigami, Kawakubo, Yamagishi, Ito and Yoshikawa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 July 2020
doi: 10.3389/fpls.2020.01098

Genome Editing of *eIF4E1* in Tomato Confers Resistance to Pepper Mottle Virus

Yoo-Joung Yoon[1†], Jelli Venkatesh[1†], Joung-Ho Lee[1], Jinhee Kim[2], Hye-Eun Lee[2], Do-Sun Kim[2] and Byoung-Cheorl Kang[1]*

[1] Department of Plant Science and Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea, [2] Vegetable Research Division, National Institute of Horticultural and Herbal Science, RDA, Jeonju-si, South Korea

Many of the recessive virus-resistance genes in plants encode eukaryotic translation initiation factors (eIFs), including eIF4E, eIF4G, and related proteins. Notably, *eIF4E* and its isoform *eIF(iso)4E* are pivotal for viral infection and act as recessive resistance genes against various potyviruses in a wide range of plants. In this study, we used Clustered Regularly Interspaced Palindromic Repeats/CRISPR-associated protein 9 (CRISPR/Cas9)-mediated targeted mutagenesis to test whether novel sequence-specific mutations at *eIF4E1* in *Solanum lycopersicum* (tomato) cv. Micro-Tom could confer enhanced resistance to potyviruses. This approach produced heritable homozygous mutations in the transgene-free E_1 generation. Sequence analysis of *eIF4E1* from E_0 transgenic plants expressing *Cas9* and *eIF4E-sgRNA* transcripts identified chimeric deletions ranging from 11 to 43 bp. Genotype analysis of the *eIF4E1*-edited lines in E_0, E_1, and E_2 transgenic tomato plants showed that the mutations were transmitted to subsequent generations. When homozygous mutant lines were tested for resistance to potyviruses, they exhibited no resistance to tobacco etch virus (TEV). Notably, however, several mutant lines showed no accumulation of viral particles upon infection with pepper mottle virus (PepMoV). These results indicate that site-specific mutation of tomato *eIF4E1* successfully conferred enhanced resistance to PepMoV. Thus, this study demonstrates the feasibility of the use of CRISPR/Cas9 approach to accelerate breeding for trait improvement in tomato plants.

Keywords: CRISPR/Cas9, eukaryotic translation initiation factor 4E (*eIF4E*), genome editing, potyvirus, phytoene desaturase (*PDS*), *Solanum lycopersicum* cv. Micro-Tom

OPEN ACCESS

Edited by:
Henryk Hanokh Czosnek, Hebrew University of Jerusalem, Israel

Reviewed by:
Aiming Wang, Agriculture and Agri-Food Canada (AAFC), Canada
Kaijun Zhao, Chinese Academy of Agricultural Sciences, China

***Correspondence:**
Byoung-Cheorl Kang
bk54@snu.ac.kr

†These authors have contributed equally to this work

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

Received: *21 May 2020*
Accepted: *03 July 2020*
Published: *24 July 2020*

Citation:
Yoon Y-J, Venkatesh J, Lee J-H, Kim J, Lee H-E, Kim D-S and Kang B-C (2020) Genome Editing of eIF4E1 in Tomato Confers Resistance to Pepper Mottle Virus. Front. Plant Sci. 11:1098. doi: 10.3389/fpls.2020.01098

INTRODUCTION

Potyviridae is the largest family of plant RNA viruses, which account for about 30% of known plant viruses, cause considerable damage to crop plants (Ward and Shukla, 1991; Cui and Wang, 2019). The potyviruses tobacco etch virus (TEV), potato virus Y (PVY), chilli veinal mottle virus (ChiVMV), pepper mottle virus (PepMoV), and pepper veinal mottle virus (PVMV) infect numerous solanaceous plants, including tomato, potato, and pepper (Kothari et al., 2010;

Zhao et al., 2014; Hančinský et al., 2020). Potyviruses, such as PepMoV and TEV are potential threat to tomato crop (Wintermantel, 2011; Melzer et al., 2012).

Eukaryotic translation initiation factor (eIF) genes, such as eukaryotic translation initiation factor 4E (eIF4E), eukaryotic translation initiation factor (Iso) 4E (eIF(iso)4E), and eukaryotic translation initiation factor 4G (eIF4G), are required for RNA viruses to maintain their lifecycle (Sanfacon, 2015; Revers and García, 2015). The number of eIF4E family members is species-dependent. In tomato, the eIF4E family consists of two *eIF4E* homologs (*eIF4E1* and *eIF4E2*), one *eIF(iso)* 4E homolog (Lebaron et al., 2016). Notably, several eIF genes confer recessive resistance to one or more potyviruses in Solanaceae crops (Kang et al., 2005a; Ruffel et al., 2006; Kang et al., 2007; Lee et al., 2013). A single recessive resistance locus *pot-1* encoding eIF4E1 protein was identified from the tomato wild relative, *Solanum habrochaites* accession PI247087. *eIF4E1* confers resistance to several potyviruses, including PVY, TEV, and PepMoV (Ruffel et al., 2005; Kang et al., 2005a; Wang, 2015). Knockout mutants of tomato *eIF4E2* and *eIF(iso)4E* were reported to be fully susceptible to potyviruses, thus suggesting prominent role of eIF4E1 in potyviral resistance in Solanaceous crops (Robaglia and Caranta, 2006; Charron et al., 2008).

The eIF protein eIF4E is a component of a multiprotein complex that aids the initiation of protein translation by enabling recognition and interaction with the mRNA cap structure and recruitment of ribosomes. The viral genome-linked protein (VPg) of potyviruses is covalently connected to the 5′ end of viral RNA and acts as an analog of the eukaryotic mRNA cap structure during protein translation (Wang, 2015). The physical interaction between the host factor, eIF4E (or its homolog eIF (iso)4E), and VPg is crucial for potyvirus infectivity (Miyoshi et al., 2006; Robaglia and Caranta, 2006; Hwang et al., 2009; Kim et al., 2013; Kim et al., 2014; Tavert-Roudet et al., 2017; de Oliveira et al., 2019). Moreover, mutations in these host factors can inhibit the interaction between the host factor and the VPg, and consequently inhibit viral proliferation and host infection (Duprat et al., 2002; Lellis et al., 2002; Sato et al., 2005; Rodríguez-Hernández et al., 2012).

The majority of naturally occurring plant recessive resistance genes have been mapped to mutations in genes encoding the isoforms of the translation initiation factors eIF4E and eIF4G that hinder their interactions with viral RNAs or proteins (Kang et al., 2005a; Ruffel et al., 2006; Le Gall et al., 2011; Wang and Krishnaswamy, 2012; Lee et al., 2013). Natural recessive resistance to potyviruses has been exploited in numerous breeding programs (Kang et al., 2005b; Ruffel et al., 2006; Lee et al., 2013). However, conventional breeding requires massive backcrossing to introgress the trait of interest into an elite background; furthermore, the availability of favorable alleles in natural populations is limited. New alleles can be created by random mutagenesis (Nicaise, 2014), but this requires labor-intensive, time-consuming screening of large populations to select mutants with desirable properties.

Advances in genome-editing tools have accelerated site-directed mutagenesis in crops. Clustered regularly interspersed palindromic repeats (CRISPR)/CRISPR-associated protein 9 (CRISPR/Cas9) is a targeted genome-editing technique derived from the adaptive immune mechanism of *Staphylococcus pyogenes* against bacteriophages (Hsu et al., 2014). Since its first report in 2012, CRISPR/Cas9 has become the technology of choice for genome editing due to its ease, low cost, and significantly shorter timeframe for construct preparation compared to those of other genome-editing tools, such as zinc finger nucleases (ZFN) and transcription activator-like effector nucleases (TALENs) (Cong and Zhang, 2015). CRISPR/Cas9 has been utilized for site-directed mutagenesis in microbes, animals, human cells, and plants (Jiang et al., 2013; Wu et al., 2013; Zhang et al., 2017; Shapiro et al., 2018). Precise genome editing of host factors can be deployed for development of recessive genetic resistance against viral diseases in plants (Wang, 2015). However, the application of CRISPR/Cas9 tools to improve plant resistance to pathogens has not been widely explored, with only a few reports to date (Chandrasekaran et al., 2016; Pyott et al., 2016; Wang et al., 2016; Peng et al., 2017; Gomez et al., 2019). Recently, CRISPR/Cas9 genome editing of plants for potyvirus resistance has been reported. For instance, knockout of the gene *eIF(iso)4E* in the model crop plant *Arabidopsis thaliana* was shown to confer resistance to turnip mosaic virus (TuMV) (Pyott et al., 2016). In another study, CRISPR/Cas9-mediated knockout of *eIF4E* in cucumber resulted in broad-spectrum viral resistance to several plant viruses, including cucumber vein yellowing virus (CVYV), zucchini yellow mosaic virus (ZYMV), and papaya ringspot mosaic virus-w (PRSV-W) (Chandrasekaran et al., 2016). Similarly novel allelic variants of rice *eIF4G* generated through CRISPR/Cas9 mediated genome editing conferred resistance to rice tungro spherical virus (Macovei et al., 2018). Thus, targeted genome editing can be expected to accelerate plant breeding for disease-resistant crop plants by facilitating the introduction of precise and predictable genetic changes directly into an elite strain background.

In this study, to introduce allelic variations in the *eIF4E* gene, we mutated *eIF4E1* in the tomato cultivar Micro-Tom using CRISPR/Cas9 technology and *Agrobacterium*-mediated transformation, and bred the E_0 transgenic plants carrying *eIF4E1* mutations to produce E_1 and E_2 progeny. We then tested the resistance of the homozygous mutant lines to the potyviruses TEV and PepMoV. The homozygous Micro-Tom mutant lines carrying genome-edited *eIF4E1* were resistant to PepMoV, although not to TEV, and showed normal plant growth and development after PepMoV challenge. Furthermore, by segregation in the E_1 generation, we were able to select virus-resistant plants that carried an edited *eIF4E1* gene or genes but not the introduced transgene. Altogether, this study thus provides important information for understanding and analyzing tomato CRISPR/Cas9 mutants, and accelerating breeding for trait improvement.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Seeds of *Solanum lycopersicum* cv. Micro-Tom were surface sterilized in 70% ethanol for 1 min and in 2% NaOCl with one

drop of Tween 20 for 15 min, and then rinsed four or five times with sterilized water. Seeds were germinated on 1/2 MS medium (Murashige and Skoog, 1962) containing 20 g/L sucrose and 8 g/L agar agar. All cultures were grown at 24°C with a 16 h light/8 h dark cycle under cool fluorescent light. Cotyledons of 7–8-day-old seedlings, before the appearance of the first true leaves, were used as explants for tissue culture.

Single Guide RNA (sgRNA) Design and Vector Construction

eIF4E is a plant cellular translation initiation factor essential for potyvirus infection, and mutations in the *eIF4E* gene can confer resistance to potyviruses. In this study, we aimed to use CRISPR/Cas9-mediated targeted genome editing of *eIF4E1* (GenBank: AY723733) to develop potyvirus-resistant tomato plants. The gene *phytoene desaturase* (*PDS*) (GenBank: EF650011), encoding the key enzyme in carotenoid biosynthesis, was used as a control gene to test genome-editing efficiency due to the easily detectable photobleached phenotype of *PDS* mutants. To design sgRNAs, we identified appropriate target sgRNA sequences using the CCTop - CRISPR/Cas9 target online predictor (https://crispr.cos.uni-heidelberg.de/index.html). sgRNAs targeting the first exon with high prediction scores were used for genome editing (**Figure 1**), since mutations in the 5′ region or first exon would increase the chance of creating nonfunctional proteins by causing frameshifts or early stop codons. These sgRNAs were cloned under the control of the AtU6-26 promoter into a binary vector (pHSE401) carrying a maize codon-optimized *Cas9* gene driven by the CaMV 35S promoter. The 23-bp sequences of the corresponding primers, PDSgRNA_F and PDSgRNA_R (or eIF4E1gRNA_F and eIF4E1gRNA_R), flanked by a *Bsa*I recognition site, were annealed and cloned into a *Bsa*I site in the vector pHSE401 by Golden Gate cloning according to a previously reported method (Xing et al., 2014). CRISPR/Cas9 vectors were transformed into *Agrobacterium tumefaciens* strain GV3101 by electroporation.

Agrobacterium-Mediated Tomato Transformation

A single *Agrobacterium* colony was inoculated into 15 mL of liquid LB medium containing 50 μg/mL kanamycin and 50 μg/mL rifampicin and incubated in a shaking incubator at 28°C for 12–16 h until it reached an OD_{600} of 0.6. The *Agrobacterium* suspension was then centrifuged at 8,000 rpm for 10 min at 20°C to collect the pellet. The pellet was completely resuspended in liquid 1/2 MS medium containing 3% sucrose and 200 μM acetosyringone. Cotyledons from 7–8-day-old seedlings were excised under sterile conditions, and then the tip of each cotyledon was removed, sectioned transversely into two fragments, and incubated adaxial side down in a preculture medium consisting of MS with 30 g/L sucrose, 1 mg/L 1-naphthaleneacetic acid (NAA), 1 mg/L benzylaminopurine (BAP) for 2 days, and 8 g/L agar agar. Explants were co-cultured in the *Agrobacterium* suspension for 20 min, transferred to sterile filter paper to briefly drain excess suspension, and then placed on the same medium used for preculture and incubated for 2 days. Then, explants were transferred to shoot induction medium consisting of MS with 30 g/L sucrose, 2 mg/L *trans*-zeatin riboside, 0.1 mg/L indole-3-acetic acid (IAA), 20 mg/L hygromycin, 250 mg/L carbenicillin, and 8 g/L agar agar for 4–6 weeks. Explants with shoot buds were moved to shoot elongation medium consisting of MS with 30 g/L sucrose, 1 mg/L *trans*-zeatin riboside, 0.1 mg/L IAA, 20 mg/L hygromycin, 250 mg/L carbenicillin, and 8 g/L agar agar. Tomato shoots of about 2 cm height were cut and transferred to rooting medium consisting of MS with 30 g/L sucrose, 1 mg/L IAA, 10 mg/L hygromycin, 250 mg/L carbenicillin, and 8 g/L agar agar. Rooted plants were transferred into plastic pots containing potting mixture (Hanarum, Minong Fertilizer, Korea) and kept in a growth room maintained at 24°C and a 16 h light/8 h dark cycle.

Nucleic Acid Extraction and Molecular Characterization

Genomic DNA (gDNA) was extracted from leaf samples of putative transgenic plants by the cetyltrimethylammonium bromide (CTAB)

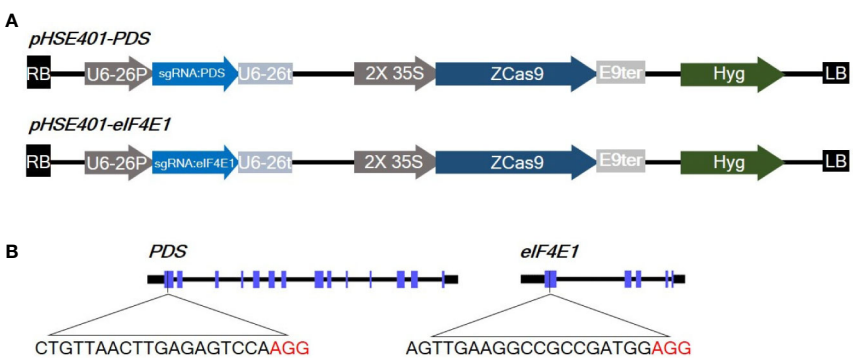


FIGURE 1 | Diagrams of the CRISPR/Cas9 vector constructs and sgRNA target sites **(A)** Diagram of the cassettes (pHSE401-PDS and pHSE401-eIF4E1) expressing the maize codon-optimized *Cas9* gene under the control of the 35S promoter, and the PDS- and eIF4E-sgRNA sequence driven by the Arabidopsis U6-26 polymerase III promoter. **(B)** Diagram of the CRISPR/Cas9 target sites within *PDS* and *eIF4E1*. The sgRNA target sequences are shown in black letters, followed by protospacer-adjacent motif (PAM) sequences in red.

method (Porebski et al., 1997). gDNA was quantified by NanoDrop spectrophotometer (Nanodrop Technology, Inc., Wilmington, DE, USA) and diluted to 100 ng/μL. The status of putative transgenic plants was confirmed by PCR using *Hpt* (*Hygromycin phosphotransferase*) and *Cas9* gene-specific primers (**Table 1**). PCR conditions were as follows: 94°C for 5 min, followed by 34 cycles of denaturing at 94°C for 30 s, annealing from 55°C for 30 s, 72°C extension for 30 s, and then final extension at 72°C 5 min. Total RNA was extracted from young leaf tissues using an MG RNAzol kit according to the manufacturer's instructions (MGmed, Seoul, Korea). The integrity and concentration of the total RNA were analyzed on a 1% agarose gel and a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA), respectively. One microgram of total RNA was used to synthesize complementary DNA (cDNA) using an EasyScript Reverse Transcriptase kit (TransGen, Beijing, China) with oligo(dT) primers. The resulting cDNAs were used for further expression analysis. RT-PCR was performed as described for gene confirmation using 1 μL cDNA as a template.

Mutation Detection

The transgenic plants were genotyped for mutation detection using primers flanking sgRNA target regions. PCR products were purified using a LaboPass PCR clean-up kit (Cosmo Genetech, Seoul, Korea). The purified amplicons were also cloned into the TA cloning vector, pMD20-T (Mighty TA-cloning kit, TAKARA, Shiga, Japan), according to the manufacturer's instructions, and positive colonies were selected by blue/white colony selection. Plasmids were extracted from least five positive clones and sequenced using M13F and M13R primers at Bionics (Seoul, Korea). To identify CRISPR/Cas9 induced mutations, DNA sequence alignments were performed using Lasergene's SeqMan program (DNASTAR, Madison, WI, USA). Mutations in E_1 progeny was detected by directly sequencing the PCR amplicons of the target region as well as by sequencing positive TA clones.

Virus Inoculations

For virus inoculum preparation, frozen stocks of TEV-HAT and PepMoV-Vb1, which were stored at –80°C, were used to inoculate 3-week-old *Nicotiana benthamiana* plants. Frozen inocula were ground in 0.1 M potassium phosphate buffer (pH 7.0), mixed with 400-grit carborundum, and rubbed on the lower leaves of *N. benthamiana*. After 10–20 min of inoculation, leaves were washed with distilled water (Hull, 2009). To inoculate the tomato plants, infected *N. benthamiana* leaves were collected and inocula were prepared as mentioned above. Tomato plants with two fully expanded leaves were used for viral inoculation. Two pairs of cotyledons were inoculated; inoculated and non-inoculated control plants (mock) were grown in a growth chamber (16 h light and 8 h night under white fluorescent light). To ensure viral infection, plants were reinoculated 7 days after the first inoculation.

Evaluation of Resistance to Potyviruses

After viral inoculation, plants were monitored regularly for the appearance of symptoms. Leaf tissue was tested for the presence of virus using double-antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA), according to the manufacturer's instructions (Agdia, Inc. Elkhart, IN, USA). Virus accumulation was tested at 7/20 days post inoculation (DPI). DAS-ELISA was performed to detect the accumulation of the coat protein (CP) of TEV or PepMoV. Three replicates of inoculated and upper non-inoculated leaves of E_1 lines were used for ELISA analysis. Absorbance of samples at 405 nm was measured using a microplate reader (Biotek, VT, USA). The statistical significance of the data was performed with Student's t-test using Microsoft Excel 2016 (Microsoft, Redmond, WA, USA).

RESULTS

Generation and Characterization of Genome-Edited Tomato Plants

To develop tomato plants with edited *PDS* and *eIF4E1* genes, we delivered the pHSE401-PDS and pHSE401-eIF4E1 CRISPR/Cas9 constructs harboring the respective sgRNAs into *S. lycopersicum* cv. Micro-Tom *via Agrobacterium* transformation. We transferred putatively genome-edited shoots regenerated from the callus to shoot elongation medium and allowed them to elongate, and then cut elongated shoots and transferred them to rooting medium to encourage root formation. We observed the photobleached phenotype in four out of 113 *PDS*-edited explants transformed with the pHSE401-PDS construct (**Figure 2A**), demonstrating

TABLE 1 | List of primers used in the present study.

Name	Primer sequence (5′ to 3′)	Amplicon size (bp)	Purpose
Sl4E_F_2	ACACTATGGTCCAAACAGTTCTTAT	330	Mutation detection
Sl4E_R_2	AACTGCTTGGGGAAGCTCAC	330	
SlPDS275_F	TGCTTCTCAACATAAATCTTGACAAAGAGAAGGA	275	
SlPDS275_R	CAAACCAAACCTTTAAAGGCCCCAAGT	275	
HygR_F	GCGAAGAATCTCGTGCTTTC	209	Transgene confirmation
HygR_R	CAACGTGACACCCTGTGAAC	209	
Cas9_pHSE_F	ATCCAATCTTCGGCAACAT	484	
Cas9_pHSE_R	TTATCCAGGTCATCGTCGTAT	484	
PDSgRNA_F	ATTGCTGTTAACTTGAGAGTCCA	–	sgRNA cloning
PDSgRNA_R	AAACTGGACTCTCAAGTTAACAG	–	
eIF4EgRNA_F	ATTGAGTTGAAGGCCGCCGATGG	–	
eIF4EgRNA_R	AAACCCATCGGCGGCCTTCAACT	–	

about 3.5% gene-editing efficiency of the *PDS* gene with both copies expected to be edited. However, the *PDS*-edited, photobleached shoots failed to develop into rooted shoots (**Figure 2A**). After sampling the photobleached shoots of the *PDS*-edited explants and putative *eIF4E1*-edited plants, we confirmed the integration of the transfer DNA (T-DNA) by genomic DNA PCR of the samples using *Cas9*- and *Hpt*-specific primers (**Supplementary Figure S1A**). We obtained PCR products with the expected amplicon sizes from the photobleached shoots, demonstrating the successful integration of *Cas9* and *Hpt* (**Supplementary Figure S1A**).

We also regenerated 22 putative *eIF4E1*-edited transgenic plants, which we tested for the presence of the transgenes, the *Cas9* and *Hpt* genes by PCR using transgene-specific primers (**Supplementary Figure S1B**). Sixteen of the 22 plants contained both *Cas9* and *Hpt*, with a transformation efficiency of 72%. Furthermore, RT-PCR results confirmed the expression of the *Cas9* gene in all the PCR positive plants (**Supplementary Figure S1C**). No morphological changes were observed in *eIF4E1*-edited plants.

Confirmation of CRISPR/Cas9-Induced Mutations

To confirm that the CRISPR/Cas9 editing had introduced mutations in the *PDS* gene, we performed PCR amplification of the target region from the albino tomato mutants using primers (**Table 1**) that flanked the sgRNA target, and then sequenced the PCR products by Sanger sequencing. Sequence analysis showed two types of sequence variations: one- and two-nucleotide deletions with breakpoints 3 bp upstream of the protospacer-adjacent motif (PAM) sequence (**Figure 2B**). These deletions resulted in frameshift mutations causing early

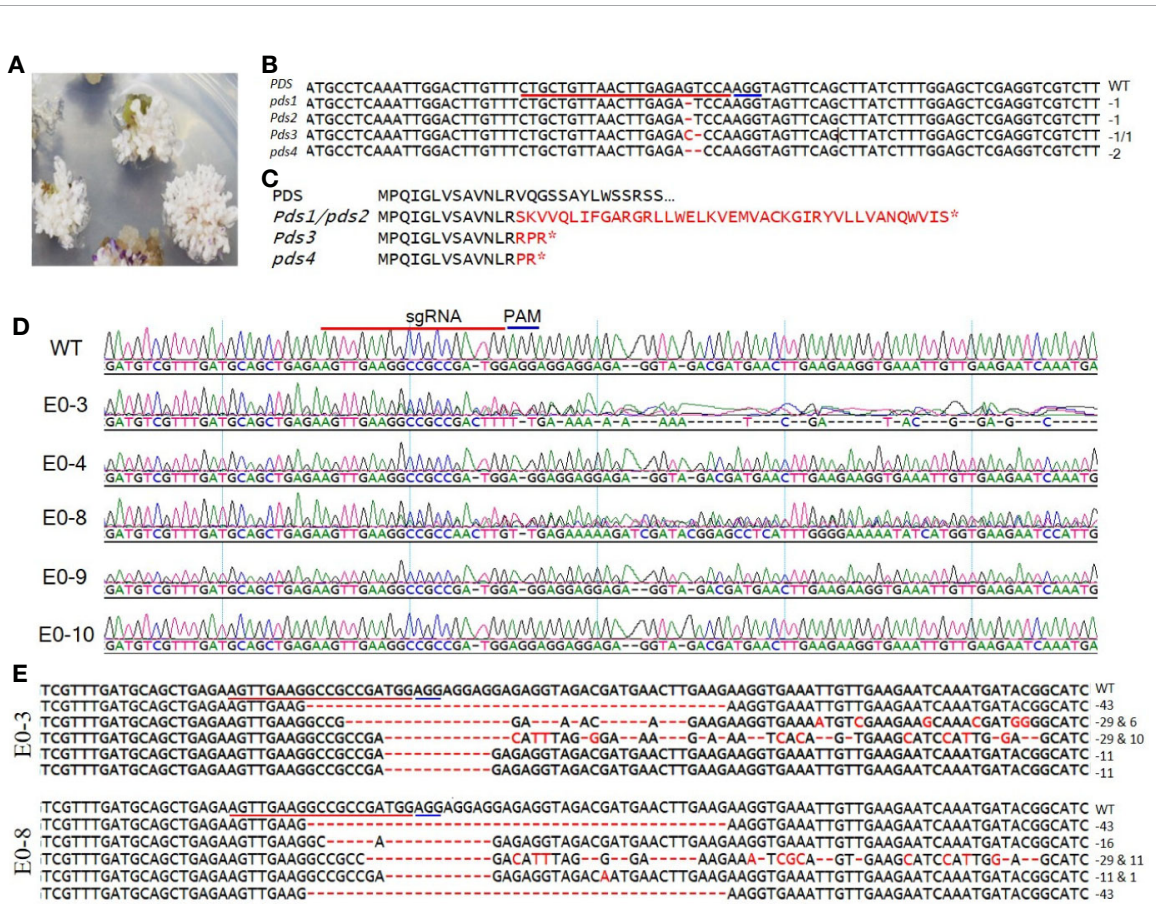


FIGURE 2 | Sequence analysis of the *PDS*- and *eIF4E1*-targeted plants. **(A)** *PDS*-targeted explants showing the photobleached phenotype. **(B)** Mutation detection in *PDS*-targeted shoots by direct sequencing of PCR product in target region. **(C)** Predicted amino acid sequence alignment of *PDS* genes from genome-edited plants. Mutations causing early stop codons or changes in protein sequence are indicated with star symbol. **(D)** Representative chromatograms produced by direct sequencing of PCR products, displayed for sequence alignment of the mutated *eIF4E1* genes from the E0-3, E0-4, E0-8, E0-9, and E0-10 transgenic line. All transgenic lines with exception of the E0-10 showed mixed peaks in the target region due to their varied indels as compared with the wild-type sequence (WT) and are regarded as CRISPR/Cas9-induced mutant lines. **(E)** DNA sequences of mutated *eIF4E1* in the E0-3 and E0-8 line. The number of mutations of each type revealed by random sequencing of TA clones of PCR products. DNA deletions are denoted by red dashes, and deletion sizes (in nucleotides) are marked on the right side of the sequence. The sgRNA target sequence is underlined in red and the PAM motif in blue.

stop codons, preventing the expression of functional PDS protein (**Figure 2C**). Similarly, we identified CRISPR/Cas9-induced mutations in *eIF4E1* by sequencing the PCR amplicons of the sgRNA target-flanking region from the genomic DNA of 16 Cas9-positive tomato lines. Sequencing of the sgRNA target region revealed no mutant homozygous lines among the E_0 plants; 15 of the 16 putative transgenic plants showed mixed sequence peaks at the sgRNA target site compared to the wild-type target sequence (**Figure 2D**), indicating that a mixture of mutant (reflective of CRISPR-Cas9-induced mutation) and wild-type alleles of *eIF4E1* were present. A single plant, E0-10 (**Figure 2D**) showed sequence peaks similar to that of wild-type target sequence, indicating no editing the in target region despite showing expression of the Cas9 gene. To further characterize the mutations in the E_0 plants, we randomly selected the E0-3 and E0-8 transgenic lines for further investigation by TA cloning and sequencing of the target region. The sequencing results for the TA clones from E0-3 and E0-8 revealed four different alleles of *eIF4E1* (**Figure 2E**), with indels and/or substitutions at various positions both in proximity to the sgRNA target and outside the immediate sgRNA target region. The presence of a mixture of different mutant alleles in each E_0 line suggested that active somatic mutation was occurring in the edited plants. Therefore, we advanced selected E_0 plants to the E_1 and E_2 generations to create homozygous mutant lines.

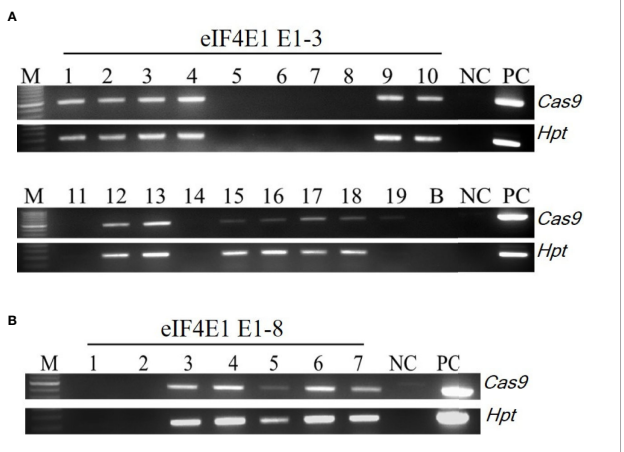


FIGURE 3 | Transgene screening of *eIF4E1*-edited E_1 progeny. **(A, B)** Transgene screening of *eIF4E1*-targeted E_1 progeny of E0-3 (**A**; 19 progeny) and E0-8 (**B**; 7 progeny) by genomic DNA PCR. NC, wild type (negative control); PC, *Cas9* plasmid (positive control), *Hpt*, Hygromycin phosphotransferase marker gene.

Inheritance of CRISPR/Cas9 Induced Mutations

Among the *eIF4E1* mutants, we selected the E0-3 and E0-8 plants from which to generate E_1 lines. First, we allowed the E0-3 and E0-8 plants to self-pollinate to produce E_1 lines. We then extracted gDNA from leaf samples of the E_1 progeny and tested these for the presence of transgene by PCR with transgene-specific primers. Among 19 plants derived from E0-3, 13 carried transgenes (1, 2, 3, 4, 9, 10, 12, 13, 15, 16, 17, 18, and 19) and 6 did not (5, 6, 7, 8, 11, and 14), and among seven plants from the E0-8 plant, five carried transgenes (3, 4, 5, 6, and 7) and two did not (1 and 2) (**Figure 3**), indicative of the transgene-heterozygous status of the E_0 plants.

To reveal sequence variations of E_1 plants, we performed TA cloning and sequencing of amplicons of the target region in E_1 plants derived from lines E0-3 and E0-8. New mutation patterns were evident in the E_1 lines as compared to the E_0 plant, and two lines (E1 3-8 and E1 3-17) derived from E0-3 were homozygous for the expected 43-bp deletion (**Figure 4A**). The E1 3-11 line was biallelic, carrying two deletions of 43 and 11 bp; two other lines (E1 3-9 and 3-15) showed a 29-bp deletion and a tandem insertion of 38 bp of adjacent repeat sequence, respectively (**Figure 4C**); and the E1 3-19 line carried mixed mutations, with deletions of the unexpected sizes of 12, 13, and 15 bp (**Figure 4A**). The three lines (E1 8-3, 8-5, 8-7) derived from the E0-8 plant were homozygous for the expected 11-bp, 43-bp, and 29-bp deletions and 38-bp insertion of a repeat sequence (**Figures 4A, C**). The E1 8-1 line was biallelic, with a 43-bp deletion and a 29-bp deletion and 38-bp repeat-sequence insertion, whereas the E1 8-4 line had mixed mutations along with a wild-type allele (**Figure 4B**). The presence of new mutation patterns in E_1 lines that differed from those in the E_0 plants suggested that active somatic mutation was occurring in the E_0 plants (**Figures 2E** and **4**). Notably, one line, E1 3-8, had a homozygous edited *eIF4E1* gene but no T-DNA. Overall, several homozygous and biallelic lines were recovered in the E_1 generation (**Table 2**), along with a smaller proportion of mosaic mutants carrying different allelic variants (**Table 2**).

Potyvirus Resistance in *eIF4E1* Edited Plants

To investigate whether these CRISPR/Cas9-derived *eIF4E1* mutants conferred resistance to TEV, we inoculated E1 3-8 (homozygous line) and E1 3-11 (biallelic line) plants with TEV (**Figure 5**). TEV symptoms appeared as early as 7 DPI; the wild-type plant showed typical TEV symptoms, including vein clearing and several small chlorotic spots in the leaves. Both *eIF4E1* mutant lines showed similar symptoms (**Figure 5A**). To confirm virus infection, we performed DAS-ELISA analysis using inoculated lower and uninoculated upper leaves. We detected similar high amounts of virus coat protein accumulation in the E1 3-8 and E1 3-11 lines, indicating that the *eIF4E1*-edited lines were susceptible to TEV-HAT (**Figure 5B**). We next assessed whether CRISPR/Cas9 *eIF4E1* mutants could confer resistance to another potyvirus, PepMoV, by challenging three E_1 mutant lines (E1 3-8, E1 3-11, and E1 3-15) with PepMoV. PepMoV coat protein accumulated to a high level in both the inoculated and uninoculated upper leaves of the susceptible wild-type control (**Figure 5**), whereas no coat protein accumulated in any of the three mutant lines in either the inoculated lower or uninoculated systemic leaves, confirming that the *eIF4E1*-edited lines were resistant to PepMoV (**Figure 5C**).

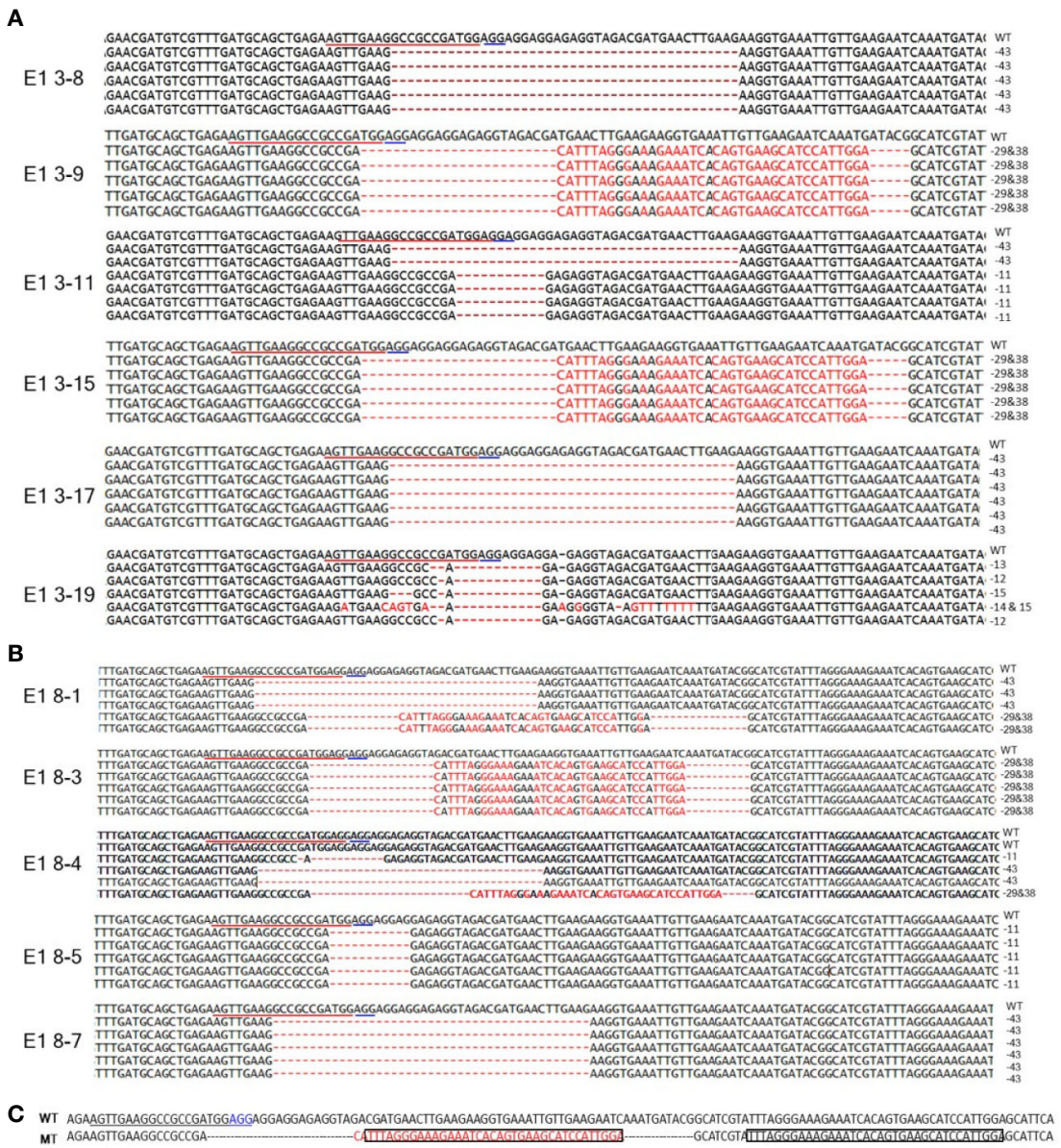


FIGURE 4 | Genome-edited sequences of *eIF4E1* from E_1 transgenic lines. **(A, B)** DNA sequences of the *eIF4E1* mutant lines derived from E0-3 **(A)** and E0-8 **(B)**. **(C)** Insertion of a 38-bp repeat sequence (boxed) from sgRNA adjacent region. The sgRNA target sequence is underlined in red and the PAM motif in blue. WT, wild type; MT, mutant lines (E1 3-9, E1 3-15, E1 8-1, and E1 8-3).

TABLE 2 | Summary of eIF4E1 CRISPR/Cas9-induced mutations in E_1 lines.

Line	Generation	Mutants analyzed	Mutant type	Number of plants	Percentage
E1 3	E_1	19	Homozygous	7	36.8%
			Biallelic	8	42.1%
			Mosaic	4	21.0%
E1 8	E_1	7	Homozygous	4	57.1%
			Biallelic	2	28.5%
			Mosaic	1	14.2%

DISCUSSION

Here we subjected the *eIF4E1* gene to CRISPR/Cas9 gene editing with the aim of creating potyvirus resistance in *S. lycopersicum* Micro-Tom, as recessive resistance genes are considered to be confer more durable resistance than dominant resistance (*R*) genes (Kang et al., 2005a; Borrelli et al., 2018). In this study, we developed CRISPR/Cas9-edited tomato *PDS* and *eIF4E1* mutants through *Agrobacterium*-mediated delivery of CRISPR/Cas9 and

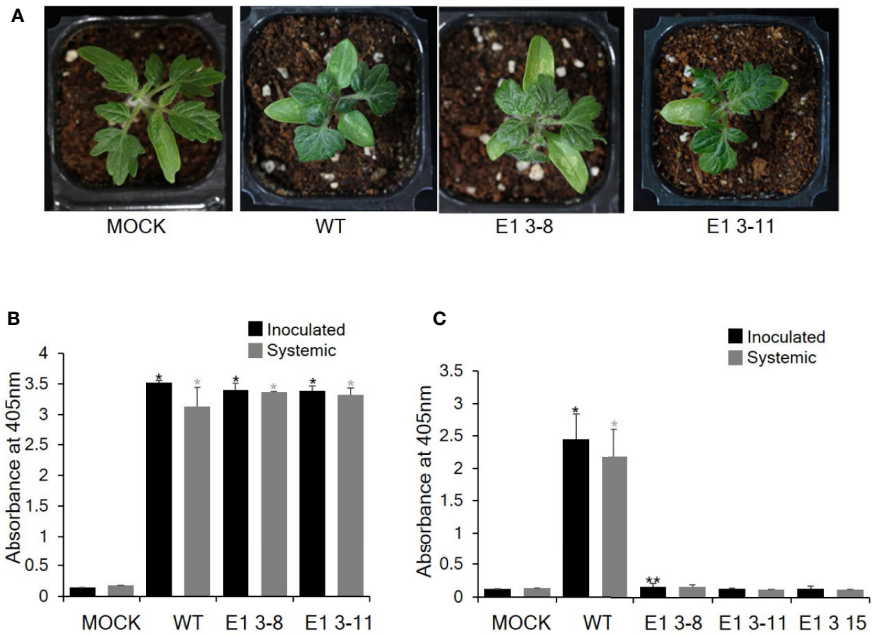


FIGURE 5 | Resistance analysis of *eIF4E1* mutant lines challenged with potyviruses. **(A)** Phenotypes of *eIF4E1* mutant lines infected with TEV at 7 DPI. **(B, C)** Analysis of the accumulation of TEV-HAT **(B)** and PepMoV **(C)** in inoculated and systemic leaves by DAS-ELISA. Error bars indicate mean values of replicates ± SD. Asterisks indicate significant differences (*P < 0.01; **P < 0.05) between 'mock' and 'treatments' according to Student's t-test.

sgRNA reagents. Sequencing of the *PDS* gene target region revealed that DNA double-strand edits occurred 3 bp upstream of the PAM site, resulting in indels of one or two nucleotides in *PDS*, as reported previously (Garneau et al., 2010; Jinek et al., 2012). All the indel mutations led to early stop codons, causing loss of *PDS* gene function. In the *eIF4E*-edited lines, indels occurred at various positions near the sgRNA target and in some cases extending beyond the sgRNA target and PAM region, creating premature stop codons and truncated eIF4E1 proteins (**Supplementary Data File 1**). Notably, we observed a 38-bp insertion of a repeat sequence as well as a 29-bp deletion in several of the gene-edited lines, E1 3-9, E1 3-15, E1 8-1, and E1 8-3; the mechanism leading to the insertion of this adjacent repeat sequence will require further investigation.

Unlike *PDS*-targeted photobleached plants, the *eIF4E1*-targeted mutant plants in the E_0 generation were chimeric, with mixtures of different allelic variations, and the E_1 generation comprised a mixture of homozygous, biallelic, and mosaic mutants. Increasing evidence suggests that transgenic plants can be chimeric in the E_0 generation. For example, chimeric lines have also been reported in T_0 (E_0) genome-edited Arabidopsis, tomato, rice, and soybean (Brooks et al., 2014; Feng et al., 2014; Ma et al., 2015; Pan et al., 2016; Pyott et al., 2016; Liu et al., 2020). One study identified only three non-mosaic mutants among nearly 300 lines of Arabidopsis created by editing of the gene *RRP42* (Yan et al., 2017). The mosaicism of the E_1 mutants could then be due to the fact that in some edited lines, one or more wild-type alleles might escape detection in the E_0 generation but then be detected in later generations (Brooks et al., 2014; Zhang et al., 2014; Zhang et al., 2019; Liu et al., 2020). Furthermore, other new alleles created from late-arising chimeric tissues might also be overlooked in E_0 plants, resulting in different flowers carrying different alleles (Brooks et al., 2014). In agreement with this contention, in the present study, some of the E_1 plants produced by gene editing of *eIF4E1* were chimeric, and furthermore mutations different from those identified in E_0 plants were observed in E_1 plants.

Notwithstanding these complications, we advanced the E_0 plants carrying diverse alleles of the target gene to the next generation, and selected E_1 lines carrying homozygous alleles from segregating populations, as reported in previous studies (Xu et al., 2015; Chandrasekaran et al., 2016; Pyott et al., 2016). Accordingly, sequence analysis of the E_1 lines identified both homozygous and biallelic *eIF4E1* mutants, both of which would be expected to stably pass on their mutant status to their descendants (Chandrasekaran et al., 2016), consistent with this, homozygous mutant alleles were stably inherited in the E_2 progeny.

Previous studies in pepper suggest that eIF4E is a key factor in the interaction between viruses and their hosts. Mutations in *eIF4E* cause conformational shifts in the encoded proteins, interrupting the interaction between VPg and eIF4E and conferring plant resistance to the virus at the cellular level (Kang et al., 2005a). This implied that site-directed mutagenesis could be used to create dominant negative mutations or targeted silencing of host genes (Azevedo et al., 2002; Kang et al., 2005b; Kang et al., 2007). We therefore assessed our eIF4E-edited E_1 lines, with or without the *Cas9* transgene, for resistance to potyviruses. Edited plants carrying mutations in *eIF4E1* showed resistant to PepMoV.

Consistent with this, we did not detect accumulation of coat protein in either inoculated or uninoculated leaves. However, E_1 plants carrying mutations in *eIF4E1* showed typical TEV symptoms, and ELISA analysis confirmed the presence of systemic infection. Viral coat protein accumulated to high levels in both inoculated and uninoculated upper leaves of the homozygous E1 3-8 and biallelic E1 3-11 lines. These results contrast with the findings of an earlier study indicating that transgenic tomato plants overexpressing a recessive resistance allele of *eIF4E (pvr1)* from *Capsicum chinense* showed dominant resistance to TEV (Kang et al., 2007). The transgenic expression of the *Capsicum eIF4E* mutant allele was suggested to perturb the interactions required for viral susceptibility in a heterologous host system (Kang et al., 2007). However, the susceptibility of the edited plants to TEV in tomato that we observed here could be due to the redundant activities of *eIF4E* homologs (*eIF4E1* and *eIF4E2*) coupled with the absence of a dominant negative allele whose protein product could interfere with the endogenous wild-type proteins and inhibit viral infection (Kang et al., 2007). Furthermore, an *eIF4E1*-knockout tomato plant selected from a TILLING population was reported to be susceptible to potyviruses because TEV could utilize either *eIF4E1* or *eIF4E2* for its replication in tomato (Gauffier et al., 2016). Thus, development of a TEV virus-resistant tomato line is hindered by the gene redundancy of *eIF4E* homologs. *eIF4E1* and *eIF4E2* double-knockout plants would be expected to show a broad-spectrum resistance to a wide range of potyviruses (Gauffier et al., 2016); however, care should be taken to ensure that mutations in both candidate genes do not impair the plant growth and development, as often occurs with double mutants (Mazier et al., 2011; Gauffier et al., 2016).

CONCLUSION

We used CRISPR/Cas9 gene editing to induce mutations in the tomato *eIF4E1* gene by *A. tumefaciens*-mediated transformation. Evaluation of the mutation and inheritance patterns of this gene in the E_0 and later generations by sequencing revealed high proportions of chimeric mutant lines in the E_0 generation and of homozygous and biallelic mutants of *eIF4E1* in the E_1 generation. The CRISPR/Cas9-mediated gene mutations were stably transmitted to the E_1 and E_2 descendants irrespective of the presence or absence of T-DNA. Although the presence of redundant *eIF4E* homologs apparently precluded TEV resistance in the *eIF4E1* genome-edited Micro-Tom plants, an edited *eIF4E1* with an early stop codon conferred resistance to PepMoV. Thus, our system enabled successful CRISPR/Cas9 editing of tomato *eIF4E1*, resulting in resistance to a common viral pathogen of Solanaceae; further research is planned to obtain homozygous E_0 plants and to extend this work to pepper (*Capsicum* species) to produce strains with broad-spectrum virus resistance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Y-JY and JV designed the constructs and performed the transformation. Y-JY, D-SK, JK, and H-EL characterized the transgenics and identified the mutants. Y-JY, J-HL and JV performed viral inoculations and ELISA. JV wrote the manuscript. B-CK, D-SK, JK, H-EL, and JV revised the manuscript. B-CK conceived and designed the research. All authors contributed to the article and approved the submitted version.

FUNDING

This work was carried out with the support of the Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01327801, PJ01327803), Rural Development Administration, Republic of Korea.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01098/full#supplementary-material

REFERENCES

Azevedo, C., Sadanandom, A., Kitagawa, K., Freialdenhoven, A., Shirasu, K., and Schulze-Lefert, P. (2002). The RAR1 interactor SGT1, an essential component of R gene-triggered disease resistance. *Science* 295, 2073–2076. doi: 10.1126/science.1067554

Borrelli, V. M., Brambilla, V., Rogowsky, P., Marocco, A., and Lanubile, A. (2018). The enhancement of plant disease resistance using CRISPR/Cas9 technology. *Front. Plant Sci.* 9:1245.S:1245. doi: 10.3389/fpls.2018.01245.S

Brooks, C., Nekrasov, V., Lippman, Z. B., and Van Eck, J. (2014). Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol.* 166, 1292–1297. doi: 10.1104/pp.114.247577

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi: 10.1111/mpp.12375

Charron, C., Nicolai, M., Gallois, J. L., Robaglia, C., Moury, B., Palloix, A., et al. (2008). Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. *Plant J.* 54, 56–68. doi: 10.1111/j.1365-313X.2008.03407.x

Cong, L., and Zhang, F. (2015). Genome engineering using CRISPR-Cas9 system. *Methods Mol. Biol.* 1239, 197–217. doi: 10.1007/978-1-4939-1862-1_10

Cui, H., and Wang, A. (2019). The biological impact of the hypervariable N-terminal region of potyviral genomes. *Annu. Rev. Virol.* 6, 255–274. doi: 10.1146/annurev-virology-092818-015843

de Oliveira, L. C., Volpon, L., Rahardjo, A. K., Osborne, M. J., Culjkovic-Kraljacic, B., Trahan, C., et al. (2019). Structural studies of the eIF4E–VPg complex reveal a direct competition for capped RNA: Implications for translation. *Proc. Natl. Acad. Sci. U. S. A.* 116, 24056–24065. doi: 10.1073/pnas.1904752116

Duprat, A., Caranta, C., Revers, F., Menand, B., Browning, K. S., and Robaglia, C. (2002). The Arabidopsis eukaryotic initiation factor (iso) 4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant J.* 32, 927–934. doi: 10.1046/j.1365-313x.2002.01481.x

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D. L., et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4632–4637. doi: 10.1073/pnas.1400822111

Garneau, J. E., Dupuis, M.È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67. doi: 10.1038/nature09523

Gauffier, C., Lebaron, C., Moretti, A., Constant, C., Moquet, F., Bonnet, G., et al. (2016). A TILLING approach to generate broad-spectrum resistance to potyviruses in tomato is hampered by *eIF4E* gene redundancy. *Plant J.* 85, 717–729. doi: 10.1111/tpj.13136

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., et al. (2019). Simultaneous CRISPR/Cas9-mediated editing of cassava eIF 4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol. J.* 17, 421–434. doi: 10.1111/pbi.12987

Hančinský, R., Mihálik, D., Mrkvová, M., Candresse, T., and Glasa, M. (2020). Plant viruses infecting Solanaceae family members in the cultivated and wild environments: A Review. *Plants* 9:667. doi: 10.3390/plants9050667

Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi: 10.1016/j.cell.2014.05.010

Hull, R. (2009). Mechanical inoculation of plant viruses. *Curr. Protoc. Microbiol.* 13, 16B–166. doi: 10.1002/9780471729259.mc16b06s13

Hwang, J., Li, J., Liu, W. Y., An, S. J., Cho, H., Her, N. H., et al. (2009). Double mutations in eIF4E and eIFiso4E confer recessive resistance to Chilli veinal mottle virus in pepper. *Mol. Cells* 27, 329–336. doi: 10.1007/s10059-009-0042-y

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 41, e188–e188. doi: 10.1093/nar/gkt780

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Kang, B. C., Yeam, I., and Jahn, M. M. (2005a). Genetics of plant virus resistance. *Annu. Rev. Phytopathol.* 43, 581–621. doi: 10.1146/annurev.phyto.43.011205.141140

Kang, B. C., Yeam, I., Frantz, J. D., Murphy, J. F., and Jahn, M. M. (2005b). The *pvr1* locus in Capsicum encodes a translation initiation factor eIF4E that interacts with *Tobacco etch virusTobacco etch virus* VPg. *Plant J.* 42, 392–405. doi: 10.1111/j.1365-313X.2005.02381.x

Kang, B. C., Yeam, I., Li, H., Perez, K. W., and Jahn, M. M. (2007). Ectopic expression of a recessive resistance gene generates dominant potyvirus resistance in plants. *Plant Biotechnol. J.* 5, 526–536. doi: 10.1111/j.1467-7652.2007.00262.x

Kim, J., Kang, W. H., Yang, H. B., Park, S., Jang, C. S., Yu, H. J., et al. (2013). Identification of a broad-spectrum recessive gene in Brassica rapa and molecular analysis of the *eIF4E* gene family to develop molecular markers. *Mol. Breed.* 32, 385–398. doi: 10.1111/mpp.12120

Kim, J., Kang, W. H., Hwang, J., Yang, H. B., Dosun, K., Oh, C. S., et al. (2014). Transgenic *Brassica rapa* plants over-expressing *eIF(iso)4E* variants show broad-spectrum Turnip mosaic virus (TuMV) resistance. *Mol. Plant Pathol.* 15, 615–626. doi: 10.1111/mpp.12120

Kothari, S. L., Joshi, A., Kachhwaha, S., and OchoaAlejo, N. (2010). Chilli peppers-a review on tissue culture and transgenesis. *Biotechnol. Adv.* 28, 35–48. doi: 10.1016/j.biotechadv.2009.08.005

Le Gall, O., Aranda, M. A., and Caranta, C. (2011). "Plant resistance to viruses mediated by translation initiation factors," in *The Recent Advances in Plant Virology*. Eds. C. Caranta, M. A. Aranda, M. Tepfer and J. López-Moya (Wymondham, Norfolk: Caister Academic Press), 177–194.

Lebaron, C., Rosado, A., Sauvage, C., Gauffier, C., German-Retana, S., Moury, B., et al. (2016). A new *eIF4E1* allele characterized by RNAseq data mining is associated with resistance to PVY in tomato albeit with a low durability. *J. Gen. Virol.* 97, 3063–3072. doi: 10.1099/jgv.0.000609

Lee, H. R., An, H. J., You, Y. G., Lee, J., Kim, H. J., Kang, B. C., et al. (2013). Development of a novel codominant molecular marker for *Chili veinal mottle virus* resistance in *Capsicum annuum* L. *Euphytica* 193, 197–205. doi: 10.1007/s10681-013-0897-z

Lellis, A. D., Kasschau, K. D., Whitham, S. A., and Carrington, J. C. (2002). Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF (iso) 4E during potyvirus infection. *Curr. Biol.* 12, 1046–1051. doi: 10.1016/s0960-9822(02)00898-9

Liu, H., Wang, K., Jia, Z., Gong, Q., Lin, Z., Du, L., et al. (2020). Efficient induction of haploid plants in wheat by editing of *TaMTL* using an optimized Agrobacterium-mediated CRISPR system. *J. Exp. Bot.* 71, 1337–1349. doi: 10.1093/jxb/erz529

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* 8, 1274–1284. doi: 10.1016/j.molp.2015.04.007

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Cermak, T., et al. (2018). Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol. J.* 16, 1918–1927. doi: 10.1111/pbi.12927

Mazier, M., Flamain, F., Nicolaï, M., Sarnette, V., and Caranta, C. (2011).). Knock-down of both *eIF4E1* and *eIF4E2* genes confers broad-spectrum resistance against potyviruses in tomato. *PLoS One* 6, e29595. doi: 10.1371/journal.pone.0029595

Melzer, M. J., Sugano, J. S., Cabanas, D., Dey, K. K., Kandouh, B., Mauro, D., et al. (2012). First report of Pepper mottle virus infecting tomato in Hawaii. *Plant Dis.* 96, 917. doi: 10.1094/PDIS-02-12-0147-PDN

Miyoshi, H., Suehiro, N., Tomoo, K., Muto, S., Takahashi, T., Tsukamoto, T., et al. (2006). Binding analyses for the interaction between plant virus genome-linked protein (VPg) and plant translational initiation factors. *Biochimie* 88, 329–340. doi: 10.1016/j.biochi.2005.09.002

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassay with tobacco tissue cultures. *Physiol. Plant* 5, 473–479. . doi: 10.1111/j.1399-3054.1962.tb08052.x

Nicaise, V. (2014). Crop immunity against viruses: outcomes and future challenges. *Front. Plant Sci.* 5:660:660. doi: 10.3389/fpls.2014.00660

Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., et al. (2016). CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. *Sci. Rep.* 6:24765. doi: 10.1038/srep24765

Peng, A., Chen, S., Lei, T., Xu, L., He, Y., Wu, L., et al. (2017). Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene Cs LOB 1 promoter in citrus. *Plant Biotechnol. J.* 15, 1509–1519. doi: 10.1111/pbi.12733

Porebski, S., Bailey, L. G., and Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15, 8–15. doi: 10.1007/BF02772108

Pyott, D. E., Sheehan, E., and Molnar, A. (2016). Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free *Arabidopsis* plants. *Mol. Plant Pathol.* 17, 1276–1288. doi: 10.1111/mpp.12417

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

Robaglia, C., and Caranta, C. (2006). Translation initiation factors: a weak link in plant RNA virus infection. *Trends Plant Sci.* 11, 40–45. doi: 10.1016/j.tplants.2005.11.004

Rodríguez-Hernández, A. M., Gosalvez, B., Sempere, R. N., Burgos, L., Aranda, M. A., and Truniger, V. (2012). Melon RNA interference (RNAi) lines silenced for *Cm-eIF4E* show broad virus resistance. *Mol. Plant Pathol.* 13, 755–763. doi: 10.1111/j.1364-3703.2012.00785.x

Ruffel, S., Gallois, J. L., Lesage, M. L., and Caranta, C. (2005). The recessive potyvirus resistance gene *pot-1* is the tomato orthologue of the pepper *pvr2-eIF4E* gene. *Mol. Genet. Genomics* 274, 346–353. doi: 10.1007/s00438-005-0003-x

Ruffel, S., Gallois, J. L., Moury, B., Robaglia, C., Palloix, A., and Caranta, C. (2006). Simultaneous mutations in translation initiation factors *eIF4E* and *eIF(iso)4E* are required to prevent pepper veinal mottle virus infection of pepper. *J. Gen. Virol.* 87, 2089–2098. doi: 10.1099/vir.0.81817-0

Sanfacon, H. (2015). Plant translation factors and virus resistance. *Viruses* 7, 3392–3419. doi: 10.3390/v7072778

Sato, M., Nakahara, K., Yoshii, M., Ishikawa, M., and Uyeda, I. (2005). Selective involvement of members of the eukaryotic initiation factor 4E family in the

infection of *Arabidopsis thaliana* by potyviruses. *FEBS Lett.* 579, 1167–1171. doi: 10.1016/j.febslet.2004.12.086

Shapiro, R. S., Chavez, A., and Collins, J. J. (2018). CRISPR-based genomic tools for the manipulation of genetically intractable microorganisms. *Nat. Rev. Microbiol.* 16, 333–339. doi: 10.1038/s41579-018-0002-7

Tavert-Roudet, G., Anne, A., Barra, A., Chovin, A., Demaille, C., and Michon, T. (2017). The Potyvirus particle recruits the plant translation initiation factor eIF4E by means of the VPg covalently linked to the viral RNA. *Mol. Plant Microbe Interact.* 30, 754–762. doi: 10.1094/MPMI-04-17-0091-R

Wang, A., and Krishnaswamy, S. (2012). Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *Mol. Plant Pathol.* 13, 795–803. doi: 10.1111/j.1364-3703.2012.00791.x

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., et al. (2016). Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene *OsERF922*. *PLoS One* 11, e0154027. doi: 10.1371/journal.pone.0154027

Wang, A. (2015). Dissecting the molecular network of virus-plant interactions: the complex roles of host factors. *Annu. Rev. Phytopathol.* 253, 45–66. doi: 10.1146/annurev-phyto-080614-120001

Ward, C. W., and Shukla, D. D. (1991). Taxonomy of potyviruses: current problems and some solutions. *Intervirology* 32, 269–296. doi: 10.1159/000150211

Wintermantel, W. M. (2011). A comparison of disinfectants to prevent spread of potyviruses in greenhouse tomato production. *Plant Health Prog.* 12, 19. doi: 10.1094/PHP-2011-0221-01-RS

Wu, Y., Liang, D., Wang, Y., Bai, M., Tan, W., Bao, S., et al. (2013). Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* 13, 659–662. doi: 10.1016/j.stem.2013.10.016

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, C. Y., Liu, B., et al. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* 14:327. doi: 10.1186/s12870-014-0327-y

Xu, R. F., Li, H., Qin, R. Y., Li, J., Qiu, C. H., Yang, Y. C., et al. (2015). Generation of inheritable and "transgene clean" targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Sci. Rep.* 5:11491. doi: 10.1038/srep11491

Yan, X., Yan, Z., and Han, Y. (2017). RRP42, a subunit of exosome, plays an important role in female gametophytes development and mesophyll cell morphogenesis in *Arabidopsis*. Front. *Plant Sci.* 8:981:981. doi: 10.3389/fpls.2017.00981

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., et al. (2014). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* 12, 797–807. doi: 10.1111/pbi.12200

Zhang, Z., Zhang, Y., Gao, F., Han, S., Cheah, K. S., Tse, H. F., et al. (2017). CRISPR/Cas9 genome-editing system in human stem cells: current status and future prospects. *Mol. Ther. Nucleic Acids* 9, 230–241. doi: 10.1016/j.omtn.2017.09.009

Zhang, S., Zhang, R., Gao, J., Gu, T., Song, G., Li, W., et al. (2019). Highly efficient and heritable targeted mutagenesis in wheat via the *Agrobacterium tumefaciens*-mediated CRISPR/Cas9 system. *Int. J. Mol. Sci.* 20, 4257. doi: 10.3390/ijms20174257

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, 1589–1589. 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.

Copyright © 2020 Yoon, Venkatesh, Lee, Kim, Lee, Kim and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 August 2020
doi: 10.3389/fmicb.2020.01903

Novel Fig-Associated Viroid-Like RNAs Containing Hammerhead Ribozymes in Both Polarity Strands Identified by High-Throughput Sequencing

Alejandro Olmedo-Velarde[1†], Beatriz Navarro[2†], John S. Hu[1], Michael J. Melzer[1] and Francesco Di Serio[2*]*

[1] Plant and Environmental Protection Sciences, University of Hawai'i at Mānoa, Honolulu, HI, United States, [2] Istituto per la Protezione Sostenibile delle Piante, Consiglio Nazionale delle Ricerche, Bari, Italy

Based on high-throughput sequencing (HTS) data, the existence of viroid-like RNAs (Vd-LRNAs) associated with fig trees grown in the Hawaiian Islands has been predicted. One of these RNAs has been characterized as a circular RNA ranging in size from 357 to 360 nucleotides. Structural and biochemical features of this RNA, tentatively named fig hammerhead viroid-like RNA (FHVd-LR), markedly resemble those previously reported for several viroids and viroid-like satellite RNAs (Vd-LsatRNAs), which are non-protein-coding RNAs infecting their hosts autonomously and in combination with a helper virus, respectively. The full-length sequence of FHVd-LR variants was determined by RT-PCR, cloning, and sequencing. Despite a low global sequence identity with known viroids and Vd-LsatRNAs, FHVd-LR contains a hammerhead ribozyme (HRz) in each polarity strand. Northern blot hybridization assays identified the circular and linear forms of both polarity strands of FHVd-LR and showed that one strand, assigned the (+) polarity, accumulates at higher levels than the (–) polarity strand *in vivo.* The (+) polarity RNA assumes a rod-like secondary structure of minimal free energy with the conserved domains of the HRzs located in opposition to each other, a feature typical of several viroids and Vd-LRNAs. The HRzs of both FHVd-LR polarity strands were shown to be active *in vitro* during transcription, self-cleaving the RNAs at the predicted sites. These data, together with the sequence variability observed in the cloned and sequenced full-length variants, indicate that FHVd-LR is a novel viroid or Vd-LsatRNA. According to HTS data, the coexistence of FHVd-LR of different sizes in the same host cannot be excluded. The relationships of FHVd-LR with previously reported viroids and Vd-LsatRNAs, and the need to perform bioassays to conclusively clarify the biological nature of this circular RNA, are discussed.

Keywords: circular RNA, infectious RNAs, *Ficus carica*, next generation sequencing, non-coding RNAs, ribozyme, viroid, virusoid

OPEN ACCESS

Edited by:
John Wesley Randles, The University of Adelaide, Australia

Reviewed by:
Peter Palukaitis, Seoul Women's University, South Korea
Zhixiang Zhang, State Key Laboratory for Biology of Plant Diseases and Insect Pests (CAAS), China

***Correspondence:**
Michael J. Melzer
melzer@hawaii.edu
Francesco Di Serio
francesco.diserio@ipsp.cnr.it

† 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: *11 June 2020*
Accepted: *20 July 2020*
Published: *18 August 2020*

Citation:
Olmedo-Velarde A, Navarro B, Hu JS, Melzer MJ and Di Serio F (2020) Novel Fig-Associated Viroid-Like RNAs Containing Hammerhead Ribozymes in Both Polarity Strands Identified by High-Throughput Sequencing. Front. Microbiol. 11:1903. doi: 10.3389/fmicb.2020.01903

INTRODUCTION

Viroids are small, infectious, non-protein-coding circular RNAs, so far identified only in plants (Kovalskaya and Hammond, 2014; Flores et al., 2015; Gago-Zachert, 2016). Other RNAs resembling viroids from a structural point of view, but differing from them at the biological level, or not conclusively characterized in this respect, have also been reported from plants and designated as viroid-like RNAs (Vd-LRNAs). Viroid-like satellite RNAs (Vd-LsatRNAs) and retroviroid-like elements are Vd-LRNAs markedly diverging from viroids at the biological level. While viroids replicate and systemically infect their host plants autonomously (in the absence of any helper virus), the infectivity of Vd-LsatRNAs depends on a coinfecting helper virus (Navarro et al., 2017), Vd-LsatRNAs are encapsidated by the helper virus capsid proteins and have been also named virusoids (Symons and Randles, 1999). Both viroids and Vd-LsatRNAs differ from carnation small Vd-LRNA, which is non-infectious and has a DNA counterpart integrated in the genome of a plant pararetrovirus (Daròs and Flores, 1995; Vera et al., 2000) or in the plant genome (Hegedus et al., 2004). For this reason, carnation small Vd-LRNA is considered a retroviroid-like element.

Viroids replicating in the nucleus and the chloroplast have been classified into the families *Pospiviroidae* and *Avsunviroidae*, respectively (Di Serio et al., 2014). Members of the two families also differ in structural and other functional features. In viroids of the family *Pospiviroidae*, one RNA polarity strand generates circular forms that assume a rod-like or quasi-rod-like secondary structure of minimal free energy containing a central conserved region (CCR) and other conserved motifs. The CCR is involved in the replication of nuclear viroids through an asymmetric rolling-circle mechanism (Flores et al., 2014). Viroids of the family *Avsunviroidae* lack the CCR and other typical structural motifs conserved in nuclear viroids. Instead, they assume rod-like, quasi-rod-like, or branched conformation and contain hammerhead ribozymes (HRzs) (Di Serio et al., 2018b). HRzs are inactive in the most stable viroid RNA conformation, but they assume an active structure responsible for RNA self-cleavage (without the catalytic contribution of any protein) during replication (Hutchins et al., 1986). In contrast to nuclear viroids, members of the family *Avsunviroidae* replicate through a symmetric rolling-circle mechanism, in which self-cleaved oligomeric RNAs of both polarity strands are circularized (Flores et al., 2014). HRzs with similar functional roles during replication also have been identified in several Vd-LsatRNAs (Navarro et al., 2017). However, in most of them the HRz is present in only one polarity, with the other polarity containing a different ribozyme (named paperclip) or no ribozyme at all. In accord with the presence of ribozymes in one or both polarity strands, asymmetric or symmetric rolling-circle replication mechanisms have been proposed for Vd-LsatRNAs, respectively (Navarro et al., 2017). Among Vd-LsatRNAs, only those of lucerne transient streak virus (LTSV, genus *Sobemovirus*), cereal yellow dwarf virus-RPV (genus *Polerovirus*), and two cherry small circular viroid-like RNAs (cscRNAs, likely associated with a mycovirus) contain two HRzs in both polarity strands (Di Serio et al., 1997, 2006; Navarro et al., 2017).

In tissues infected by viroids or Vd-LsatRNAs, the two polarity strands of the infectious RNAs are detectable, but generally one strand accumulates at a higher level and, by convention, is designated the (+) strand. Akin to viruses, viroids and Vd-LsatRNAs have the typical features of quasispecies (Codoñer et al., 2006; Flores et al., 2014), which is consistent with the observation that these infectious agents accumulate in their hosts as populations of closely related sequence variants, among which one (the master sequence) or a few may prevail depending on the selection pressures imposed by the host and environment (Di Serio et al., 2017).

Edible fig (*Ficus carica* L.) trees are known to be natural hosts of several viruses (Elbeaino et al., 2010, 2012; Laney et al., 2012; Ale-Agha and Rakhshandehroo, 2014; Minafra et al., 2017). Viroids of the family *Pospiviroidae*, including hop stunt viroid, citrus exocortis viroid, and a viroid resembling apple dimple fruit viroid, have also been reported in fig trees (Yakoubi et al., 2007; Chiumenti et al., 2014). The latter viroid was initially identified in fig by high-throughput sequencing (HTS), a powerful technology that has been applied to investigate the virome of several plant species (Hadidi et al., 2016; Villamor et al., 2019). Starting with HTS data, we report the identification of Vd-LRNAs containing HRzs in fig. One of these RNAs has been molecularly characterized, and its relationships with previously reported viroids and Vd-LsatRNAs are discussed.

MATERIALS AND METHODS

Plant Material, RNA Isolation, and HTS

A fig plant from a commercial nursery on the island of Kauai, Hawaii, displaying symptoms of severe mosaic and leaf distortion, was analyzed by HTS. Nucleic acid preparations enriched in double-stranded RNAs (dsRNAs) were obtained from root tissue as described in Navarro and Di Serio (2018) and used as a source for generating a random-amplified cDNA library (Melzer et al., 2010). Sequencing of the cDNA library was performed using a 454 GS FLX Titanium platform (Roche, Branford, CT, United States) at the University of Hawaii's Advanced Studies in Genomics, Proteomics and Bioinformatics (ASGPB) laboratory.

In 2018 and 2019, samples composed of mixed leaves, petioles, and green bark from twelve fig trees displaying symptoms resembling those of the Kauai fig sample were collected from three locations on the island of Oahu (Hawaii). Total nucleic acids were extracted as described by Li et al. (2008) and tested by RT-PCR (see below). Nucleic acids enriched in highly structured RNAs were obtained from young leaf, bark, and petiole tissues of the fig trees with buffer-saturated phenol and partitioning the nucleic acids by chromatography on non-ionic cellulose CF-11 (Whatman, Maidstone, United Kigdom) as described previously (Pallás et al., 1987).

Bioinformatic Analysis

Analyses of HTS data were performed as described earlier (Olmedo-Velarde et al., 2019). Briefly, after trimming, quality

control, and *de novo* assembly using Trinity (Grabherr et al., 2011), Velvet (Zerbino and Birney, 2008), and Unicycler (Wick et al., 2017), all the assembled contigs were sorted by length. Duplicate contigs and those ≤100 nucleotides (nt) were discarded using Geneious v.10.1.3[1]. The remaining contigs were screened for viral and viroid sequence homology using BlastX and BlastN, respectively[2]. Alignments of HTS reads with a sequence reference and reassembling of selected reads were performed using Bowtie (Langmead et al., 2009) and Phrap (Machado et al., 2011), respectively, and then implemented by the MacVector Assembler platform (17.5.3, MacVector, Inc., Apex, NC, United States). Multiple alignments of nucleotide sequences were performed using Clustal Omega (Sievers et al., 2011). RNAfold software (Lorenz et al., 2011) was used to predict the secondary structure of minimal free energy of the RNAs.

RT-PCR and Cloning

Total nucleic acids were reverse transcribed using random hexamers and Superscript IV reverse transcriptase (Invitrogen, Thermo Fisher Scientific, Waltham, MA, United States), following the manufacturer's instructions. Two microliters of the cDNA reaction served as template for PCR amplification using 0.4 units of Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA, United States) in a 20-μl mixture containing 1 × reaction buffer, 0.25 mM of dNTPs, and 0.5 μM of two FHVd-LR-specific primers (**Table 1**) and using the following cycling conditions: initial denaturation at 98°C for 30 s, followed by 33 cycles at 98°C for 15 s, 59°C for 15 s, 72°C for 15 s, and a final extension step at 72°C for 7 min. After agarose gel purification, an adenine-residue overhang was added at the 5′ end of the amplicons using GoTaq DNA polymerase (Promega, Madison, WI, United States). Amplification products of the expected size (monomeric and dimeric) were then cloned into a pGEMT-Easy vector (Promega, Madison, WI, United States) and sequenced by Sanger Sequencing Custom Service (Macrogen, Amsterdam, Netherlands).

Northern Blot Hybridization

Nucleic acid preparations enriched in highly structured RNAs were subjected to double polyacrylamide gel electrophoresis (PAGE) (Flores et al., 1985). Briefly, the nucleic acids were separated by two consecutive 5% PAGE, the first under non-denaturing conditions (TAE buffer: 40 mM Tris, 20 mM sodium acetate, 1 mM EDTA, pH 7.2) and the second under denaturing conditions (8 M urea and 0.25 × TBE buffer). After staining the second gel with ethidium bromide, the nucleic acids were electroblotted to a nylon membrane (Hybond-N, Amersham, Little Chalfont, United Kingdom) in 0.5 × TBE buffer and immobilized by UV cross-linking. The membranes were then hybridized with DIG-labeled riboprobes complementary to (+) or (−) polarity strands of FHVd-LR as described previously (Hajizadeh et al., 2012). The hybridization signals were revealed with an anti-DIG alkaline phosphatase conjugate and the chemiluminescence substrate CSPD (Roche Applied Science) and visualized with a ChemiDoc Touch Imaging system (Bio-Rad, Hercules, CA, United States). The riboprobes were generated by *in vitro* transcription from linearized plasmids containing the full-length cDNA of FHVd-VL (see below) using a commercial Dig-labeling kit (Roche Diagnostics GmbH, Germany).

TABLE 1 | Primers used in this study.

Name	Strand	Primer sequence (5′ to 3′)	Position
FHVd-1F	(+)	CTCTGCCTGGAACGCTATGC	79–98
FHVd-2R	(−)	GAGCCGAAGAGGTGAGAGTC	79–60
FHVd-3F	(+)	GGAAAACACATTCCTAGACTTC	228–249
FHVd-4R	(−)	TACTGATGAGTCCAAAAGGACG	227–206
FHVd-5F	(+)	CTGATGAGAACAAAAGTTCGAAAC	19–42
FHVd-6R	(−)	TTGGATCACACAATCCAATACCTT	353–18
RACE 5′		CGCGGATCCCCCCCCCCCC	

Analyses of RNA Self-Cleavage

Monomeric and dimeric transcripts of both polarity strands were obtained by *in vitro* transcription of plasmids containing monomeric and head-to-tail dimeric FHVd-LR cDNA sequences in both orientations. Recombinant plasmids were linearized by digestion with the appropriate restriction enzyme (*Sal*I or *Nco*I) and, after phenol–chloroform extraction and ethanol precipitation, used as a template for the *in vitro* transcription with T7 or SP6 RNA polymerase (Forster et al., 1990). The transcription reactions containing the primary transcripts and their self-cleavage products were separated in 5% PAGE containing 8 M urea and 1 × TBE (89 mM Tris, 89 mM boric acid, 2.5 mM EDTA, pH 8.3), stained with ethidium bromide and UV visualized.

The 5′ terminal sequence of the 3′ self-cleavage product of the FHVd-LR monomeric transcripts of both polarity strands was determined by 5′ RACE experiments. Briefly, 3′ RNA fragments were eluted from the denaturing 5% PAGE by phenol–chloroform extraction and ethanol precipitation. The eluted RNA was reverse transcribed (as described above) using FHVd-4R and FHVd-5F (see **Table 1**) for the (+) and (−) polarity RNA fragment, respectively. Following the addition of a poly(dG) tail, the tailed cDNAs were PCR amplified using GoTaq DNA polymerase (Promega, Madison, WI, United States) with a 5′ RACE primer (**Table 1**) and the same primer used for the cDNA synthesis. PCR amplicons were gel-purified, cloned, and sequenced as reported above.

RESULTS

A cDNA library was generated from dsRNAs extracted from a fig tree with symptoms of mosaic and leaf distortion, grown on Kauai, Hawaii. HTS of such a library produced 262,700 reads with an average size of approximately 500 base pair (bp), which were filtered for quality and *de novo* assembled, generating 1,183 contigs (size 100 to 17,165 nt). BlastX and BlastN searches against the viral database revealed contigs with significant sequence identity to several viruses, including

[1]https://www.geneious.com

[2]http://www.ncbi.nlm.nih.gov/

badnaviruses, closteroviruses, emaraviruses, endornaviruses, totiviruses, trichoviruses, and umbraviruses (Olmedo-Velarde et al., unpublished data). Moreover, BlastN analysis revealed two contigs of 462 and 674 nt that shared sequence similarity with some viroids and Vd-LRNAs. In particular, a short region of 40–50 nt of both contigs shared 84–87% sequence identity with HRzs contained in eggplant latent viroid (Fadda et al., 2003), grapevine hammerhead Vd-LRNA (Wu et al., 2012), and cscRNAs (Di Serio et al., 1997, 2006). Interestingly, each contig consisted of a partial direct repeat of a monomeric sequence of 358 and 390 nt, respectively (**Supplementary Figure S1**), thus suggesting a possible multimeric or circular nature of the corresponding RNAs. The monomeric sequences of the two contigs shared about 84% sequence identity with each other. More specifically, they shared almost identical sequences spanning a region of 260 nt, while the remaining part of the molecules (about 120 nt) largely diverged (**Supplementary Figure S2**). Despite the sequence diversity, the secondary structure of lowest free energy of both RNAs was of the rod-like class, with the largely divergent sequences located in the terminal right regions (**Supplementary Figure S2**). The two RNAs contained HRzs in both polarity strands (see below), thus displaying typical structural features previously reported for members of the family *Avsunviroidae* and in some Vd-LsatRNAs (Di Serio et al., 2018b; Navarro et al., 2017). Although these data strongly suggested the possible existence of at least two novel Vd-LRNAs in fig, it was not possible to confirm these preliminary results because the original fig source from Kauai was destroyed. In an attempt to identify additional isolates containing similar Vd-LRNAs, specific primers were designed to test other fig trees by RT-PCR.

Identification of a Novel Fig Viroid-Like RNA on the Island of Oahu

An RT-PCR-based preliminary survey was performed on the island of Oahu, Hawaii, using primers FHVd-1F and FHVd-2R (**Table 1**) derived from two adjacent regions of a sequence common to the two contigs and, therefore, expected to amplify full-length cDNAs of the potential circular RNAs (Candresse et al., 1998). Three out of the twelve fig trees examined, all grown in the same location on Oahu, generated amplification products of about 360 bp. Cloning and sequencing of the amplicons showed sequence variants ranging in size from 357 to 359 nt, which differed from each other in a few positions (**Figure 1**), The conserved nucleotides reported in almost all natural HRzs were found in both polarity strands of the cloned RNA variants (**Figure 1**), and BlastN searches confirmed that only the short-sequence fragment corresponding to the ribozyme domain matched with some members of the family *Avsunviroidae* and with csc-RNAs. Importantly, the sequenced variants from Oahu and the short Vd-LRNA from the fig tree on the island of Kauai shared high identity (96%), with only 14 polymorphic positions observed in a pairwise alignment between them (**Supplementary Figure S3**). Altogether, these data suggested that the RNA identified from Oahu was closely related to those found by HTS in the Kauai sample and could represent a novel Vd-LRNA, hereafter named fig hammerhead viroid-like RNA (FHVd-LR).

Primary and Proposed Secondary Structure of FHVd-LR

To further assess the nucleotide sequence composition in the region covered by the first primer set and the possible circularity of the RNA, full-length cDNAs were also amplified and cloned using two additional pairs of adjacent primers (FHVd-3F/4R and FHVd-5F/6R, **Table 1**). Sequencing of 10 and 8 clones of the RT-PCR amplicons generated with the two respective primer sets revealed variants ranging in size from 356 to 360 nt. They showed high sequence identity (97.8–100%) between them and to those obtained previously using the primer set FHVd-1F/2R. A total of nine different sequence variants were annotated in GenBank (with the accession numbers from MT57734 to MT577542). The most frequently sequenced variant was FHVd-LR_A1, which differed from the others in up to six positions and will be considered as the reference variant for this new Vd-LRNA (**Figure 1**). These results support the circular and quasispecies nature of the RNA. Multiple-sequence alignments also revealed the absence of nucleotide variability in the region covered by FHVd-1F/2R, while only two polymorphic positions were found in the regions targeted by the other two primer pairs (FHVd-3F/4R and FHVd-5F/6R). Interestingly, 9 out of 15 changes mapped in a region covering almost 100 nt, between positions 265 and 358 of the multiple sequence alignment (**Figure 1**). FHVd-LR was composed of 22.6, 28.2, 24.6, and 24.6% of A, U, C, and G, respectively. Therefore, its G + C content (49.2%) resembled that of most viroids (ranging from 52.2% to 61.6%) and Vd-LsatRNAs (ranging from 50.0 to 63.6%). However, the G + C content differed from avocado sunblotch viroid (ASBVd) (Symons, 1981), which is unique in its low G + C content of about 38%. The existence of a DNA counterpart of FHVd-LR, and therefore its potential retroviroid-like nature, was excluded based on the negative results of amplification by PCR without previous reverse transcription using nucleic acid preparations from a FHVd-LR-positive tree and different primer pairs (**Figure 2**).

When the secondary structure of minimal free energy was calculated, FHVd-LR variants assumed a rod-like conformation with about 70% paired residues (**Figure 3**). In such a structure, the conserved sequences in the natural HRzs were located in a central region and opposite to each other (**Figure 3**), as previously observed in most members of the family *Avsunviroidae* and several Vd-LsatRNAs. The proposed rod-like secondary structure of FHVd-LR was also supported by most of the heterogeneity found in the sequenced variants consisting of changes in loops or compensatory mutations, which did not result in major modifications of the proposed conformation (**Figure 3**). It is interesting that the nucleotide changes were asymmetrically distributed in the proposed structure, with most of them mapping at the terminal right domain of the molecule and covering about 120 nt.

As reported above, FHVd-LR and the short Vd-LRNA from Kauai shared high sequence identity. Interestingly, most nucleotide changes in the latter RNA mapped in the terminal right domain of the secondary structure proposed for FHVd-LR and were compensatory mutations or covariations preserving the rod-like conformation (**Figure 3**). These findings highlighted an

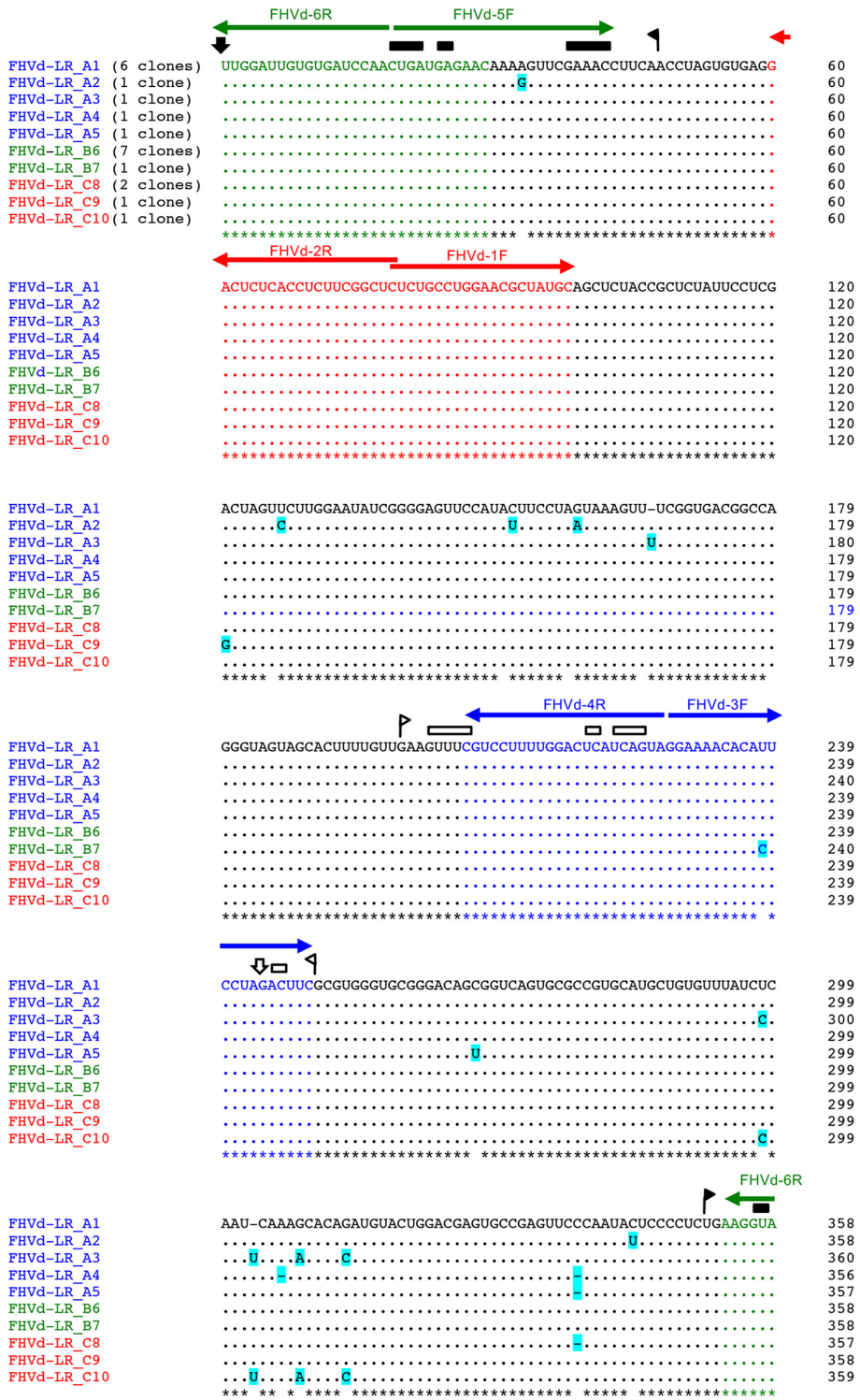


FIGURE 1 | Multiple-sequence alignment of full-length cDNAs of fig hammerhead viroid-like RNA (FHVd-LR) variants amplified with primer pairs FHVd-1F/2R (in red), FHVd-3F/4R (in blue), and FHVd-5F/6R (in green). Names of the sequence variants and primers (horizontal arrows) used to amplify the corresponding cDNAs are reported with the same colors; the numbers of clones containing the same variant are indicated in brackets. The reference variant (FHVd-LR_A1) reported at the top is identical to the variant FHVd-LR_B6 and, being the most frequently sequenced (found in 13 independent clones), also corresponds to the master sequence in the population. Nucleotide identity and gaps with respect to the FHVd-LR_A1 variant are indicated by dots and dashes, respectively. The light blue background highlights mutations. Regions involved in the formation of (+) and (–) hammerhead structures are delimited by flags; arrows and bars indicate the self-cleavage sites and the nucleotides conserved in most natural hammerhead structures, respectively; filled (black) and open symbols refer to (+) and (–) polarity, respectively. Nucleotide positions in the multiple alignments are reported on the left.

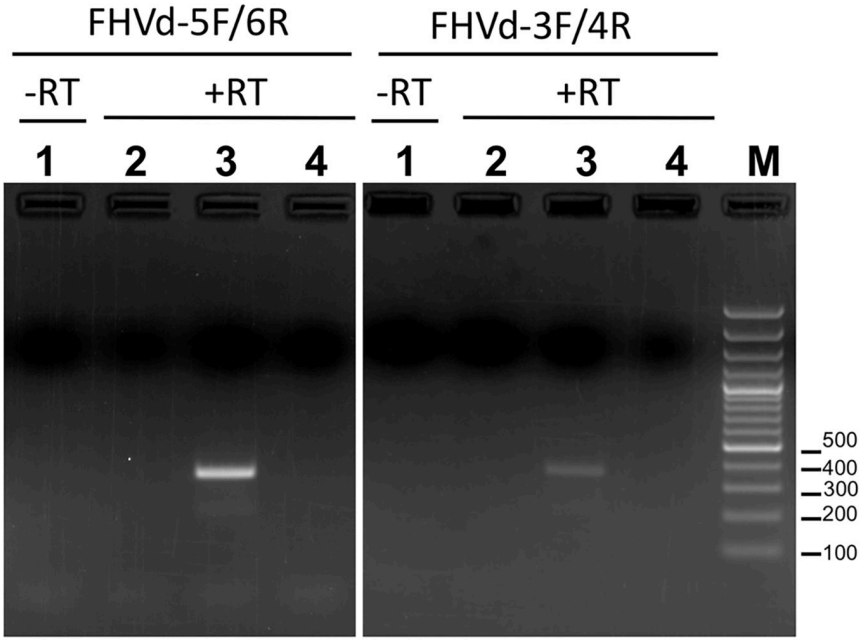


FIGURE 2 | Amplification assays using RNA preparations from FHVd-LR-positive and -negative samples, and primer pairs FHVd-5F/6R (left) or FHVd-3F/4R (right). Lane 1, amplification was performed in the absence of reverse transcription (−RT) using an FHVd-LR-positive sample. Lanes 2 and 3, amplifications were performed after reverse transcription (+RT) using RNA preparations from FHVd-LR-negative or -positive samples, respectively. Lane 4 is a non-template control. The same FHVd-LR-positive sample was tested in lanes 1 and 3.

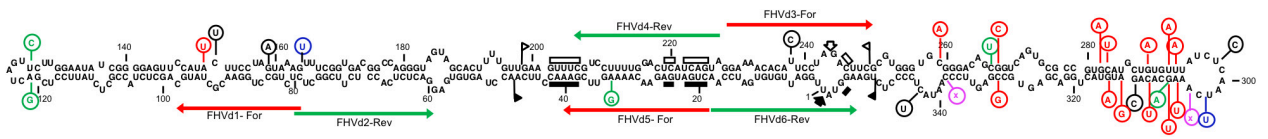


FIGURE 3 | Primary and predicted secondary structures of lowest free energy of the fig hammerhead viroid-like RNA variant A1 (FHVd-LR_A1) collected from the island of Oahu. Regions involved in the formation of (+) and (−) hammerhead structures are delimited by flags; arrows and bars indicate the self-cleavage sites and the nucleotides conserved in most natural hammerhead structures, respectively; filled (black) and open symbols refer to (+) and (−) polarity, respectively. Mutated positions in other FHVd-LR variants are indicated in circles, with compensatory mutations, insertions, and deletions reported in green, blue, and pink, respectively. Nucleotide changes with respect to the short VL-RNA from the island of Kauai are in red. Horizontal arrows delimit the regions targeted by primer pairs used in the amplification steps of the cloning and sequencing strategy.

additional structural parallelism between the variants from the Oahu samples, which we confirmed by cloning and sequencing, and the Vd-LRNAs from the Kauai isolate that were detected only *in silico* due to destruction of the original tree.

Circularity and Accumulation Levels of Both Polarity Strands of FHVd-LR

The presence of FHVd-LR in fig and evidence of its circularity were further inferred by northern blot hybridization. Nucleic acid preparations enriched in highly structured RNAs from fig trees that tested positive and negative in the RT-PCR assays were loaded side by side in a non-denaturing PAGE followed by a denaturing PAGE (double PAGE), a method specifically developed to separate circular from linear RNAs of the same size (Flores et al., 1985). Moreover, to test whether both polarity strands of FHVd-LR generated circular forms and to ascertain whether one of them accumulated at a higher level *in vivo*, equalized riboprobes specific for detecting each FHVd-LR polarity strand were used separately in parallel hybridization experiments. As expected, the circular and linear forms of FHVd-LR were only detected in the samples from the fig trees that previously tested positive by RT-PCR, thus confirming the association of this RNA with some fig plants only (**Figure 4**). Circular forms of both FHVd-LR polarity strands were detected. Based on the intensity of the hybridization signals, one polarity strand accumulated at higher level *in vivo* (**Figure 4**). Therefore, this strand, corresponding to the sequences in **Figures 1**, **2**, was considered as the (+) FHVd-LR polarity. The possibility that one probe could cross-hybridize with the strand of the same polarity due to the self-complementarity of the FHVd-LR sequence was excluded by northern blot assays showing that this did not happen under the experimental conditions used (**Supplementary Figure S4**).

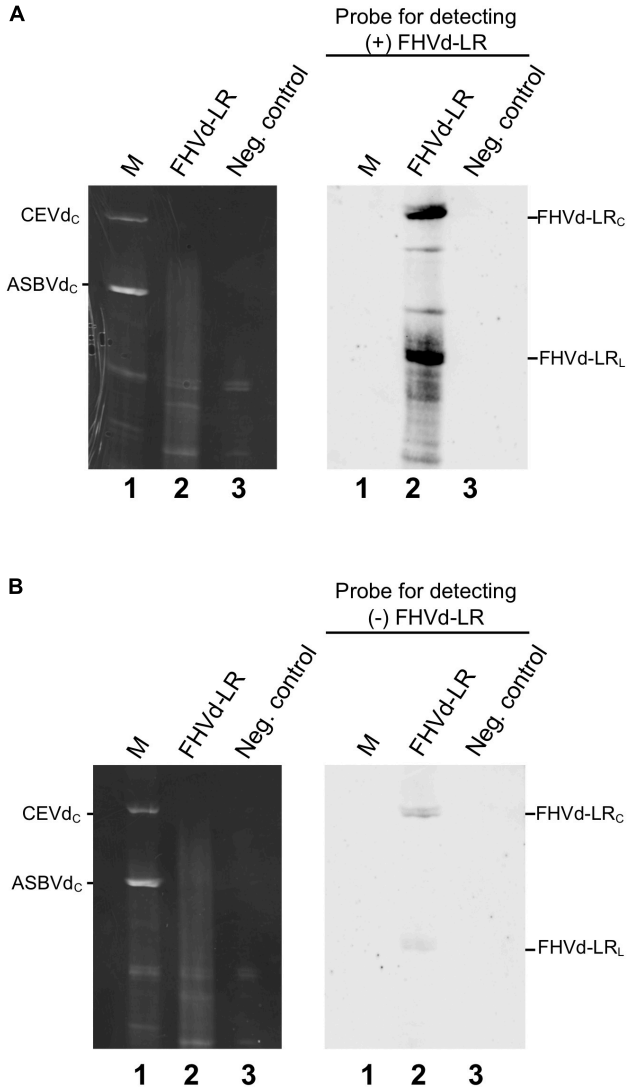


FIGURE 4 | Northern blot hybridization assays to detect circular RNAs and identify the most abundant FHVd-LR strand denoted (+) polarity strand. Fig RNA preparations separated by double PAGE (left) were transferred to nylon membranes and hybridized with equalized full-length digoxigenin-labeled riboprobes for detecting the (+) and (−) FHVd-LR strands **(A,B)**, respectively. The second (denaturing) gel of the double PAGE, stained with ethidium bromide, and the northern blot results are respectively reported on the left and on the right of each panel. Lane 1 corresponds to RNA preparations from *Gynura aurantiaca* and avocado infected by citrus exocortis viroid (CEVd, 371 nt) and avocado sunblotch viroid (ASBVd, 246 nt) respectively, used as molecular markers (M); lane 2 is an RNA preparation from a fig tree that tested positive to FHVd-LR by RT-PCR assay; lane 3 is an RNA preparation from FHVd-LR fig tree that tested negative to FHVd-LR and used as negative control. Identical aliquots of the same RNA preparations were loaded in the gels shown in panels **(A,B)**. The positions of circular forms of CEVd and ASBVd (CEVdc and ASBVdc), visible in the gel stained with ethidium bromide, are reported on the left; the positions of circular and linear forms of FHVd-LR detected by the specific probes for the (+) and (−) polarity strands [panels **(A,B)**, respectively] are indicated on the right. Bands between the circular and linear forms could correspond to artifacts generated by UV-induced cross-links during the visualization of the first non-denaturing gel of the double PAGE, as observed previously for some viroids (Hernández et al., 2006; Serra et al., 2018).

Hammerhead Ribozymes of FHVd-LR Are Active During Transcription

Fig hammerhead viroid-like RNA HRzs of both polarity strands were composed of three hairpins, two of them closed by short loops, which are located around the central core containing the predicted self-cleavage site (**Figure 5A**). In the (−) HRz, this site is preceded by a GUC trinucleotide, as in most HRzs of other viroids and Vd-LRNAs (Hutchins et al., 1986; Navarro et al., 2017). In contrast, a GUA trinucleotide was found at the same position in the (+) HRz of FHVd-LR, a situation previously reported only in the (−) HRz of the Vd-LsatRNAs of velvet tobacco mottle virus and LTSV (Forster et al., 1988; Collins et al., 1998) and in the (+) HRz of cscRNAs (Di Serio et al., 2006). The relevance *in vivo* of the FHVd-LR HRzs was confirmed by the limited sequence variability observed in these catalytic domains in the FHVd-LR variants (**Figures 1**, **3**). Indeed, the single mutation detected in the (+) HRz mapped in the loop that closes the hairpin II, while that found in the (−) HRz was a compensatory mutation located in the stem II (**Figure 5A**). Therefore, both mutations preserved the typical hammerhead structures.

In vitro transcription of recombinant plasmids containing monomeric and dimeric head-to-tail constructs of FHVd-LR indicated that HRzs of both polarity strands were active, generating RNA fragments of sizes consistent with those expected and in accord with the self-cleavage of transcripts at the predicted sites (**Figures 5B,C** and **Supplementary Figure S5**). Moreover, the self-cleavage site of both polarity strands was further confirmed by 5′ RACE experiments, showing that the 3′ fragments generated by the ribozymes in each polarity strand had the expected 5′ terminal termini (**Figure 5D**).

DISCUSSION

Analyses of HTS data from a library prepared from dsRNAs allowed us to predict the existence of two novel Vd-LRNAs in a fig tree grown on the island of Kauai, Hawaii. The two potential small circular RNAs contained HRzs in both polarity strands and had closely related sequences of 358 nt and 390 nt that mainly diverged in a specific region of about 120 nt. Discrimination between the long and short Vd-LRNAs *in silico* was possible due to the availability of long reads, up to 382 nt, which covered common and divergent sequences of the two *de novo* assembled contigs. This finding suggests that HTS technologies providing long sequence reads, like the 454 platform used in the present study, increase the chances of detecting coexisting but slightly divergent sequence variants of Vd-LRNAs.

In the absence of additional material from the original plant isolate, *in silico* data were used to design specific primers and perform a preliminary RT-PCR survey of figs on the island of Oahu, Hawaii. This analysis resulted in the identification of several fig trees associated with FHVd-LR, a small circular RNA with HRz in both polarity strands closely related to the Vd-LRNAs from Kauai. The presence of FHVd-LR only in some fig trees from Oahu, but not in others, was confirmed by northern blot hybridization. Northern blotting also showed that linear

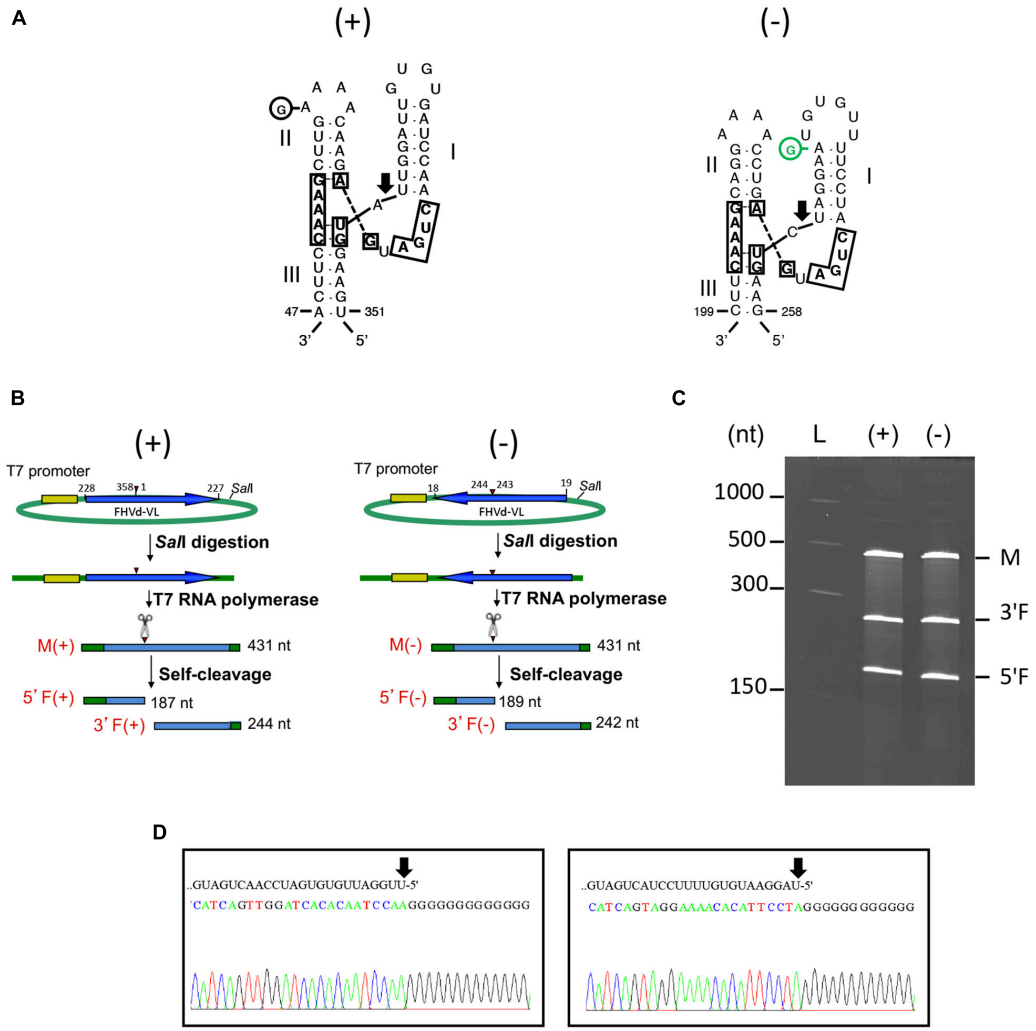


FIGURE 5 | (A) Primary and Y-shaped secondary structure of hammerhead ribozymes (HRzs) of (+) and (–) FHVd-LR are presented considering the existing X-ray crystallography data on this class of ribozymes (Pley et al., 1994). Stems I and II are closed by loops. The nucleotides of the catalytic core conserved in most natural hammerhead structures are boxed. The cleavage site of each ribozyme is indicated by an arrow. Nucleotides in the (+) and (–) polarity are numerated considering their positions in the variant FHVd-LR_A1 (+). Mutations are indicated in circles, with the compensatory mutation reported in green. **(B)** Schematic representation of the DNA templates and monomeric RNA products generated by *in vitro* transcription. Plasmids containing the monomeric sequence of FHVd-LR in opposite orientations were linearized with *Sal*I and transcribed with T7 RNA polymerase to produce monomeric transcripts (M) of opposite polarity and the respective 5′ and 3′ fragments (5′F and 3′F, respectively) derived from the HRz self-cleaving activity. In green, plasmid sequences; in yellow, polymerase promoter; in blue, FHVd-LR sequence, with the arrows indicating the (+) orientation; arrowheads and scissors mark the position of the self-cleavage sites. Numbers on the right of each RNA fragment indicate its expected size. **(C)** Analyses by PAGE of the *in vitro* transcription of the plasmids containing (+) and (–) monomeric FHVd-LR cDNA. L, RNA ladder with sizes indicated on the left; see panel **(B)** for M, 3′F and 5′F abbreviations. **(D)** Determination of self-cleavage site by 5′ RACE of 3′F fragment resulting from the HRz-mediated cleavage of FHVd-LR monomeric transcripts [generated as reported in panels **(B,C)**]. Sequencing electropherograms of 5′ RACE products of the (+) and (–) 3′F fragments are shown on the left and right, respectively, with the self-cleaved RNA sequence reported on the top and the 5′ terminal nucleotide indicated by the arrow.

and circular forms of both polarity strands of this RNA exist *in vivo*, with one strand accumulating at a higher level than the other. Further experiments showed that this RNA has no DNA counterpart and that it accumulates in infected tissues as a population of sequence variants differing from each other in a few positions. These are typical features of quasispecies previously reported for other RNA replicons, including viruses, viroids, and VL-sat RNAs (Biebricher and Eigen, 2006; Codoñer et al., 2006; Flores et al., 2014).

Hammerhead ribozyme-mediated self-cleavage of both polarity strands of FHVd-LR during transcription was demonstrated and correspondence between the predicted and actual self-cleavage sites confirmed by RACE experiments. Interestingly, the nucleotide changes observed in the HRzs of both polarity strands of FHVd-LR did not impair the formation of the stable hammerhead structures, thus upholding the relevance *in vivo* of these ribozymes. Altogether, these data strongly support that FHVd-LR is a novel Vd-LRNA that likely

replicates through a symmetric rolling-circle mechanism as previously proposed for some viroids and Vd-LsatRNAs (Flores et al., 2011). However, these structural and biochemical features did not provide any indication on whether FHVd-LR is a viroid or a Vd-LsatRNA. Indeed, HRzs in both polarity strands have been reported in three Vd-LsatRNAs and in all viroids classified in the family *Avsunviroidae*. HRzs were also found in both polarity strands of two Vd-LRNAs whose nature remains unclear (grapevine hammerhead Vd-LRNA and cscRNAs, which possibly are a viroid and two Vd-LsatRNAs of a mycovirus, respectively) (Alioto et al., 2003; Wu et al., 2012; Minoia et al., 2014).

Fig hammerhead viroid-like RNA assumes a rod-like conformation of minimal free energy, the relevance *in vivo* of which is supported by the variability observed in the FHVd-LR population accumulating in fig tissues. Indeed, most nucleotide changes detected in the FHVd-LR sequence variants mapped at loops or were compensatory mutations or covariations preserving the proposed structure. Rod-like structures have been proposed for most viroids and Vd-LsatRNAs. In members of the family *Avsunviroidae*, the G + C content, morphology, and thermodynamic stability of the HRzs and the secondary structure are of taxonomic relevance (Di Serio et al., 2018b). When these structural elements are taken into consideration, FHVd-LR seems to diverge from all members of the family *Avsunviroidae*. Among them, only ASBVd adopts a rod-like conformation *in silico*, *in vitro*, and *in vivo* (Symons, 1981; López-Carrasco and Flores, 2017). However, with a 38% G + C content and thermodynamically unstable HRzs due to a short stem III (Forster et al., 1988), this viroid appears different from FHVd-LR, which has about 50% G + C and stable HRzs. In contrast, the remaining members of the family *Avsunviroidae*, which share with FHVd-LR a high G + C content and stable HRzs (Di Serio et al., 2018b; Serra et al., 2018), assume branched or bifurcated conformations. Actually, among all known Vd-LRNAs, only the Vd-LsatRNA of LTSV and the two Vd-LRNAs from cherry (cscRNAs) resemble FHVd-LR in their rod-like secondary structure, high G + C content (about 50%), and stable HRzs in both polarity strands (Forster and Symons, 1990; Di Serio et al., 2006). Additionally, akin to the HRzs of FHVd-LR, the (+) and (−) HRzs of the two cscRNAs and the (−) HRz of VdL-satRNA of LTSV also contain an atypical adenine residue preceding the self-cleavage site.

In this context, note that the *in silico* data reported here supported the coexistence of at least two Vd-LRNAs of different size in the fig sample from Kauai, a hypothesis that was not conclusively tested due to the destruction of the original isolate. Attempted amplifications by RT-PCR, using different primer pairs, of Vd-LRNAs larger than FHVd-LR in the Oahu isolates have so far been unsuccessful, suggesting the absence of larger coinfecting Vd-LRNAs. Moreover, the two cscRNAs coinfecting cherry trees were not simultaneously observed in all the studied isolates, with some of them being infected by only one (Minoia et al., 2014).

Although the FHVd-LR structural features are closer to the VL-satRNA of LTSV and to cscRNAs than to viroids, we are unable to predict the biological nature of this novel Vd-LRNA from fig. Only bioassays may answer the specific question of whether FHVd-LR is able to replicate autonomously and systemically infect fig (Di Serio et al., 2018a). In this respect, experiments to test the transmission of FHVd-LR to virus-free fig plants by grafting and slash-inoculation of RNA preparations enriched in Vd-LRNAs and of *in vitro*-generated dimeric head-to tail transcripts are ongoing. Slash-inoculated plants were negative by RT-PCR and northern blot assays one and two months post-inoculation, but such bioassays in woody hosts may require longer times for systemic infection (Di Serio et al., 2018a).

Although all viruses identified by HTS as potentially infecting the original fig tree from Kauai belong to viral taxa not containing helper viruses of Vd-LsatRNAs, the association of FHVd-LR with a plant virus (known or unknown) cannot be ruled out; further studies are needed. The association of FHVd-LR with a virus infecting a fungus or another organism associated with fig cannot be excluded either. It has been in fact reported that HTS applied to cDNA libraries from plant tissues may detect viruses (and possibly Vd-LRNAs) infecting other transiently plant-associated organisms (Hily et al., 2018).

CONCLUSION

In conclusion, HTS allowed the initial identification of two novel Vd-LRNAs associated with fig, providing the necessary information for molecular studies to confirm the existence, self-cleaving activity mediated by HRzs, circularity of both polarity strands, and quasispecies nature of FHVd-LR. The data reported here make available detection methods to further investigate the biology and epidemiology of this and possibly other coinfecting novel Vd-LRNAs in fig trees.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in NCBI database GenBank, accession numbers from MT57734 to MT577542.

AUTHOR CONTRIBUTIONS

FD and MM supervised the project. AO-V and BN performed the experiments. AO-V, MM, JH, FD, and BN conceived the study and analyzed the data. FD, AO-V, and BN wrote the manuscript. All authors revised the manuscript.

FUNDING

This work was supported by USDA NIFA Hatch project HAW09030-H managed by the College of Tropical Agriculture and Human Resources at the University of Hawai'i at Mānoa. This project has also received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 734736. This article

reflects only the author's view, and the agency is not responsible for any use that may be made of the information it contains.

ACKNOWLEDGMENTS

We sincerely thank Professor Ricardo Flores for critical reading of the manuscript and for suggestions. We would also like to thank the reviewers and the editor for their comments contributing to improve the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01903/full#supplementary-material

FIGURE S1 | Contigs of different size consisting of direct repeats (green arrows) generated by overlapping reads (in pink) obtained by HTS of a cDNA library generated from dsRNA extracted from a fig tree grown on the island of Kauai, Hawaii.

FIGURE S2 | (A) Pairwise comparison between a short and a long viroid-like RNA (Vd-LRNA) identified *in silico* in a fig sample from the island of Kauai. Positions with identical nucleotides are denoted by asterisks. **(B)** Primary and predicted secondary structure of the free energy of the long and short Vd-LRNA variants from Kauai. Differences between both RNA molecules are denoted in red. Regions involved in the formation of (+) and (−) hammerhead structures are delimited by flags; arrows and bars indicate the self-cleavage sites and the nucleotides conserved in most natural hammerhead structures, respectively; filled (black) and open symbols refer to (+) and (−) polarity, respectively.

FIGURE S3 | Pairwise comparison between the short viroid-like RNA (Vd-LRNA) from island of Kauai and FHVd-LR (master sequence) from the island of Oahu. The sequence identity is 96.1%. Positions with identical nucleotides are denoted by asterisks.

FIGURE S4 | Northern blot hybridization with equalized full-length digoxigenine-labeled riboprobes for detecting FHVd-LR (+) or (−) strands (left and right panel, respectively). In each panel, lanes 1 and 2 correspond to equal amounts of dimeric FHVd-LR (+) and (−) transcripts, respectively. Positions of the FHVd-LR (+) and (−) linear monomeric forms (m) are indicated by an arrow. Position and size (in nt) of the RNAs used as molecular markers are reported on the left.

FIGURE S5 | (A) Schematic representation of plasmids containing head-to-tail dimeric constructs of FHVd-LR and of the products generated by *in vitro* transcription. Plasmids containing the dimeric sequence of FHVd-LR in opposite orientations were linearized with *Sal*I or *Nco*I and transcribed with T7 or SP6 RNA polymerase, respectively. Transcription of these templates is expected to produce complete dimeric transcripts (D), fragments longer than a monomer (5′F-M and 3′F-M) generated by the self-cleavage of one hammerhead ribozyme (HRz), monomeric RNAs (M) and fragments (5′F and 3′F) generated by the self-cleavage of both HRzs. In green, plasmid sequences; in yellow, polymerase promoter; in blue, FHVd-LR sequence, with the arrows indicating the (+) orientation; arrowheads and scissors mark the positions of the self-cleavage sites. Numbers on the left of each RNA fragment indicate its expected size. **(B)** Analysis by PAGE of the *in vitro* transcription of plasmids containing (+) and (−) dimeric FHVd-LR cDNA; L1 and L2, RNA ladders with sizes indicated on the left and on the right, respectively; for other symbols see panel **(A)**.

REFERENCES

Ale-Agha, G. N., and Rakhshandehroo, F. (2014). Detection and molecular variability of fig fleck-associated virus and fig cryptic virus in Iran. *J. Phytopathol.* 162, 417–425. doi: 10.1111/jph.12204

Alioto, D., Zaccaria, F., Covelli, L., Di Serio, F., Ragozzino, A., and Milne, R. G. (2003). Light and electron microscope observations on chlorotic rusty spot, a disorder of cherry in Italy. *J. Plant Pathol.* 85, 215–218. doi: 10.4454/jpp.v85i3.1033

Biebricher, C. K., and Eigen, M. (2006). What is a quasispecies? *Curr. Top. Microbiol.* 229, 1–31. doi: 10.1007/3-540-26397-7_1

Candresse, T., Hammond, R. W., and Hadidi, A. (1998). "Detection and identification of plant viruses and viroids using polymerase chain reaction (PCR)," in *Plant Virus Disease Control*, eds A. Hadidi, R. K. Khetarpal, and H. Koganezawa (St Paul: APS Press), 399–416.

Chiumenti, M., Torchetti, E. M., Di Serio, F., and Minafra, A. (2014). Identification and characterization of a viroid resembling apple dimple fruit viroid in fig (*Ficus carica* L.) by next generation sequencing of small RNAs. *Virus Res.* 188, 54–59. doi: 10.1016/j.virusres.2014.03.026

Codoñer, F. M., Daròs, J. A., Sole, R. V., and Elena, S. F. (2006). The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* 2:0020136. doi: 10.1371/journal.ppat.0020136

Collins, R. F., Gellatly, D. L., Sehgal, O. P., and Abouhaidar, M. G. (1998). Self-cleaving circular RNA associated with rice yellow mottle virus is the smallest viroid-like RNA. *Virology* 241, 269–275. doi: 10.1006/viro.1997.8962

Daròs, J. A., and Flores, R. (1995). Identification of a retroviroid-like element from plants. *Proc. Natl. Acad. Sc.i U.S.A.* 92, 6856–6860. doi: 10.1073/pnas.92.15.6856

Di Serio, F., Ambrós, S., Sano, T., Flores, R., and Navarro, B. (2018a). Viroid diseases in pome and stone fruit trees and Koch's postulates: a critical assessment. *Viruses* 10:612. doi: 10.3390/v10110612

Di Serio, F., Daròs, J. A., Ragozzino, A., and Flores, R. (1997). A 451-nucleotide circular RNA from cherry with hammerhead ribozymes in its strands of both polarities. *J. Virol.* 71, 6603–6610. doi: 10.1128/jvi.71.9.6603-6610.1997

Di Serio, F., Daròs, J. A., Ragozzino, A., and Flores, R. (2006). Close structural relationship between two hammerhead viroid-like RNAs associated with cherry chlorotic rusty spot disease. *Arch. Virol.* 151, 1539–1549. doi: 10.1007/s00705-006-0732-0

Di Serio, F., Flores, R., Verhoeven, J. T., Li, S. F., Pallás, V., Randles, J. W., et al. (2014). Current status of viroid taxonomy. *Arch. Virol.* 159, 3467–3478. doi: 10.1007/s00705-014-2200-6

Di Serio, F., Li, S. F., Matousek, J., Owens, R. A., Pallás, V., Randles, J. W., et al. (2018b). ICTV virus taxonomy profile: *Avsunviroidae*. *J. Gen. Virol.* 99, 611–612. doi: 10.1099/jgv.0.001045

Di Serio, F., Navarro, B., and Flores, R. (2017). "Origin and evolution of viroids," in *Viroids and Satellites*, eds A. Hadidi, R. Flores, J. W. Randles, and P. Palukaitis (Cambridge: Academic Press), 125–134. doi: 10.1016/b978-0-12-801498-1.00012-7

Elbeaino, T., Abou Kubaa, R., Ismaeil, E., Mando, J., and Digiaro, M. (2012). Viruses and hop stunt viroid of fig trees in Syria. *J. Plant Pathol.* 94, 687–691. doi: 10.4454/JPP.V94I3.002

Elbeaino, T., Digiaro, M., Heinoun, K., De Stradis, A., and Martelli, G. P. (2010). Fig mild mottle-associated virus, a novel closterovirus infecting fig. *J. Plant Pathol.* 92, 165–172. doi: 10.4454/jpp.v92i1.26

Fadda, Z., Daròs, J. A., Flores, R., and Durán-Vila, N. (2003). Identification in eggplant of a variant of citrus exocortis viroid (CEVd) with a 96 nucleotide duplication in the right terminal region of the rod-like secondary structure. *Virus Res.* 97, 145–149. doi: 10.1016/j.virusres.2003.08.002

Flores, R., Durán-Vila, N., Pallás, V., and Semancik, J. S. (1985). Detection of viroid and viroid-like RNAs from grapevine. *J. Gen. Virol.* 66, 2095–2102. doi: 10.1099/0022-1317-66-10-2095

Flores, R., Gago-Zachert, S., Serra, P., Sanjuán, R., and Elena, S. F. (2014). Viroids: survivors from the RNA world? *Annu. Rev. Microbiol.* 68, 395–414. doi: 10.1146/annurev-micro-091313-103416

Flores, R., Grubb, D., Elleuch, A., Nohales, M. A., Delgado, S., and Gago, S. (2011). Rolling-circle replication of viroids, viroid-like satellite RNAs and hepatitis delta virus: variations on a theme. *RNA Biol.* 8, 200–206. doi: 10.4161/rna.8.2.14238

Flores, R., Minoia, S., Carbonell, A., Gisel, A., Delgado, S., López-Carrasco, A., et al. (2015). Viroids, the simplest RNA replicons: how they manipulate their hosts for being propagated and how their hosts react for containing the infection. *Virus Res.* 209, 136–145. doi: 10.1016/j.virusres.2015.02.027

Forster, A. C., Davies, C., Hutchins, C. J., and Symons, R. H. (1990). Characterization of self-cleavage of viroid and virusoid RNAs. *Methods Enzymol.* 181, 583–607. doi: 10.1016/0076-6879(90)81153-l

Forster, A. C., Davies, C., Sheldon, C. C., Jeffries, A. C., and Symons, R. H. (1988). Self-cleaving viroid and newt RNAs may only be active as dimers. *Nature* 334, 265–267. doi: 10.1038/334265a0

Forster, A. C., and Symons, R. H. (1990). Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell* 49, 211–220. doi: 10.1016/0092-8674(87)90562-9

Gago-Zachert, S. (2016). Viroids, infectious long non-coding RNAs with autonomous replication. *Virus Res.* 212, 12–24. doi: 10.1016/j.virusres.2015.08.018

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Hadidi, A., Flores, R., Candresse, T., and Barba, M. (2016). Next-generation sequencing and genome editing in plant virology. *Front. Microbiol.* 7:1325. doi: 10.3389/fmicb.2016.01325

Hajizadeh, M., Navarro, B., Bashir, N. S., Torchetti, E. M., and Di Serio, F. (2012). Development and validation of a multiplex RT-PCR method for simultaneous detecting five grapevine viroids. *J. Virol. Methods* 179, 62–69. doi: 10.1016/j.jviromet.2011.09.022

Hegedus, K., Dallmann, G., and Balazs, E. (2004). The DNA form of a retroviroid-like element is involved in recombination events with itself and with the plant genome. *Virology* 325, 277–286. doi: 10.1016/j.virol.2004.04.035

Hernández, C., Di Serio, F., Ambrós, S., Daròs, J. A., and Flores, R. (2006). An element of the tertiary structure of Peach latent mosaic viroid RNA revealed by UV irradiation. *J. Virol.* 80, 9336–9340. doi: 10.1128/JVI.00630-06

Hily, J. M., Candresse, T., Garcia, S., Vigne, E., Tannière, M., Komar, V., et al. (2018). High-throughput sequencing and the viromic study of grapevine leaves: from the detection of grapevine-infecting viruses to the description of a new environmental *Tymovirales* member. *Front. Microbiol.* 9:1782. doi: 10.3389/fmicb.2018.01782

Hutchins, C. J., Rathjen, P. D., Forster, A. C., and Symons, R. H. (1986). Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res.* 14, 3627–3640. doi: 10.1093/nar/14.9.3627

Kovalskaya, N., and Hammond, R. W. (2014). Molecular biology of viroid-host interactions and disease control strategies. *Plant Sci.* 228, 48–60. doi: 10.1016/j.plantsci.2014.05.006

Laney, A. G., Hassan, M., and Tzanetakis, I. E. (2012). An integrated badnavirus is prevalent in fig germplasm. *Phytopathology* 102, 1182–1189. doi: 10.1094/PHYTO-12-11-0351

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Li, R., Mock, R., Huang, Q., Abad, J., Hartung, J., and Kinard, G. (2008). 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

López-Carrasco, A., and Flores, R. (2017). The predominant circular form of avocado sunblotch viroid accumulates in planta as a free RNA adopting a rod-shaped secondary structure unprotected by tightly bound host proteins. *J. Gen. Virol.* 98, 1913–1922. doi: 10.1099/jgv.0.000846

Lorenz, R., Bernhart, S. H., Höner, Z., Siederdissen, C., Tafer, H., Flamm, C., et al. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26. doi: 10.1186/1748-7188-6-26

Machado, M., Magalhães, W. C., Sene, A., Araújo, B., Faria-Campos, A. C., Chanock, S. J., et al. (2011). Phred-Phrap package to analyses tools: a pipeline to facilitate population genetics re-sequencing studies. *Investig. Genet.* 2:3. doi: 10.1186/2041-2223-2-3

Melzer, M. J., Borth, W. B., Sether, D. M., Ferreira, S., Gonsalves, D., and Hu, J. S. (2010). Genetic diversity and evidence for recent modular recombination in Hawaiian Citrus tristeza virus. *Virus Genes* 40, 111–118. doi: 10.1007/s11262-009-0409-3

Minafra, A., Savino, V., and Martelli, G. P. (2017). Virus diseases of fig and their control. *Acta Horticulturae* 1173, 237–244. doi: 10.17660/actahortic.2017.1173.41

Minoia, S., Navarro, B., Covelli, L., Barone, M., Garcia-Becedas, M. T., Ragozzino, A., et al. (2014). Viroid-like RNAs from cherry trees affected by leaf scorch disease: further data supporting their association with mycoviral double-stranded RNAs. *Arch. Virol.* 159, 589–593. doi: 10.1007/s00705-013-1843-z

Navarro, B., and Di Serio, F. (2018). "Double-stranded RNA-enriched preparations to identify viroids by next-generation sequencing," in *Viral Metagenomics: Methods and Protocols. Methods in Molecular Biology*, Vol. 1746, eds V. Pantaleo and M. Chiumenti (New York, NY: Humana Press), 37–43. doi: 10.1007/978-1-4939-7683-6_3

Navarro, B., Rubino, L., and Di Serio, F. (2017). "Small circular satellite RNAs," in *Viroids and Satellites*, eds A. Hadidi, R. Flores, J. W. Randles, and P. Palukaitis (Cambridge: Academic Press), 659–669. doi: 10.1016/b978-0-12-801498-1.00061-9

Olmedo-Velarde, A., Park, A. C., Sugano, J., Uchida, J. Y., Kawate, M., Borth, W. B., et al. (2019). Characterization of ti ringspot-associated virus, a novel *Emaravirus* associated with an emergent ringspot disease of *Cordyline fruticose*. *Plant Dis.* 103, 2345–2352. doi: 10.1094/PDIS-09-18-1513-RE

Pallás, V., Navarro, A., and Flores, R. (1987). Isolation of a viroid-like RNA from hop different from hop stunt viroid. *J. Gen. Virol.* 68, 3201–3205. doi: 10.1099/0022-1317-68-12-3201

Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature* 372, 68–74. doi: 10.1038/372068a0

Serra, P., Messmer, A., Sanderson, D., James, D., and Flores, R. (2018). Apple hammerhead viroid-like RNA is a bona fide viroid: autonomous replication and structural features support its inclusion as a new member in the genus *Pelamoviroid*. *Virus Res.* 249, 8–15. doi: 10.1016/j.virusres.2018.03.001

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Symons, R. H. (1981). Avocado sunblotch viroid: primary sequence and proposed secondary structure. *Nucleic Acids Res.* 9, 6527–6537. doi: 10.1093/nar/9.23.6527

Symons, R. H., and Randles, J. W. (1999). "Encapsidated circular viroid-like satellite RNAs (virusoids) of plants," in *Satellites and Defective Viral RNAs*, eds P. K. Vogt and A. O. Jackson (Berlin: Springer-Verlag).

Vera, A., Daròs, J. A., Flores, R., and Hernández, C. (2000). The DNA of a plant retroviroid-like element is fused to different sites in the genome of a plant pararetrovirus and shows multiple forms with sequence deletions. *J. Virol.* 74, 10390–10400. doi: 10.1128/jvi.74.22.10390-10400.2000

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109, 716–725. doi: 10.1094/PHYTO-07-18-0257-RVW

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Wu, Q., Wang, Y., Cao, M., Pantaleo, V., Burgyan, J., Li, W. X., et al. (2012). Homology-independent discovery of replicating pathogenic circular RNAs by deep sequencing and a new computational algorithm. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3938–3943. doi: 10.1073/pnas.1117815109

Yakoubi, S., Elleuch, A., Besaies, N., Marrakchi, M., and Fakhfakh, H. (2007). First report of Hop stunt viroid and Citrus exocortis viroid on fig with symptoms of fig mosaic disease. *J. Phytopathol.* 155, 125–128. doi: 10.1111/j.1439-0434.2007.01205.x

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

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.

Copyright © 2020 Olmedo-Velarde, Navarro, Hu, Melzer and Di Serio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

REVIEW
published: 11 September 2020
doi: 10.3389/fmicb.2020.553109

High-Throughput Sequencing for Deciphering the Virome of Alfalfa (*Medicago sativa* L.)

*Nicolas Bejerman[1], Philippe Roumagnac[2,3] and Lev G. Nemchinov[4]**

[1] IPAVE-CIAP-INTA and UFyMA-INTA-CONICET, Córdoba, Argentina, [2] CIRAD, BGPI, Montpellier, France, [3] BGPI, INRAE, CIRAD, Institut Agro, Université Montpellier, Montpellier, France, [4] Molecular Plant Pathology Laboratory, USDA-ARS-BARC, Beltsville, MD, United States

Alfalfa (*Medicago sativa* L.), also known as lucerne, is a major forage crop worldwide. In the United States, it has recently become the third most valuable field crop, with an estimated value of over $9.3 billion. Alfalfa is naturally infected by many different pathogens, including viruses, obligate parasites that reproduce only inside living host cells. Traditionally, viral infections of alfalfa have been considered by breeders, growers, producers and researchers to be diseases of limited importance, although they are widespread in all major cultivation areas. However, over the past few years, due to the rapid development of high-throughput sequencing (HTS), viral metagenomics, bioinformatics tools for interpreting massive amounts of HTS data and the increasing accessibility of public data repositories for transcriptomic discoveries, several emerging viruses of alfalfa with the potential to cause serious yield losses have been described. They include alfalfa leaf curl virus (family *Geminiviridae*), alfalfa dwarf virus (family *Rhabdoviridae*), alfalfa enamovirus 1 (family *Luteoviridae*), alfalfa virus S (family *Alphaflexiviridae*) and others. These discoveries have called into question the assumed low economic impact of viral diseases in alfalfa and further suggested their possible contribution to the severity of complex infections involving multiple pathogens. In this review, we will focus on viruses of alfalfa recently described in different laboratories on the basis of the above research methodologies.

Keywords: alfalfa, *Medicago sativa* L., virome, high throughput sequencing, emerging viruses

OPEN ACCESS

Edited by:
Henryk Hanokh Czosnek, The Hebrew University of Jerusalem, Israel

Reviewed by:
Xiao-Wei Wang, Zhejiang University, China
Elvira Fiallo-Olivé, Institute for Mediterranean and Subtropical Horticulture La Mayora, Spain

***Correspondence:**
Lev G. Nemchinov
lev.nemchinov@usda.gov

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

Received: *17 April 2020*
Accepted: *12 August 2020*
Published: *11 September 2020*

Citation:
Bejerman N, Roumagnac P and Nemchinov LG (2020) High-Throughput Sequencing for Deciphering the Virome of Alfalfa (Medicago sativa L.). Front. Microbiol. 11:553109. doi: 10.3389/fmicb.2020.553109

INTRODUCTION

Importance of Alfalfa Worldwide

Alfalfa (*Medicago sativa* L.), also known as lucerne, is a major forage crop worldwide cultivated in more than 80 countries, where it is mainly used as silage for grazing livestock (Samarfard et al., 2020). In the United States, it has recently become the third most valuable field crop planted on 22 million acres, with an estimated value of over $9.3 billion (National Alfalfa and Forage Alliance, 2018; Miller, 2019). Argentina is the second largest producer of alfalfa in the world, with alfalfa cultivation covering approximately 17 million acres (Basigalup and Ustarroz, 2007; Miller, 2019). Alfalfa is the principal forage crop in 15 countries of Southern, Eastern and Western Europe, where it is grown on nearly 2.5 million hectares (Julier et al., 2017). Most of these alfalfa fields (65%) are located in Italy, France, Romania, and Spain (Julier et al., 2017).

Underestimation of Viral Diseases in Alfalfa

Like most agricultural crops, alfalfa is a natural host of many plant viruses (Samac et al., 2015). However, among the groups of pathogens that infect alfalfa, viruses are the least recognized members (Samac et al., 2015; Malvick, 2020). Many, if not all, field management guides for alfalfa growers list acute diseases and pests of the crop without mentioning viral pathogens (Undersander et al., 2011, 2020; Vincelli and Smith, 2014). The reason for this lack of mention is that viruses alone are not capable of killing alfalfa plants and do not appear to cause any major damage or yield losses in the crop. Therefore, they are considered of minor importance to alfalfa production (Samac et al., 2015).

Multipathogen Infections as a Norm

The field pathology, disease management and studies of host–pathogen interactions in alfalfa, similar to many other plant species and crops, are often limited to a conventional two-way approach involving the host and a single disease-causing biological agent. In natural and cultivated populations, however, plants are frequently infected with a diverse array of pathogens, including many coinfecting viruses, that form multispecies within-host communities (Lamichhane and Venturi, 2015; Tollenaere et al., 2016; Bernardo et al., 2018). The coinfection of plants with different pathogens that may exhibit distinct life cycles, biology and modes of action can alter host susceptibility, affect the severity and dynamics of the disease and create selective pressure, driving the evolution of pathogen virulence (Alizon et al., 2013; Tollenaere et al., 2016; Abdullah et al., 2017). Both symbiotic and antagonistic relationships can occur between microbial species coinfecting a single plant (Lamichhane and Venturi, 2015; Moore and Jaykus, 2018). This is especially true for plant viruses, for which the outcomes of mixed viral infections are unforeseeable and range from coexistence to antagonism (Pruss et al., 1997; Syller, 2012; Elena et al., 2014; Syller and Grupa, 2016; Moreno and López-Moya, 2020). Virus-infected plants, for instance, can be predisposed to secondary infections with non-viral pathogens: the systemic infection of *Arabidopsis thaliana* with cauliflower mosaic virus leads to increased susceptibility to the bacterial pathogen *Pseudomonas syringae* (Zvereva et al., 2016). In contrast, virus-bacteria interactions may have beneficial effects on the host: healthy wild gourd plants (*Cucurbita pepo* ssp. *texana*) were found to contract a bacterial wilt infection at higher rates than plants already infected with zucchini yellow mosaic virus (Shapiro et al., 2013). Overall, the impact of polymicrobial infections on plant health cannot be underestimated. This applies equally to alfalfa, in which the interactions between several pathogens in "disease complexes" are poorly understood (Samac et al., 2015). In most, if not all cases, any association of viral infections with these multispecies consortia is generally unrecognized and, consequently, not assessed in detail. Nevertheless, viral infections of alfalfa represent a ubiquitous and abundant background for all other host–pathogen interactions. Quite reasonably, the same could be true for all *Plantae*.

HTS as a Tool for the Discovery of New Viruses Infecting Alfalfa

Many of the viruses infecting alfalfa have long been known (Babadoost, 1990; Samac et al., 2015). In recent years, however, a number of new viral pathogens or pathogens that were not previously described in the crop have been discovered (Bejerman et al., 2011, 2015, 2016, 2019; Roumagnac et al., 2015; Nemchinov et al., 2017a,b, 2018a; Kim et al., 2018; Gaafar et al., 2019; Jiang et al., 2019a,b; Samarfard et al., 2020). This became possible due to revolutionary advances in nucleic acid sequencing that have been leading to the replacement of traditional detection methods in plant virology with comprehensive, large-scale, unbiased, reliable high-throughput sequencing (HTS) technologies (Wren et al., 2006; Roossinck, 2012b, 2017; Nagano et al., 2015; Roossinck et al., 2015; Jones et al., 2017; Villamor et al., 2019). The goal of this review is not to address the applications of HTS for the diagnosis and characterization of plant viruses in general because there are numerous good-quality surveys and opinions covering all aspects of this subject in much detail (Barba et al., 2014; Jones et al., 2017; Roossinck, 2017; Maree et al., 2018; Olmos et al., 2018; Villamor et al., 2019). Instead, we will focus on the most recent developments that contributed to elucidating the alfalfa virome and have been implemented in the crop-specific field of alfalfa virology using HTS and the exploration of public transcriptome data repositories.

HTS METHODOLOGIES

Since no virus genes are universally conserved, virologists have developed metagenomics-based approaches that potentially detect plant viromes without *a priori* information (Roossinck, 2012a; Roossinck et al., 2015). These metagenomics-based approaches have targeted several classes of nucleic acids, including total RNA or DNA, virion-associated nucleic acids (VANA) purified from virus-like particles, double-stranded RNAs (dsRNA), and virus-derived small interfering RNAs (siRNAs) (reviewed in Roossinck et al., 2015). Interestingly, the massive acquisition of transcriptomic and genomic data has paved the way for the recovery of virus sequences hidden within these databases. We will hereafter focus on three HTS methodologies that have targeted different viral nucleic acids extracted from alfalfa plants.

Small Interfering (si) RNA Method

This approach, which was initially described by Donaire et al. (2009) and Kreuze et al. (2009), is based on the analysis of 21- to 24-nucleotide siRNAs that are processed by Dicer-like proteins during the RNA silencing process, which plays a critical role in plant resistance against viruses. It has proven effective during the last decade for detection of known and previously uncharacterized plant viruses (Pooggin, 2018). Specifically, this approach was successful in identifying viruses from alfalfa plants displaying symptoms of alfalfa dwarf disease (Bejerman et al., 2011, 2015), which is prevalent (over 70%) in several growing regions of Argentina and may lead to yield reductions of up to 30% (Lenardon, personal communication). siRNAs extracted

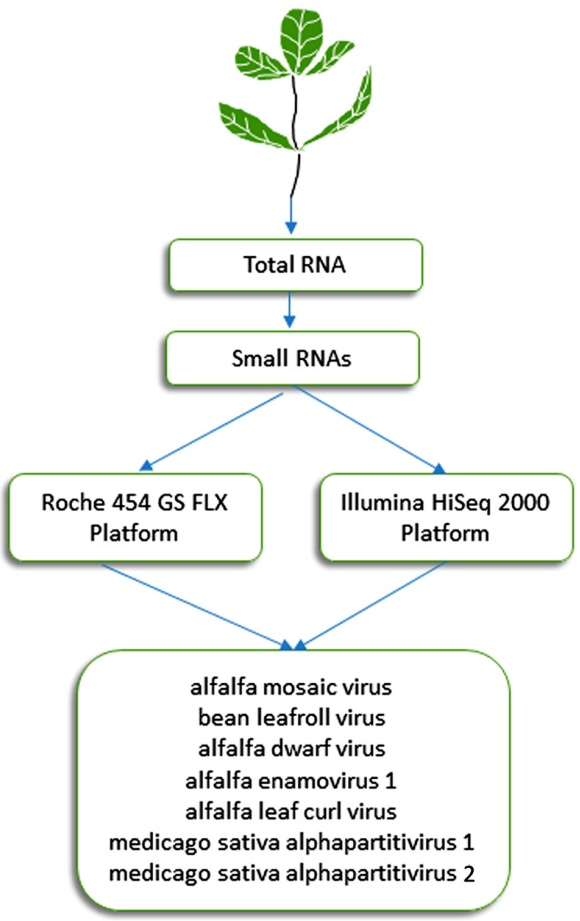


FIGURE 1 | Schematic diagram of the HTS approaches used in Argentina for the identification of seven viruses in alfalfa (Bejerman et al., 2011, 2015, 2016, 2019; Trucco et al., 2014, 2016).

from symptomatic alfalfa plants with shortened internodes, a bushy appearance, leaf puckering and varying-sized vein enations on abaxial leaf surfaces were subjected to HTS on an Illumina HiSeq 2000 system (Illumina, United States). This methodology resulted in the identification of six viruses (**Figure 1**) (Bejerman et al., 2015, 2016, 2018, 2019), which are described below in more detail. The sequencing of siRNAs was also carried out by Guo et al. (2019), who extracted total RNA from pooled alfalfa samples showing different symptomatology using the EASYspin Plant Micro RNA Rapid Extraction Kit and subjected it to the high-throughput sequencing of small RNAs on the Illumina HiSeq 4000 platform. Short reads were assembled into contigs using Velvet 1.0 software. This work resulted in the identification of three known alfalfa viruses (Guo et al., 2019).

Enrichment of Viral Particles and Virion-Associated Nucleic Acid Sequencing

High-throughput sequencing approaches allowing the enrichment of viral sequences and, thus, improving the sensitivity of virus detection, have also gained popularity during the last decade for inventorying plant virus diversity (Bernardo et al., 2018; Ma et al., 2019). One of these strategies is the VANA metagenomics-based approach (Candresse et al., 2014; Filloux et al., 2015; François et al., 2018), which was successfully used to screen and identify alfalfa viruses in France (Roumagnac et al., 2015). VANA is a combination of several technical steps leading to the concentration of viral particles. These steps include filtration, differential centrifugation and the removal of non-encapsidated material by DNase and RNase digestion. Encapsidated viral DNA and RNA are then extracted from enriched viral particles, RNA is converted to cDNA using a 26 nt primer (Dodeca Linker), and double-stranded (ds) DNA is synthesized using large (Klenow)-fragment of DNA polymerase (**Figure 2**). dsDNAs are further PCR amplified, and the amplicon libraries are sequenced on either the 454 GS FLX Titanium platform (Bernardo et al., 2018) or the Illumina MiSeq platform as 2 × 300 bp paired-end reads (Nemchinov et al., 2018a). Viral genomes are assembled using Spades (François et al., 2018), CLC Genomics Workbench (Candresse et al., 2014), IDBA-UD (Peng et al., 2012; Ma et al., 2019), CAP3 (Huang and Madan, 1999; Bernardo et al., 2018), and other assemblers. Another approach allowing the enrichment of viral sequences was carried out by Samarfard et al. (2020), who applied the dsRNA immunocapture technique (Blouin et al., 2016) to characterize the virome of alfalfa plants in Australia. dsRNA was captured with specific monoclonal antibodies prior to HTS on an Illumina platform. This work led to the identification of several known alfalfa viruses and the discovery of a new emaravirus tentatively named alfalfa ringspot-associated virus (Samarfard et al., 2020).

Standard Protocols With Total Plant RNA

Conventional procedures for HTS utilizing total plant RNA have been shown to be efficient for the identification of alfalfa viruses (Nemchinov et al., 2015, 2017a,b; Gaafar et al., 2019). Total RNA is generally extracted from alfalfa samples using the TRIzol RNA isolation reagent (Thermo Fisher Scientific, United States), innuPREP RNA Mini Kit (Analitik Jena AG, Germany) or the RNAeasy Plant Mini Kit (Qiagen, MD, United States) as described by the manufacturers. RNA-Seq libraries are prepared from total RNA using a selection of polyadenylated RNA transcripts for the removal of ribosomal RNA and to obtain higher coverage and better accuracy (Zhao et al., 2018; Gaafar et al., 2019). High-throughput sequencing assays are outsourced and performed on an Illumina platform. Similarly to other protocols, complete genomes are obtained by the assembly of paired-end reads using tools such as SPAdes (Bankevich et al., 2012), Velvet (Zerbino and Birney, 2008), Geneious software (Biomatters Limited, Auckland, New Zealand) or a Qiagen CLC Genomics Workbench and by mapping to the reference genomes, when available, using the Bowtie 2 aligner (Langmead and Salzberg, 2012). Assembly first and mapping-first approaches are usually alternated, depending on the number of short reads and the availability of annotated reference genomes.

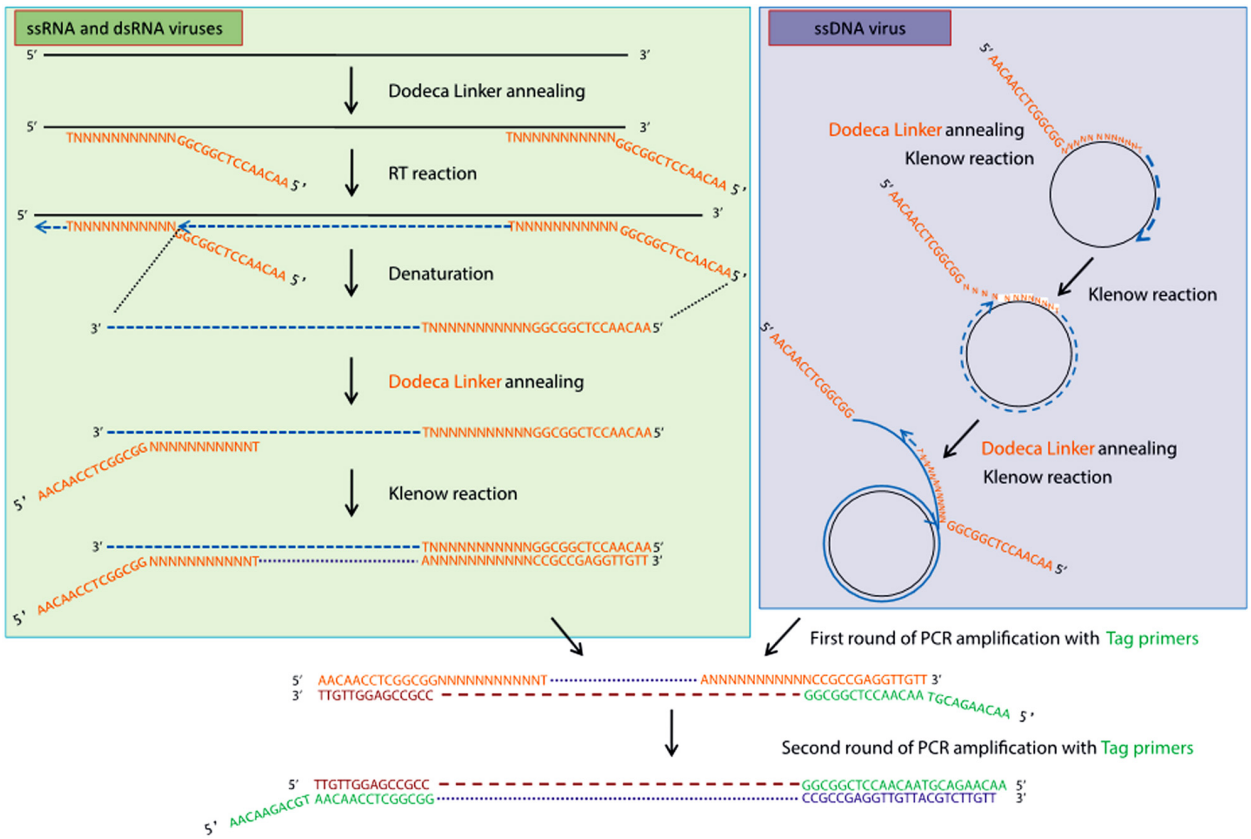


FIGURE 2 | Schematic outline of the VANA-based metagenomics method (François et al., 2018). **Upper left panel:** conversion by the random priming of viral ssRNA or dsRNA with sequence-ready cDNA, including a reverse transcription step, followed by a Klenow reaction step. **Upper right panel:** conversion by the random priming of viral DNA (e.g., circular ssDNA) with sequence-ready cDNA, including a Klenow reaction step (strand displacement amplification). **Bottom panel:** double-stranded DNA is amplified using a single PCR multiplex identifier primer, which yields amplicons that are all tagged at both ends with the same multiplex identifier primer (François et al., 2018). License to publish this figure obtained from the publisher (license #4897140253092).

USING PUBLIC HTS REPOSITORIES AS AN OPEN SOURCE FOR THE DISCOVERY OF VIRAL PATHOGENS

HTS in general, including not only technologies specifically focused on virome research but also those fine-tuned for profiling transcriptional activity in biological systems, provides an immense amount of raw sequencing data. In many cases, the scientific interests of the original submitters of sequences are limited to their respective fields of study, which are frequently unrelated to the discovery of novel transcripts. In other words, only a small portion of the deposited data is subjected to scientific scrutiny, often leaving a considerably larger amount of information untouched, unknown and freely available to the research community. For virologists, this leads to exciting opportunities for the exploration and retrieval of new viral genomes, thus improving the understanding of the diversity and host range of these pathogens and providing essential tools for their diagnosis and characterization. Unsurprisingly, public repositories have become an invaluable tool for the discovery of new pathogens, particularly viral sequences (Bekal et al., 2014; Mushegian et al., 2016; Nibert et al., 2016; Kim et al., 2018; Debat and Bejerman, 2019; Gilbert et al., 2019; Jiang et al., 2019a,b; Lauber et al., 2019; Vieira and Nemchinov, 2019). Computational pipelines for the identification of viral sequences in transcriptomic datasets are numerous and range from custom software packages (Ho and Tzanetakis, 2014; Li et al., 2016; Yamashita et al., 2016; Rott et al., 2017; Zhao et al., 2017; Zheng et al., 2017) (which are often not available for download, may not have an online interface or be regularly updated, or may require the creation of an account with the software providers) to various combinations of freely available short read assemblers, sequence-mapping algorithms and queries for virus-specific domains (Kreuze et al., 2009; Ho and Tzanetakis, 2014; Lambert et al., 2018; Massart et al., 2019). For the purpose of this review, we will discuss in more detail an uncomplicated pipeline used in one of our laboratories for the retrieval of viral sequences from public alfalfa datasets (Jiang et al., 2019a).

As a first step, alfalfa transcriptomic datasets are retrieved from the NCBI Sequence Read Archive (SRA)[1]. The raw sequencing reads are then mapped to the reference genomes

[1]https://www.ncbi.nlm.nih.gov/sra

of *Medicago sativa*[2] and a close relative of alfalfa with an annotated genome, *Medicago truncatula*[3]. This is necessary to locate and remove host-derived sequences (host filtering step). Read mapping is usually performed with Bowtie 2 tool (Langmead and Salzberg, 2012). Those reads that are not mapped to the reference genomes are further assembled into contigs using the SPAdes open source software (Bankevich et al., 2012) and searched by BLASTn with default settings and customized parameters (Jiang et al., 2019a) against other plant genomes for a second time to ensure that there is no possible cross-run contamination. Reads that are not mapped to any plant species are next aligned to the NCBI viral genome database, a public resource for virus genome sequence data[4]. Alignments are performed using BBMap[5], DNASTAR SeqMan NGen software[6] and Bowtie 2 with very sensitive settings (Jiang et al., 2019a). The very sensitive settings improve the search effort of Bowtie 2 by increasing the cutoffs for which Bowtie will stop searching. The reads that are mapped to the reference viral sequences related to the viruses of interest and to the assembled viral contigs from the datasets are then sequestered and assembled *de novo* using SPAdes (Bankevich et al., 2012). A simplified illustration of the bioinformatics pipeline is presented in **Figure 3**.

ALFALFA VIROME: RECENT DEVELOPMENTS

Previously Known but Not Fully Characterized Alfalfa Viruses Identified by HTS

Alfalfa Mosaic Virus

Alfalfa mosaic virus (AMV) is one of the most important plant viruses distributed worldwide, with a very broad host range (Maina et al., 2019). Despite the large amount of data accumulated on AMV, the application of HTS was necessary to obtain the first complete nucleotide sequence of AMV isolated from alfalfa as a natural host (Trucco et al., 2014). Diseased alfalfa plants exhibited shortening of the internodes, chlorosis at the margins and ribs of the leaflets and vein enations of varying sizes on their abaxial surfaces (**Figure 4**). Total RNA was purified from enriched viral particles (**Figure 1**) and used as a template to build libraries that were sequenced on the Roche 454 GS FLX platform (Trucco et al., 2014). The isolate of the virus, designated AMV-Argentina (AMV-Arg), shared a high identity with and presented a similar genome structure to AMV isolates infecting other hosts (Trucco et al., 2014). It is worth noting that the symptoms described above, referred to as alfalfa dwarf disease, were caused by coinfection with several other viruses, as shown below (Bejerman et al., 2011, 2015, 2016, 2018, 2019; Trucco et al., 2016). More recently, complete genomes of two novel AMV isolates from alfalfa plants were obtained in Australia (Samarfard et al., 2020) and China (Guo et al., 2019) employing enrichment of viral nucleic acids prior to HTS and siRNA sequencing, respectively.

Bean Leaf Roll Virus

Bean leaf roll virus (BLRV), a member of the genus *Luteovirus* (family *Luteoviridae*), is a phloem-limited virus that has been reported to infect a wide range of legume species worldwide, including alfalfa (Van Leur and Kumari, 2011). It was not diagnosed in Argentina in alfalfa or other crops prior to a study by Trucco et al. (2016). BLRV was isolated from alfalfa plants displaying symptoms of dwarf disease complex (**Figure 4**) that were coinfected with other viruses. HTS was employed to obtain the first complete nucleotide sequence of BLRV isolated from alfalfa as a natural host. Deep sequencing was performed on the Illumina HiSeq 2000 platform using siRNAs as a template. The complete genome of the BLRV isolate from Argentina (Manfredi BLRV-Arg) was highly identical to BLRV isolates infecting other legume hosts and presented a similar genome structure (Trucco et al., 2016). BLRV-Arg showed a prevalence of over 50% and a wide distribution in Argentinian alfalfa fields.

Alfalfa Latent Virus

Alfalfa latent virus (ALV) was first reported as a distinct species and a new member of the carlavirus group (Veerisetty and Brakke, 1977). It was subsequently recognized as a strain of *Pea streak mosaic virus*, genus *Carlavirus*, family *Betaflexiviridae* (Hampton, 1981). In the United States, pea streak mosaic virus (PeSV) is common in Nebraska and Wisconsin (Veerisetty, 1979). Although PeSV was first described in 1938 (Zaumeyer, 1938) and ALV was first described in 1977 (Veerisetty and Brakke, 1977), no complete genomic sequences of PeSV and its alfalfa latent strain were available prior to 2015. The first complete genomic sequence of the virus, determined using HTS and primer walking techniques, showed substantial differences from other members of the genus *Carlavirus* (Nemchinov et al., 2015). In addition to a low nucleotide identity to the most closely related species from the genus, *Shallot latent virus* (48.5%, PASC tool, NCBI[7]), the ALV genome did not appear to encode a typical-for-carlaviruses 3′ proximal, cysteine-rich protein of ∼11–16 kDa (Martelli et al., 2007). Instead, it encoded a hypothetical protein of ∼6 kDa that was different from the putative nucleic-acid binding proteins of known carlaviruses. A similar genome structure was later reported for another isolate of PeSV that was 77.9% identical to the alfalfa strain of the virus at the nucleotide level and originated from an unknown host (Su et al., 2015). To confirm that the genomic sequence of ALV was accurate, an infectious cDNA clone of the virus was constructed (Nemchinov, 2017). Rub inoculation of pea plants (*Pisum sativum*) with transcripts generated from the cDNA clone resulted in symptom development within 3 weeks after inoculation. The inoculated plants exhibited chlorotic leaves with noticeable mosaic patterns and necrotic zones along the leaf margins (**Figure 5**). The plants subsequently showed a decline, developing extensive necrosis and severe browning. When viral preparations purified from the

[2]http://www.medicagohapmap.org/downloads/cadl

[3]http://www.medicagogenome.org/

[4]https://www.ncbi.nlm.nih.gov/genome/viruses/

[5]https://sourceforge.net/projects/bbmap/

[6]http://www.dnastar.com/t-nextgen-seqman-ngen.aspx

[7]https://www.ncbi.nlm.nih.gov/sutils/pasc/viridty.cgi

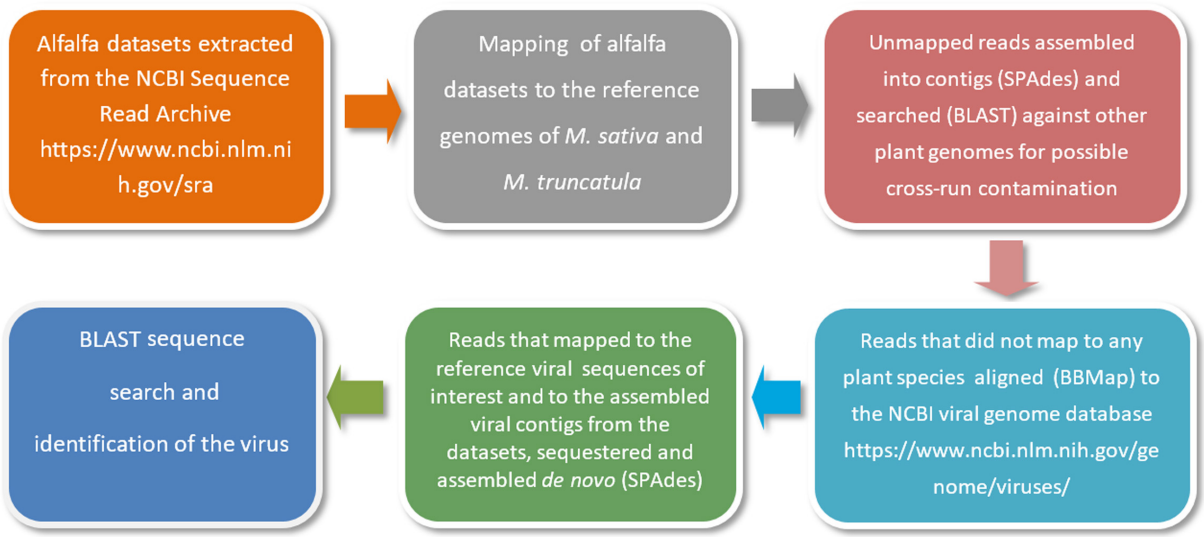


FIGURE 3 | A simplified illustration of the bioinformatics pipeline used for the identification of viral pathogens in alfalfa transcriptome datasets (Jiang et al., 2019a).

FIGURE 4 | Alfalfa plants exhibiting the alfalfa dwarf disease complex (shortened internodes, bushy appearance, chlorosis at the margins and ribs of the leaflets, leaf puckering and vein enations of varying sizes on abaxial leaf surfaces) (Bejerman et al., 2011).

transcript-inoculated pea plants were used to inoculate alfalfa, the alfalfa plants became infected, and the virus was detected in non-inoculated leaves by western blotting with PeSV antiserum, virus-specific RT-PCR and transmission electron microscopy (Nemchinov, 2017). Although ALV is asymptomatic in alfalfa, it can be transmitted mechanically or by aphids to other crops in

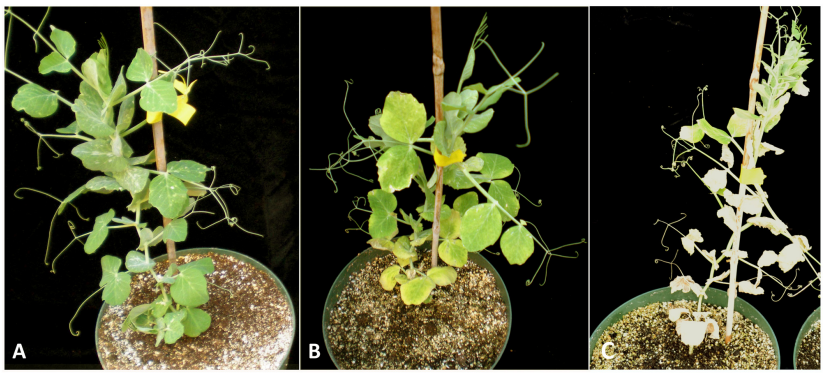


FIGURE 5 | Symptoms of alfalfa latent virus on pea plants (*Pisum sativum*) (Nemchinov, 2017). **(A)** Control uninfected plant. **(B)** Plants infected with ALV showing symptoms of a chlorotic mosaic pattern and necrotic zones along the leaf margins. **(C)** Plants infected with ALV developed extensive necrosis and severe browning (Nemchinov, 2017). License to publish this figure obtained from the publisher (license #4897140253092).

which it produces symptoms (Veerisetty and Brakke, 1977). This makes alfalfa a natural reservoir of the virus.

Novel Alfalfa Viruses Identified and Characterized by HTS

Alfalfa Dwarf Virus

In 2011, a fragment of the polymerase gene of a cytorhabdovirus was amplified from alfalfa samples showing alfalfa dwarf disease symptomatology by RT-PCR assays with degenerate primers for conserved regions of plant rhabdovirus polymerase (L) genes (**Figure 4**) (Bejerman et al., 2011). The pathogen, designated alfalfa dwarf virus (ADV), was the first cytorhabdovirus reported to infect alfalfa (Bejerman et al., 2011). Since most of the genes of plant rhabdoviruses described so far have been highly divergent, the application of HTS was essential to elucidate the genomic sequence of ADV. Using siRNA as a template to build libraries to be sequenced on the Illumina HiSeq 2000 platform, the complete genomic sequence of ADV was obtained (Bejerman et al., 2015). The genome of ADV encoded six proteins characteristic of all cytorhabdoviruses as well as one accessory protein, which was also described in ADV-related viruses (Bejerman et al., 2015). The subcellular localization of each ADV-encoded protein, determined by means of transient expression as fusions with green fluorescent protein in the leaves of *Nicotiana benthamiana*, showed that ADV is an unusual rhabdovirus that combines the properties of both cytoplasmic and nuclear rhabdoviruses (Bejerman et al., 2015). Interestingly, the HTS of small RNAs from alfalfa samples collected in Henan Province, China, displaying symptoms such as dwarfism, shrinkage and mottle mosaic revealed fragments of the ADV genome (Guo et al., 2019). Moreover, an analysis of the transcriptome generated in this work, which is publicly available from the NCBI (SRA057663), resulted in the assembly of a novel ADV genome (Humberto Debat, personal communication), thus confirming that this virus was present in China. Furthermore, a contig related to ADV was identified by Jiang et al. (2019a) in the publicly available SRR7252502 alfalfa transcriptome dataset, which was deposited at the NCBI by researchers from Beijing Forestry University in China. Therefore, the use of HTS was critical to determine the complete genome sequence of ADV; it also raised questions about the evolution and the extant geographic distribution of the virus.

Alfalfa Enamovirus

The deep sequencing of small RNAs isolated from alfalfa samples showing dwarfism disease (**Figure 4**) also resulted in the identification of a novel enamovirus, designated alfalfa enamovirus 1 (AEV-1) (Bejerman et al., 2016). AEV-1 was the first enamovirus reported to infect alfalfa. It has a genomic structure characteristic of other known legume-infecting enamoviruses and is phylogenetically related to them (Bejerman et al., 2016; Debat and Bejerman, 2019).

In 2017, the HTS of total RNAs extracted from alfalfa plants collected in Sudan resulted in the assembly of the complete genome of a novel AEV-1 strain, which was designated alfalfa enation virus 2 (AEV-2) (Nemchinov et al., 2017b). At the nucleotide level, AEV-2 was 95.3% identical to AEV-1 from Argentina; its amino acid identity to AEV-1 varied from 94 to 98% for different viral proteins. Phylogenetic analyses of the predicted RNA-dependent RNA polymerase (RdRp) amino acid sequences and the complete nucleotide sequences of AEV-2 and other members of the family *Luteoviridae* clustered AEV-2 and AEV-1 together (Nemchinov et al., 2017b). Although the exact origin of the Sudanese isolate of alfalfa enamovirus is not known, it is possible that its evolution and dissemination into new areas are related to the geography of the host. The first occurrence of the virus outside of Argentina indicated that it might be widespread and can potentially emerge in Iran and southwestern Asia, the geographic origin of alfalfa (Brough et al., 1977), as well as in other alfalfa cultivation regions worldwide.

Alfalfa Leaf Curl Virus

Using a virion-associated nucleic acid (VANA)-based metagenomics approach described above (Candresse et al., 2014), a novel single-stranded DNA (ssDNA) virus designated alfalfa leaf curl virus (ALCV) was identified in alfalfa samples (Roumagnac et al., 2015). ALCV was discovered in plants collected in Southern France that exhibited leaf curl symptoms

(**Figure 6**). The species *Alfalfa leaf curl virus* belongs to the recently designated genus *Capulavirus* within the *Geminiviridae* family. It is unique among geminiviruses in that it is aphid transmitted (Roumagnac et al., 2015). More recently, the siRNA-based HTS of alfalfa samples displaying symptoms of dwarfism disease in Argentina and symptoms of dwarfism, shrinkage and mottle mosaic in China resulted in the identification of novel ALCV isolates (Bejerman et al., 2018; Guo et al., 2019) belonging to ALCV genotype D, which is divergent from three other genotypes (A, B, and C) found in many European countries, Northern Africa and the Middle East (Davoodi et al., 2018). Thus far, one hundred and twenty complete ALCV genome sequences have been recovered from ten countries, and four ALCV genotypes (ALCV-A, ALCV-B, ALCV-C, and ALCV-D) have been clearly distinguished (Davoodi et al., 2018). The identification of these isolates expanded the known geographical range of ALCV and shed more light on the distribution of this emergent alfalfa virus. ALCV isolates were found to be highly recombinogenic and it was suggested that recombination has been a determining factor in the origin of the different viral genotypes. The ALCV sequence data support the hypothesis that the virus likely emerged and diversified in the Middle East before spreading to the western Mediterranean basin and Argentina (Davoodi et al., 2018). The international research conducted on ALCV is a good example highlighting the need to combine the power of HTS for the detection and identification of poorly known viruses with the reliability of classical molecular methodology to obtain complete genomes of novel viruses and understand their evolution (Bernardo et al., 2016).

Alfalfa Virus S

Alfalfa virus S was discovered in alfalfa samples received from Sudan, Northern Africa, where commercial pivot-irrigated fields were planted with alfalfa seeds originating from the United States (Nemchinov et al., 2017a). Although the plants exhibited chlorosis and stunting in the field, upon arrival at the laboratory, the samples had deteriorated and showed no visual signs of symptoms. Transmission electron microscopy (TEM) observations of the infected tissues revealed the presence of filamentous virions similar to allexiviruses in their length and appearance (**Figure 7**). The samples were subjected to a standard HTS protocol to detect all viruses that were potentially present in an unbiased manner. Several coding-complete viral genomes were identified in the sequenced sample, including a novel flexivirus with the highest bit score for shallot virus X (ShVX), a virus with ∼98% identity to peanut stunt virus (PSV, genus *Cucumovirus*, family *Bromoviridae*), and a virus with 90–97% identity to alfalfa enamovirus-1 (AEV-1, tentative member of the *Luteoviridae* family) (Bejerman et al., 2016). While PSV and AEV-1 had been known to infect alfalfa, a novel flexivirus, which we referred to as alfalfa virus S, for Sudan (AVS), represented a previously undescribed species. A complete nucleotide sequence of the viral genome consisting of 8,349 nucleotides was obtained by the *de novo* assembly of the HTS-generated reads, supplemented with 5′RACE and the sequencing of the RT-PCR-amplified 3′ terminus (Nemchinov et al., 2017a). At the nucleotide level, AVS was most similar to *Arachis pintoi virus*, *Blackberry virus E* (BVE), and the type member of the *Allexivirus* genus, ShVX. Phylogenetic analyses grouped AVS together with *Arachis pintoi virus* and BVE in a distinct cluster associated with known allexiviruses (Nemchinov et al., 2017a). Similarly to *Arachis pintoi virus* and BVE, the assembled genome of AVS did not appear to have sequences homologous to the 3′ proximal nucleic acid-binding protein of allexiviruses. Currently, *Alfalfa virus S* is recognized by the ICTV as a species in the genus *Allexivirus*.

Recently, two more isolates of the AVS were identified by HTS in alfalfa plants: an isolate from China (GenBank ID MN864567) and an isolate from the USA (GenBank ID MT094142). These results suggest that AVS is more widespread than originally thought. Apart from the fact that AVS often appears to be

FIGURE 6 | Symptoms of alfalfa leaf curl virus on alfalfa (Roumagnac et al., 2015).

FIGURE 7 | Filamentous virions observed by transmission electron microscopy in alfalfa tissues infected with alfalfa virus S (Nemchinov et al., 2017a).

associated with other pathogens, the economic importance of this virus is largely unknown.

Alfalfa Virus F

This novel virus, provisionally designated alfalfa virus F (AVF), was identified using a VANA metagenomics-based approach in symptomless alfalfa samples collected in Southern France (Nemchinov et al., 2018a). The distribution of AVF is currently unknown, and it is unlikely to be restricted to a single area. In accordance with the current species demarcation criteria, the virus represented a distinct species in genus *Marafivirus*, family *Tymoviridae*. Until 2018, marafiviruses were not known to infect alfalfa. The virus shared the highest degree of sequence identity (∼78%) with medicago sativa marafivirus 1 (MsMV1), which was computationally deduced from alfalfa transcriptomic datasets (Kim et al., 2018). Phylogenetic analysis of the complete nucleotide sequences of AVF and other viruses of the *Tymoviridae* family clustered AVF together with MsMV1 in a sister group connected to the ICTV-proposed marafiviruses: grapevine rupestris vein feathering virus and grapevine Syrah virus-1, thus supporting the classification of AFV as a new member of the *Marafivirus* genus (Nemchinov et al., 2018a).

Alfalfa Ringspot-Associated Virus

A novel putative emaravirus, provisionally named alfalfa ringspot-associated virus (ARaV), was recently discovered in Australia (Samarfard et al., 2020) using a dsRNA immunocapture technique (Blouin et al., 2016). Emaraviruses (family *Fimoviridae*, order *Bunyavirales*) have negative-sense, linear, segmented RNA genomes. Only the partial RNA1, 3 and 4 of ARaV, which shared 56–96% amino acid identity with the RdRP, nucleocapsid, and movement (MP) proteins (RNA1, RNA3, and RNA4, respectively) of several emaraviruses, were assembled in the study (Samarfard et al., 2020).

Medicago Sativa Alphapartitiviruses

While examining HTS-derived transcriptomic data from two U.S. alfalfa cultivars, cv. Maverick and cv. ZG 9830, it was found that all plants used in the experiment (n = 36) contained short reads related to alphapartitiviruses (Nemchinov et al., 2018b). Members of the genus *Alphapartitivirus* infect either plants or ascomycetous and basidiomycetous fungi (Vainio et al., 2018). In plants, partitiviruses cause persistent infections, remaining with their hosts for many generations and having no visible effects on their hosts (Vainio et al., 2018). Plant partitiviruses are transmitted by ovules and pollen to the seed embryo (Boccardo et al., 1987) and are assumed to be mutualistic (Roossinck, 2015). Complete viral genomes were obtained from both cultivars by the assembly of HTS-generated paired-end reads and 5′/3′ rapid amplification of cDNA ends (RACE). The genomes were characteristic of the genus *Alphapartitivirus* and contained two monocistronic segments: double-stranded RNA1 (dsRNA1), encoding an RdRP, and dsRNA2, encoding a viral coat protein (CP). The study was the first to demonstrate that alfalfa cultivars in the United States could be frequently infected with a seed-transmitted cryptic virus of the *Alphapartitivirus* genus.

Alphapartitiviruses were also diagnosed in alfalfa plants with dwarfism disease symptoms in Argentina (Bejerman et al., 2019). HTS of small RNAs isolated from these alfalfa samples led to the identification of two alphapartitiviruses, which were designated medicago sativa alphapartitivirus 1 (MsAPV1) and medicago sativa alphapartitivirus 2 (MsAPV2) (Bejerman et al., 2019). Characterization of the MsAPV1 discovered in this study resulted in the redefinition of the previously reported MsAPV1 genome (Kim et al., 2018), which was likely reconstructed from mixed genomic segments of two different alphapartitiviruses (Bejerman et al., 2019). The MsAPV2 represented a new member of the genus *Alphapartitivirus*, based on the low identity of its CP and RdRp with those of MsAPV1. Thereafter, MsAPV1 was also identified in alfalfa plants originating from Australia (Samarfard et al., 2020). The biological significance and any negative effects of the partitiviruses on alfalfa are currently unknown and require further investigation.

Alfalfa-Associated Nucleorhabdovirus

To determine the genome sequence of a rhabdovirus-like pathogen found by electron microscopy in alfalfa samples from Austria, Gaafar et al. (2019) performed high-throughput sequencing followed by RT-PCR to confirm virus infection. HTS was performed on the Illumina MiSeq platform using ribosomal RNA-depleted total RNA as a template. As a result, the authors were able to identify and characterize a new nucleorhabdovirus that shared 39.8% nucleotide sequence identity with its closest known relative, black currant-associated rhabdovirus 1 (Gaafar et al., 2019). This alfalfa-associated nucleorhabdovirus (AaNV) exhibited a unique genomic organization and encoded a new accessory ORF (U) with unknown function located between the matrix and glycoprotein-encoding genes, in addition to the six main nucleorhabdovirus proteins (N, P, P3, M, G, and L) (Gaafar et al., 2019). According to the species demarcation criteria set by the ICTV, alfalfa-associated nucleorhabdovirus (AaNV) represents a new species (Gaafar et al., 2019). AaNV is the first alfalfa-infecting nucleorhabdovirus whose sequence has been characterized. A putative nucleorhabdovirus known as lucerne enation virus (LEV) was previously characterized biologically (Alliot and Signoret, 1972; Leclant et al., 1973). According to its phylogenetic relationships, the AaNV vector could be an aphid, and LEV is likely transmitted by the aphid *Aphis craccivora* (Alliot and Signoret, 1972; Leclant et al., 1973). Therefore, it would be interesting to apply HTS for the characterization of the molecular properties of LEV to determine whether AaNV and LEV represent the same virus species.

ALFALFA VIRUSES FOUND IN PUBLIC REPOSITORIES

Medicago Sativa Alphapartitivirus 1, Medicago Sativa Deltapartitivirus 1, and Medicago Sativa Marafivirus 1

Kim et al. (2018) analyzed an alfalfa transcriptome dataset downloaded from the NCBI Sequence Read Archive

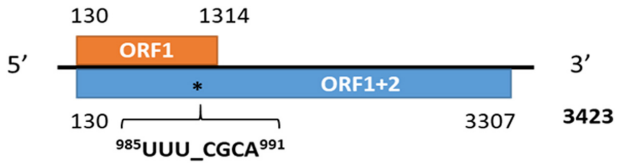


FIGURE 8 | The genome organization of medicago sativa amalgavirus 1 and the putative + 1 programmed ribosomal frameshifting motif in MsAV1 (Jiang et al., 2019a).

(SRA057663) and identified the genome sequences of three new RNA viruses designated medicago sativa alphapartitivirus 1 (MsAPV1), medicago sativa deltapartitivirus 1(MsDPV1), and medicago sativa marafivirus 1 (MsMV1, belonging to the genera *Alphapartitivirus* and *Deltapartitivirus* of the family *Partitiviridae* and the genus *Marafivirus* of the family *Tymoviridae*, respectively. In this study, high-quality sequence reads were collected using the Sickle program, followed by their assembly into contigs using SPAdes software (Kim et al., 2018). For the identification of the viral sequences, the authors first searched the assembled contigs with BLASTX against a custom database of non-redundant RdRp motifs and then used matched candidates as a query in the NCBI BLASTX search. The RdRp of MsAPV1, MsDPV1, and MsMV1 shared only 68, 58, and 46% amino acid sequence identity with the closest virus species, respectively. The authors concluded that the protocol devised in the study may facilitate the identification of new persistent plant RNA viruses from various plant transcriptome data (Kim et al., 2018). However, as emphasized by Bejerman et al. (2019), who also found MsAPV1 as well as a new alphapartitivirus MsAPV2 in alfalfa HTS repositories, it is critical to pay special attention to the assembly of multisegmented viral genomes when they are obtained from publicly available transcriptomes. The biological significance and any negative effects of the partitiviruses infecting alfalfa are currently unknown and require further investigation.

Alfalfa Isolate of Cycas Necrotic Stunt Virus

Cycas necrotic stunt virus, a member of the genus *Nepovirus*, family *Secoviridae*, was first identified in the gymnosperm *Cycas revoluta* in Japan (Kusunoki et al., 1986; Han et al., 2002). The severely affected plants deteriorated and subsequently died (Kusunoki et al., 1986). Cycas necrotic stunt virus (CNSV) and similar viruses have also been isolated from gladiolus (*Gladiolus* spp.), peony (*Paeonia lactiflora* Pall.), Easter lily (*Lilium longiflorum*), aucuba (*Aucuba japonica*), daphne (*Daphne odora*) and spring onions (*Allium fistlosum*) (Kusunoki et al., 1986; Han et al., 2002; Ochoa-Corona et al., 2003; Hanada et al., 2006; Wylie et al., 2012; Lim et al., 2019; Shaffer et al., 2019). Until recently, CNSV had not been found in alfalfa. CNSV sequences were identified in three NCBI accessions/datasets (SRR7751381, SRR7751384 and SRR7751386) generated from alfalfa plants by third parties using the Illumina HiSeq platform (Dong et al., 2018; Jiang et al., 2019b). The assembled virus had a bipartite (RNA1 and RNA2 segments) single-stranded positive-sense RNA genome that appeared to be coding-complete. Polyproteins P1 and P2, encoded by RNA1 and RNA2 segments, showed 94.3% and 91.3% amino acid identity to the respective polyprotein of the reference CNSV sequence (Jiang et al., 2019b). Phylogenic analyses grouped the alfalfa strain together with CNSV isolated from other species, indicating their origin from the same ancestral virus (Jiang et al., 2019b). Jiang et al. (2019b) concluded that the virus represented a new strain of CNSV adapted to alfalfa, for which the name CNSV-A was proposed. Further research is underway to verify the *in silico* identification of the virus and assess its symptomatology, geographic distribution and economic importance to the alfalfa industry.

Medicago Sativa Amalgavirus 1

Amalgaviruses are members of the recently established *Amalgaviridae* family that have monopartite double-stranded RNA genomes and encode two proteins: RdRp and CP (Sabanadzovic et al., 2009; Martin et al., 2011; Krupovic et al., 2015). The medicago sativa amalgavirus 1 (MsAV1) sequence was initially communicated by Wang and Zhang in 2013 under GenBank accession number GAFF01077243.1 and subsequently analyzed by Nibert et al. (2016). The virus was not found in the U.S. prior to the study by Jiang et al. (2019a), and its sequence had not been validated experimentally.

The RNA-seq data in which the virus reads were identified originated from the publicly available datasets SRR6050922 to SRR6050957 generated from the U.S. alfalfa cultivars Maverick and ZG9830 (Nemchinov et al., 2017c). The subject of the original study was not related to virology research, and these datasets were evaluated a second time as part of an effort to identify emerging viral genomes in publicly available alfalfa transcriptomic repositories. Among the 36 screened alfalfa datasets, half included MsAV1 reads. The raw viral reads were mapped to the reference genome of MsAV1 (GAFF01077243.1; NC_040591.1) (Zhang et al., 2015; Nibert et al., 2016) and assembled into a complete viral genome (**Figure 8**). The U.S. isolate of MsAV1 was found to be 100% identical to the GAFF01077243.1/NC040591 isolate from China (Zhang et al., 2015; Nibert et al., 2016) at both the nucleotide and amino acid levels, indicating the same origin of the virus. It is likely that the alfalfa strain of the virus originated in the U.S., since cv. Maverick, used in the study by Zhang et al. (2015), was introduced to China from the United States. In 2020, MsAV1 was also identified in alfalfa plants from Australia (Samarfard et al., 2020). Although the economic significance of amalgaviruses is currently unknown, and with few exceptions, they do not cause any symptoms, amalgaviruses are vertically transmitted through seeds and are persistent in plants. It has been suggested that persistent viruses may represent cytoplasmic epigenetic elements that provide a selective advantage and genetic information to their hosts (Roossinck, 2012a).

Alfalfa Isolate of Cnidium Vein Yellowing Virus

Cnidium vein yellowing virus (CnVYV) is a bipartite, linear, positive-sense ssRNA virus that is a tentative member of family

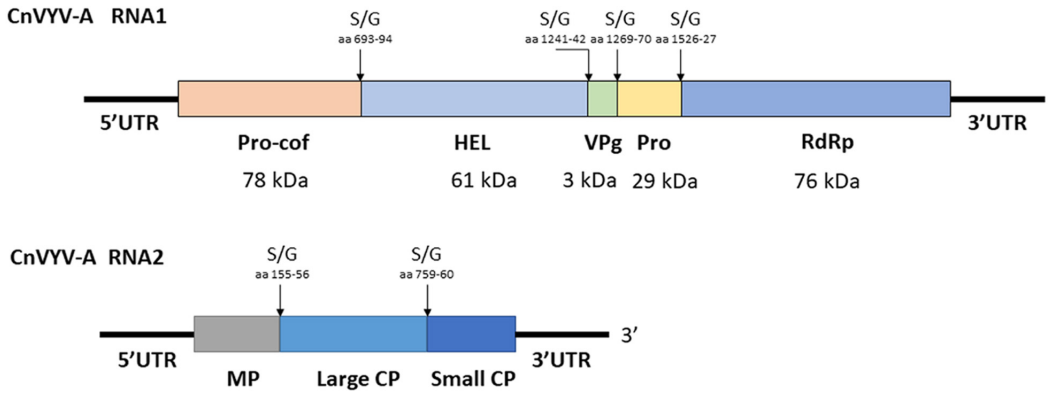


FIGURE 9 | Putative genomic organization of the alfalfa strains of cnidium vein yellowing virus (CnVYV-A). The open reading frames are indicated by boxes, and the putative serine/glycine (S/G) cleavage sites and their amino acid positions are indicated by arrows (Jiang et al., 2019a).

A

CnVYV-A1	100%							
CnVYV-A2	98.23%	100%						
CnVYV-1_AKN59243	91.15%	90.7%	100%					
CnVYV-2_AKN59245	91.15%	91.59%	94.69%	100%				
LycMoV-A	98.45%	98.67%	90.7%	91.15%	100%			
LycMoV_AKN59247	92.03%	92.92%	94.91%	97.34%	92.47%	100%		
LycMoV-J_LC382243	91.15%	91.59%	95.57%	94.46%	91.15%	95.13%	100%	
SLRSV_YP227367	91.15%	91.15%	90.7%	92.03%	90.92%	92.03%	91.37%	100%
	CnVYV-A1	**CnVYV-A2**	**CnVYV-1_AKN59243**	**CnVYV-2_AKN59245**	**LycMoV-A**	**LycMoV_AKN59247**	**LycMoV-J_LC382243**	**SLRSV_YP227367**

B

CnVYV-A1	100%							
CvVYV-A2	96.94%	100%						
CnVYV-1_AKN59244	89.31%	89.05%	100%					
CnVYV-2_AKN59246	92.36%	91.09%	91.09%	100%				
LycMoV-A	98.47%	98.47%	90.07%	92.62%	100%			
LycMoV_AKN59248	92.36%	91.6%	92.62%	93.63%	92.62%	100%		
LycMoV-J_BBE07891	89.82%	89.56%	95.92%	92.62%	90.58%	94.4%	100%	
SLRSV_YP227375	6.7%	7.21%	7.21%	7.73%	6.95%	7.47%	7.47%	100%
	CnVYV-A1	**CvVYV-A2**	**CnVYV-1_AKN59244**	**CnVYV-2_AKN59246**	**LycMoV-A**	**LycMoV_AKN59248**	**LycMoV-J_BBE07891**	**SLRSV_YP227375**

FIGURE 10 | Amino acid identities between the Pro-Pol **(A)** and CP **(B)** regions of the cnidium vein yellowing virus CnVYV-A strains, the CnVYV1 and CnVYV2 strains, lychnis mottle virus LycMoV-A, LycMoV, and LycMoV-J strains and strawberry latent ringspot virus (SLRSV), as predicted by the SIAS tool (http://imed.med.ucm.es/Tools/sias.html) (Jiang et al., 2019a).

Secoviridae, order *Picornavirales* (Yoo et al., 2015a). Two isolates of the virus, CnVYV-1 and CnVYV-2, were previously found to infect cnidium plants (*Cnidium officinale*) in Korea (Yoo et al., 2015a) and no other hosts of CnVYV have been reported. Presently, the virus is not listed by the ICTV as either an established or unassigned species (Thompson et al., 2017). In the study by Jiang et al. (2019a), the datasets were retrieved from the SRA accessions SRR2089795 and SRR2089796 of the BioProject PRJNA289195 (Song et al., 2016) and mapped to the genome of CnVYV (GenBank accession numbers: KR011028, KR011029, KR011030, and KR011031). Both datasets were scanned for possible cross-run contamination, and no plant species other than *Medicago* spp. were detected. The virus reads from both accessions were assembled into coding-complete bipartite genomes consisting of RNA1 and RNA2 segments from two different strains of CnVYV-related viruses. The virus strains were provisionally named CnVYV-A1 and CnVYV-A2 (Jiang et al., 2019a). The RNA1 and RNA2 sequences of the CnVYV-A strains translated into two polyproteins, P1 and P2, which exhibited ∼78–79% (P1) and ∼90–92% (P2) identity to the respective polyproteins of the reference genomes (Jiang et al., 2019a). Based on the predictions made by the Pfam, InterPro and SIAS tools, it was anticipated that CnVYV-A strains share a similar genome organization (**Figure 9**) with other closely related viruses of the family *Secoviridae* and represent isolates of the same virus strain adapted to alfalfa, for which the name CnVYV-A (alfalfa) was proposed.

Alfalfa Isolate of Lychnis Mottle Virus

Lychnis mottle virus (LycMoV) is a tentative member of the family *Secoviridae* that was first described in *Lychnis cognata*, a flowering plant in the family *Caryophyllaceae* (Yoo et al.,

2015b). In 2017, the virus was also isolated from the leaves of *Vincetoxicum acuminatum* in Japan, and the complete nucleotide sequence of LycMoV-J was reported (Fujimoto et al., 2018). While performing a survey of alfalfa transcriptome datasets available at NCBI, transcripts that mapped to the genome of the LycMoV isolate Andong (KR011032 and KR011033) under accession number SRR2089796 (Song et al., 2016) were found (Jiang et al., 2019a). The assembled coding-complete genome of the alfalfa isolate of LycMoV (LycMoV-A) consisted of two segments, corresponding to RNA1 and RNA2 (Jiang et al., 2019a). Putative polyproteins P1 and P2 of LycMoV-A exhibited the top BLAST hits to CnVYV-1 (78.7%) (Yoo et al., 2015b) and LycMoV-J (90.9%) isolates, (Fujimoto et al., 2018), suggesting that CnVYV and LycMoV may potentially belong to the same viral species. To further assess whether LycMoV-A and CnVYV-A are strains of the same virus species or represent different species, as suggested by Yoo et al. (2015a; 2015b) for the Yeongyang isolates of CnVYV from *Cnidium officinale* and the Andong isolate of LycMoV from *Lychnis cognata*, the amino acid identities between their Pro-Pol and CP regions were compared, which are currently used as species demarcation criteria by the ICTV. The SIAS tool[8] predicted that all the conserved Pro-Pol values were higher than 80% (the ICTV criterion for species demarcation is less than 80% identity), and all the CP values except for that of SLRSV were higher than 75% (the ICTV criterion for species demarcation is less than 75% identity, **Figure 10**). The authors speculated that CnVYV-A and all the other viruses in the study group (CnVYV-1, CnVYV-2, LycMoV-A, LycMoV and LycMoV-J) with the exception of SLRSV due to the low identity of its CP region, represent individual strains of the same species isolated from different hosts, for which a common name reflecting the taxonomic position and biological characteristics of the species is needed (Jiang et al., 2019a). The following provisional name was suggested for the species: cnidium vein yellowing-like virus (CnVYLV). Phylogenetic analyses of the Pro-Pol region of these viruses between the protease CG motif and the RdRp GDD motif (CG/GDD) in RNA1 fully supported this conclusion (Jiang et al., 2019a).

[8]http://imed.med.ucm.es/Tools/sias.html

Cactus Virus X

Cactus virus X is a member of genus *Potexvirus* (family *Alphaflexiviridae*) that infects various species in the *Cactaceae* plant family worldwide (Koenig, 1987). The reference sequence for cactus virus X (CVX) was reported under two identical accessions, NC_002815.2 and AF308158 (Liou et al., 2004). Prior to the work of Jiang et al. (2019a), CVX had not been identified in alfalfa. CVX-related reads were detected in alfalfa transcriptome dataset SRR7751381, BioProject PRJNA487676 (Dong et al., 2018). The raw reads were assembled *de novo* into a coding-complete genome consisting of a single molecule of linear ssRNA of 6,603 nucleotides in length, excluding the poly(A) tail. At the nucleotide level, the alfalfa isolate of CVX (CVX-A) was 97% identical to the reference sequence (AF308158.2).

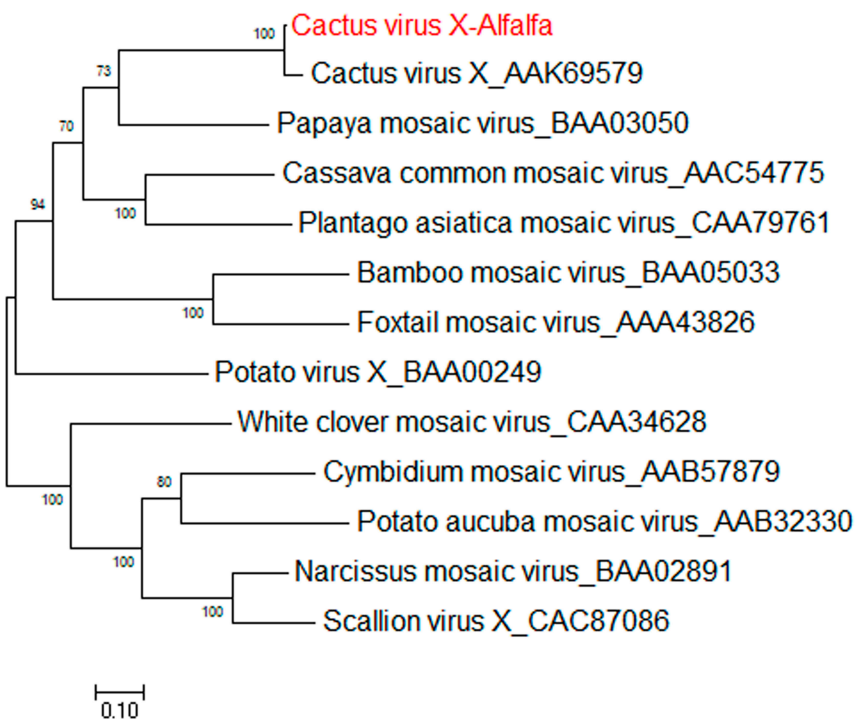


FIGURE 11 | Phylogenetic analyses based on the amino acid alignments of the predicted RdRp sequence of CVX-A and other members of the *Potexviridae* family. The trees were generated using the Maximum Likelihood method of MEGA7 with 1000 bootstrap replicates (Jiang et al., 2019a).

TABLE 1 | Viruses identified in alfalfa by HTS and by analysis of publicly available transcriptome datasets.

Virus name	Genus	Genome	References
Alfalfa mosaic virus	Alfamovirus	ssRNA+	Trucco et al., 2014; Guo et al., 2019; Jiang et al., 2019b; Samarfard et al., 2020
Bean leafroll virus	Luteovirus	ssRNA+	Trucco et al., 2016; Jiang et al., 2019b; Samarfard et al., 2020
Alfalfa dwarf virus	Cytorhabdovirus	ssRNA−	Bejerman et al., 2015; Guo et al., 2019; Jiang et al., 2019b
Alfalfa enamovirus	Enamovirus	ssRNA+	Bejerman et al., 2016; Nemchinov et al., 2017b
Alfalfa leaf curl virus	Capulavirus	ssDNA+	Roumagnac et al., 2015; Bejerman et al., 2018; Guo et al., 2019; Jiang et al., 2019a
Medicago sativa alphapartitivirus 1	Alphapartitivirus	dsRNA+	Kim et al., 2018; Nemchinov et al., 2018b; Bejerman et al., 2019; Samarfard et al., 2020
Medicago sativa alphapartitivirus 2	Alphapartitivirus	dsRNA+	Bejerman et al., 2019; Jiang et al., 2019b
Alfalfa-associated nucleorhabdovirus	Nucleorhabdovirus	ssRNA−	Gaafar et al., 2019
Alfalfa virus S	Allexivirus	ssRNA+	Nemchinov et al., 2017a
Peanut stunt virus	Cucumovirus	ssRNA+	Nemchinov et al., 2017a
Alfalfa virus F	Marafivirus	ssRNA+	Nemchinov et al., 2018a
Medicago sativa marafivirus 1	Marafivirus	ssRNA+	Kim et al., 2018; Nemchinov et al., 2018a
Medicago sativa deltapartitivirus 1	Deltapartitivirus	dsRNA+	Kim et al., 2018; Jiang et al., 2019a
Medicago sativa amalgavirus 1	Amalgavirus	dsRNA+	Nibert et al., 2016; Jiang et al., 2019a; Samarfard et al., 2020
Alfalfa ringspot associated virus	Emaravirus	ssRNA−	Samarfard et al., 2020
Cnidium vein yellowing virus	Unassigned	ssRNA+	Jiang et al., 2019a
Lychnis mottle virus	Unassigned	ssRNA+	Jiang et al., 2019a
Pea streak virus	Carlavirus	ssRNA+	Nemchinov et al., 2015; Jiang et al., 2019a
Cactus virus X	Potexvirus	ssRNA+	Jiang et al., 2019a
Zhuye pepper nepovirus	Nepovirus	ssRNA+	Jiang et al., 2019a
Cowpea mild mottle virus	Carlavirus	ssRNA+	Jiang et al., 2019a
Strawberry latent ringspot virus	Unassigned	ssRNA+	Jiang et al., 2019a
Cycas necrotic stunt virus	Nepovirus	ssRNA+	Jiang et al., 2019b

The CVX-A RNA is translated into five putative open reading frames (ORF) encoding RdRp, triple gene block proteins 1, 2, and 3 and a putative coat protein (Jiang et al., 2019a). It was therefore anticipated that the genome organization of CVX-A is similar to that of CVX (Liou et al., 2004). BLASTP analysis of the amino acid identities of the putative CVX-A proteins with the corresponding proteins of the reference genome indicated a close relationship. Phylogenetic analysis performed with the amino acid sequences of CVX-A RdRp and CP placed CVX-A in the same subcluster with CVX (**Figure 11**) (Jiang et al., 2019a). Accordingly, it was concluded that CVX-A represents a strain of CVX adapted to alfalfa.

Overall, a systematic survey of more than 600 publicly available alfalfa transcriptome datasets conducted by Jiang et al. (2019a) indicated that approximately 90% of the plant samples employed in the generation of the deposited datasets contained viruses. This analysis identified 23 different viruses, including several emerging viral pathogens that had not been previously reported or experimentally confirmed in *M. sativa*, such as two strains of cnidium vein yellowing virus, lychnis mottle virus and *Cactus virus X*, for which coding-complete genomic sequences were obtained by *de novo* and reference-based assembly (Jiang et al., 2019a). Although further research is needed to confirm the *in silico* identification of these viruses and to specify their symptomatology, geographic distribution and economic importance to the alfalfa industry, the transcriptomic survey improved the knowledge of the host range and diversity and of the viruses infecting alfalfa and provided essential tools for their diagnosis and characterization.

CONCLUDING REMARKS

It is thus becoming increasingly obvious that in research on alfalfa virology, similar to research on the virology of any other plant species or agricultural crop, HTS technologies and their derivatives, such as the exploration of public transcriptomic datasets, are making a major contribution toward the discovery of novel viral genomes, the sequences of emerging pathogens transitioning to new host species and the detection of known viruses. The employment of HTS in the field of alfalfa virology has not only empowered and significantly deepened the understanding of the virome of this strategic legume crop but has also increased the understanding of the geographical range of emergent viruses such as alfalfa dwarf virus, alfalfa enamovirus, alfalfa leaf curl virus, and alfalfa virus S (**Table 1**). Meanwhile, the virome of alfalfa plants from many other geographic locations, other than those described in this review (Western Europe, United States, China, Australia, and Argentina) remains to be characterized.

Needless to say, further disregarding the role of viruses in alfalfa health could be unwise and impractical for alfalfa improvement and production. This is especially true for polymicrobial infections of alfalfa, in which viruses may

constitute a substantial and thus far unrecognized part of a disease complex or be solely responsible for coinfections consisting of multiple viral pathogens, as appears to be the case for alfalfa dwarfism disease. Alfalfa may also serve as a natural reservoir for the dissemination of viruses to other agriculturally important crops, although its exact role in the epidemiology of viruses in other crops is not well documented (Babadoost, 1990; Frate and Davis, 2008; Van Leur and Kumari, 2011). Thus, the characterization of the virome of asymptomatic alfalfa plants as well as other *Medicago* species will be crucial to unravel their role as natural viral reservoirs.

Taken together, the continuous discoveries of new viruses in alfalfa made by HTS and associated technologies have called into question the assumed low economic impact of viral diseases in alfalfa and further suggested their potential contribution to the severity of complex infections involving multiple pathogens.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Publication of this review was supported by the United States Department of Agriculture, Agricultural Research Service, CRIS number 8042-21000-300-00-D and the Agropolis Fondation (Labex Agro – Montpellier, E-SPACE project number 1504-004).

REFERENCES

Abdullah, A. S., Moffat, C. S., Lopez-Ruiz, F. J., Gibberd, M. R., Hamblin, J., and Zerihun, A. (2017). Host-multi-pathogen warfare: pathogen interactions in co-infected plants. *Front. Plant Sci.* 8:1806. doi: 10.3389/fpls.2017.01806

Alizon, S., de Roode, J. C., and Michalakis, Y. (2013). Multiple infections and the evolution of virulence. *Ecol. Lett.* 16, 556–567. doi: 10.1111/ele.12076

Alliot, B., and Signoret, P. A. (1972). La maladie a enations de la lucerne, une maladie nouvelle pour la France. *Phytopathol. Z* 74:69. doi: 10.1111/j.1439-0434.1972.tb04647.x

Babadoost, M. (1990). *Virus Diseases of Alfalfa and Clovers in Illinois*. Reports on Plant Diseases, RPD No. 307, Champaign: University of Illinois at Urbana-Champaign.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, L., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barba, M., Czosnek, H., and Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106–136. doi: 10.3390/v6010106

Basigalup, D. H., and Ustarroz, E. (2007). "Grazing alfalfa systems in the argentinean pampas," in *Proceedings of the 37th California Alfalfa & Forage Symposium*, Monterey, CA.

Bejerman, N., Debat, H., Nome, C., Cabrera-Mederos, D., Trucco, V., de Breuil, S., et al. (2019). Redefining the *Medicago sativa* alphapartitiviruses genome sequences. *Virus Res.* 265, 156–161. doi: 10.1016/j.virusres.2019.03.021

Bejerman, N., Giolitti, F., de Breuil, S., Trucco, V., Nome, C., Lenardon, S., et al. (2015). Complete genome sequence and integrated protein localization and interaction map for alfalfa dwarf virus, which combines properties of both cytoplasmic and nuclear plant rhabdoviruses. *Virology* 483, 275–283. doi: 10.1016/j.virol.2015.05.001

Bejerman, N., Giolitti, F., Trucco, V., de Breuil, S., Dietzgen, R. G., and Lenardon, S. (2016). Complete genome sequence of a new enamovirus from argentina infecting alfalfa plants showing dwarfism symptoms. *Arch. Virol.* 161, 2029–2032. doi: 10.1007/s00705-016-2854-3

Bejerman, N., Nome, C., Giolitti, F., Kitajima, E., de Breuil, S., Fernandez, J. P., et al. (2011). First report of a rhabdovirus infecting alfalfa in argentina. *Plant Dis.* 95:771. doi: 10.1094/pdis-10-10-0764

Bejerman, N., Trucco, V., de Breuil, S., Pardina, P. R., Lenardon, S., and Giolitti, F. (2018). Genome characterization of an argentinean isolate of alfalfa leaf curl virus. *Arch. Virol.* 163, 799–803. doi: 10.1007/s00705-017-3673-x

Bekal, S., Domier, L. L., Gonfa, B., McCoppin, N. K., Lambert, K. N., and Bhalerao, K. (2014). A novel flavivirus in the soybean cyst nematode. *J. Gen. Virol.* 95, 1272–1280. doi: 10.1099/vir.0.060889-0

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

Bernardo, P., Muhire, B., François, S., Deshoux, M., Hartnady, P., Farkas, K., et al. (2016). Molecular characterization and prevalence of two capulaviruses: alfalfa leaf curl virus from France and Euphorbia caput-medusae latent virus from South Africa. *Virology* 493, 142–153. doi: 10.1016/j.virol.2016.03.016

Blouin, A. G., Ross, H. A., Hobson-Peters, J., O'Brien, C. A., Warren, B., and MacDiarmid, R. (2016). A new virus discovered by immunocapture of double-stranded RNA, a rapid method for virus enrichment in metagenomic studies. *Mol. Ecol. Res.* 16, 1255–1263. doi: 10.1111/1755-0998.12525

Boccardo, G., Lisa, V., Luisoni, E., and Milne, R. G. (1987). Cryptic plant viruses. *Adv. Virus Res.* 32, 171–214. doi: 10.1016/S0065-3527(08)60477-7

Brough, R. C., Robison, L. R., and Jackson, R. H. (1977). The historical diffusion of alfalfa. *J. Agron. Educ.* 6, 13–19. doi: 10.2134/jae.1977.0013

Candresse, T., Filloux, D., Muhire, B., Julian, C., Galzi, S., Fort, G., et al. (2014). Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLoS One* 9:e102945. doi: 10.1371/journal.pone.0102945

Davoodi, Z., Bejerman, N., Richet, C., Filloux, D., Kumari, S. G., Chatzivassiliou, E. K., et al. (2018). The westward journey of alfalfa leaf curl virus. *Viruses* 10:542.

Debat, H. J., and Bejerman, N. (2019). Novel bird's-foot trefoil RNA viruses provide insights into a clade of legume-associated enamoviruses and rhabdoviruses. *Arch. Virol.* 164, 1419–1426. doi: 10.1007/s00705-019-04193-1

Donaire, L., Wang, Y., Gonzalez-Ibeas, D., Mayer, K. F., Aranda, M. A., and Llave, C. (2009). Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology* 392, 203–214. doi: 10.1016/j.virol.2009.07.005

Dong, W., Liu, X., Li, D., Gao, T., and Song, Y. (2018). Transcriptional profiling reveals that a MYB transcription factor MsMYB4 contributes to the salinity stress response of alfalfa. *PLoS One* 3:e0204033. doi: 10.1371/journal.pone.0204033

Elena, S. F., Bernet, G. P., and Carrasco, J. L. (2014). The games plant viruses play. *Curr. Opin. Virol.* 8, 62–67. doi: 10.1016/j.coviro.2014.07.003

Filloux, D., Dallot, S., Delaunay, A., Galzi, S., Jacquot, E., and Roumagnac, P. (2015). Metagenomics approaches based on virion-associated nucleic acids (vana): an innovative tool for assessing without a priori viral diversity of plants. *Methods Mol. Biol.* 1302, 249–257. doi: 10.1007/978-1-4939-2620-6_18

François, S., Filloux, D., Fernandez, E., Ogliastro, M., and Roumagnac, P. (2018). Viral metagenomics approaches for high-resolution screening of multiplexed arthropod and plant viral communities. *Methods Mol. Biol.* 1746, 77–95. doi: 10.1007/978-1-4939-7683-6_7

Frate, C. A., and Davis, R. M. (2008). *Irrigated Alfalfa Management for Mediterranean and Desert Zones*. Davis, CA: University of California, Division of Agriculture and Natural Recourses.

Fujimoto, Y., Nijo, T., Hosoe, N., Watanabe, K., Maejima, K., Yamaji, Y., et al. (2018). Complete genome sequence of lychnis mottle virus isolated in Japan. *Genome Announc.* 6:e0535-18.

Gaafar, Y. Z. A., Richert-Pöggeler, K. R., Maaß, C., Vetten, H. J., and Ziebell, H. (2019). Characterization of a novel nucleorhabdovirus infecting alfalfa (*Medicago sativa*). *Virol. J.* 16:55. doi: 10.1186/s12985-019-1147-3

Gilbert, K. B., Holcomb, E. E., Allscheid, R. L., and Carrington, J. C. (2019). Hiding in plain sight: new virus genomes discovered via a systematic analysis of fungal public transcriptomes. *PLoS One* 14:e0219207. doi: 10.1371/journal.pone.0219207

Guo, Z. P., Zhang, J. X., Wang, M. L., Guan, Y. Z., Qu, G., Liu, J. Y., et al. (2019). First report of alfalfa leaf curl virus infecting alfalfa (*Medicago sativa*) in China. *Plant Dis.* 104:1001. doi: 10.1094/PDIS-02-19-0318-PDN

Hampton, R. O. (1981). Evidence suggesting identity between alfalfa latent andpea streak viruses. *Phytopathology* 71:223.

Han, S. S., Karasev, A. V., Ieki, H., and Iwanami, T. (2002). Nucleotide sequence and taxonomy of cycas necrotic stunt virus. *Arch. Virol.* 147, 2207–2214. doi: 10.1007/s00705-002-0876-5

Hanada, K., Fukumoto, F., Kusunoki, M., Kameya-Iwaki, M., Tanaka, Y., and Iwanami, T. (2006). Cycas necrotic stunt virus isolated from gladiolus plants in Japan. *J. Gen. Plant. Pathol.* 72, 383–386. doi: 10.1007/s10327-006-0304-x

Ho, T., and Tzanetakis, I. E. (2014). Development of a virus detection and discovery pipeline using next generation sequencing. *Virology* 471–473, 54–60. doi: 10.1016/j.virol.2014.09.019

Huang, X. Q., and Madan, A. (1999). CAP3: a DNA sequence assembly program. *Genome Res.* 9, 868–877. doi: 10.1101/gr.9.9.868

Jiang, P., Shao, J., and Nemchinov, L. G. (2019a). Identification of emerging viral genomes in transcriptomic datasets of alfalfa (*Medicago sativa* L.). *Virol. J.* 16:153. doi: 10.1186/s12985-019-1257-y

Jiang, P., Shao, J., and Nemchinov, L. G. (2019b). Identification of the coding-complete genome of cycas necrotic stunt virus in transcriptomic data sets of alfalfa (*Medicago sativa*). *Microbiol. Resour. Announc.* 8, e981–e919. doi: 10.1128/MRA.00981-19

Jones, S., Baizan-Edge, A., MacFarlane, S., and Torrance, L. (2017). Viral diagnostics in plants using next generation sequencing: computational analysis in practice. *Front. Plant Sci.* 8:1770. doi: 10.3389/fpls.2017.01770

Joshi, N. A., and Fass, J. N. (2011). *Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33) [Software]*. Available at: https://github.com/najoshi/sickle

Julier, B., Gastal, F., Louarn, G., Badenhausser, I., Annicchiarico, P., Crocq, G., et al. (2017). *Legumes in Cropping Systems*. Wallingford: CABI.

Kim, H., Park, D., and Hahn, Y. (2018). Identification a novel RNA viruses in alfalfa (*Medicago sativa*): an *Alphapartitivirus*, *Deltapartitivirus*, and a *Marafivirus*. *Gene* 638, 7–12. doi: 10.1016/j.gene.2017.09.069

Koenig, R. (1987). "Cactus X potexvirus," in *Plant Viruses Online. Descriptions and Lists From the VIDE Database*, eds A. A. Brunt, K. Crabtree, M. J. Dallwitz, A. J. Gibbs, L. Watson, and E. J. Zurcher, (Canberra: The Australian National University).

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., and Barker, I. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1–7. doi: 10.1016/j.virol.2009.03.024

Krupovic, M., Dolja, V. V., and Koonin, E. V. (2015). Plant viruses of the Amalgaviridae family evolved via recombination between viruses with double-stranded and negative-strand RNA genomes. *Biol. Direct.* 10:12. doi: 10.1186/s13062-015-0047-8

Kusunoki, M., Hanada, K., Iwaki, M., Chang, M. U., Doi, Y., and Yora, K. (1986). Cycas necrotic stunt virus, a new member of nepoviruses found in *Cycas revoluta* host range, purification, serology and some other properties. *Jpn. J. Phytopathol.* 2, 302–311. doi: 10.3186/jjphytopath.52.302

Lambert, C., Braxton, C., Charlebois, R. L., Deyati, A., Duncan, P., La Neve, F., et al. (2018). Considerations for optimization of high-throughput sequencing bioinformatics pipelines for virus detection. *Viruses* 10:E528. doi: 10.3390/v10100528

Lamichhane, J. R., and Venturi, V. (2015). Synergisms between microbial pathogens in plant disease complexes: a growing trend. *Front. Plant Sci.* 6:385. doi: 10.3389/fpls.2015.00385

Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lauber, C., Seifert, M., Bartenschlager, R., and Seitz, S. (2019). Discovery of highly divergent lineages of plant-associated astro-like viruses sheds light on the emergence of potyviruses. *Virus Res.* 260, 38–48. doi: 10.1016/j.virusres.2018.11.009

Leclant, F., Alliot, B., and Signoret, P. A. (1973). Transmission et épidémiologie de la maladie à énations de la luzerne (LEV), *Premiers résultats*. *Ann. Phytopathol.* 5, 441–445.

Li, Y., Wang, H., Nie, K., Zhang, C., Zhang, Y., Wang, J., et al. (2016). VIP: an integrated pipeline for metagenomics of virus identification and discovery. *Sci. Rep.* 6, 23774. doi: 10.1038/srep23774

Lim, S., Kwon, S. Y., Lee, J. H., Cho, H. S., Kim, H. S., Park, J. M., et al. (2019). Genomic detection and molecular characterization of two distinct isolates of cycas necrotic stunt virus from *Paeonia suffruticosa* and *Daphne odora*. *Virus Genes* 55, 734–737. doi: 10.1007/s11262-019-01687-7

Liou, M. R., Chen, Y. R., and Liou, R. F. (2004). Complete nucleotide sequence and genome organization of a Cactus virus X strain from *Hylocereus undatus* (Cactaceae). *Arch. Virol.* 149, 1037–1043. doi: 10.1007/s00705-003-0251-1

Ma, Y., Marais, A., Lefebvre, M., Theil, S., Svanella-Dumas, L., Faure, C., et al. (2019). Phytovirome analysis of wild plant populations: comparison of double-stranded RNA and virion-associated nucleic acid metagenomic approaches. *J. Virol.* 94:e01462-19. doi: 10.1128/JVI.01462-19

Maina, S., Zheng, L., Kinoti, W. M., Aftab, M., Nancarrow, N., Trębicki, P., et al. (2019). Metagenomic analysis reveals a nearly complete genome sequence of *alfalfa mosaic virus* from a field pea in Australia. *Microbiol. Res. Announc.* 8:e00766-19. doi: 10.1128/MRA.00766-19

Malvick, D. (2020). *Root Rot and Crown Rots and Virus Diseases of Alfalfa and Clover*. Available at: https://fyi.extension.wisc.edu/forage/root-rot-and-crown-rots-and-virus-diseases-of-alfalfa-and-clover/ (accessed June 18, 2020).

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

Martelli, G. P., Adams, M. J., Kreuze, J. F., and Dolja, V. V. (2007). Family flexiviridae: a case study in virion and genome plasticity. *Annu. Rev. Phytopathol.* 45, 73–100. doi: 10.1146/annurev.phyto.45.062806.094401

Martin, R. R., Zhou, J., and Tzanetakis, I. E. (2011). Blueberry latent virus: an amalgam of the Partitiviridae and Totiviridae. *Virus Res.* 155, 175–180. doi: 10.1016/j.virusres.2010.09.020

Massart, S., Chiumenti, M., De Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., et al. (2019). Virus detection by high-throughput sequencing of small RNAs: large-scale performance testing of sequence analysis strategies. *Phytopathology* 109, 488–497. doi: 10.1094/PHYTO-02-18-0067-R

Miller, D. (2019). "Progressive forage," in *Proceedings of the Second World Alfalfa Conference - Argentina*, Jerome, ID.

Moore, M. D., and Jaykus, L. A. (2018). Virus-bacteria interactions: implications and potential for the applied and agricultural sciences. *Viruses* 10:E61. doi: 10.3390/v10020061

Moreno, A. B., and López-Moya, J. J. (2020). When viruses play team sports: mixed infections in plants. *Phytopathology* 110, 29–48. doi: 10.1094/PHYTO-07-19-0250-FI

Mushegian, A., Shipunov, A., and Elena, S. F. (2016). Changes in the composition of the RNA virome mark evolutionary transitions in green plants. *BMC Biol.* 14:68. doi: 10.1186/s12915-016-0288-8

Nagano, A. J., Honjo, M. N., Mihara, M., Sato, M., and Kudoh, H. (2015). Detection of plant viruses in natural environments by using RNA-Seq. *Methods Mol. Biol.* 1236, 89–98. doi: 10.1007/978-1-4939-1743-3_8

National Alfalfa and Forage Alliance (2018). Available online at: http://alfalfa.org/newsletter/180327nafanews.htm (accessed April 3, 2020).

Nemchinov, L. G. (2017). Development and characterization of the first infectious clone of alfalfa latent virus, a strain of Pea streak virus. *Eur. J. Plant Pathol.* 149, 1019–1022. doi: 10.1007/s10658-017-1237-2

Nemchinov, L. G., François, S., Roumagnac, P., Ogliastro, M., Hammond, R. W., Mollov, D. S., et al. (2018a). Characterization of alfalfa virus F, a new member of the genus *Marafivirus*. *PLoS One* 13:e0203477. doi: 10.1371/journal.pone.0203477

Nemchinov, L. G., Lee, M. N., and Shao, J. (2018b). First report of alphapartitiviruses infecting alfalfa (*Medicago sativa* L.) in the United States. *Microbiol. Resour. Announc.* 7:e01266-18. doi: 10.1128/MRA.01266-18

Nemchinov, L. G., Grinstead, S. C., and Mollov, D. S. (2017a). Alfalfa virus S, a new species in the family *Alphaflexiviridae*. *PLoS One* 12:e0178222. doi: 10.1371/journal.pone.0178222

Nemchinov, L. G., Grinstead, S. C., and Mollov, D. S. (2017b). First report and complete genome sequence of alfalfa enamovirus from Sudan. *Genome Announc.* 5:e0531-17.

Nemchinov, L. G., Shao, J., Lee, M. N., Postnikova, O. A., and Samac, D. A. (2017c). Resistant and susceptible responses in alfalfa (*Medicago sativa*) to bacterial stem blight caused by *Pseudomonas syringae* pv. syringae. *PLoS One* 12:e0189781. doi: 10.1371/journal.pone.0189781

Nemchinov, L. G., Shao, J., and Postnikova, O. A. (2015). Complete genome sequence of the alfalfa latent virus. *Genome Announc.* 3:e00250-15.

Nibert, M. L., Pyle, J. D., and Firth, A. E. (2016). A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. *Virology* 498, 201–220. doi: 10.1016/j.virol.2016.07.002

Ochoa-Corona, F. M., Elliot, D. R., Tang, Z., Lebas, B. S. M., and Alexander, B. J. R. (2003). Detection of Cycas necrotic stunt virus (CNSV) in post-entry quarantine stocks of ornamentals in New Zealand. *Phytopathology* 93:S67.

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. *Bull. OEPP/EPPO Bull.* 48, 219–224. doi: 10.1111/epp.12472

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Pooggin, M. M. (2018). Small rna-omics for plant virus identification, virome reconstruction, and antiviral defense characterization. *Front. Microbiol.* 9:2779. doi: 10.3389/fmicb.2018.02779

Pruss, G., Ge, X., Shi, X. M., Carrington, J. C., and Bowman, V. V. (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

Roossinck, M. J. (2012a). Plant virus metagenomics: biodiversity and ecology. *Annu. Rev. Genet.* 46, 359–369. doi: 10.1146/annurev-genet-110711-155600

Roossinck, M. J. (2012b). "Persistent plant viruses: molecular hitchhikers or epigenetic elements?," in *Viruses: Essential Agents of Life*, ed. G. Witzany, (Berlin: Springer), 177–186. doi: 10.1007/978-94-007-4899-6_8

Roossinck, M. J. (2015). Plants, viruses and the environment: ecology and mutualism. *Virology* 479–480, 271–277. doi: 10.1016/j.virol.2015.03.041

Roossinck, M. J. (2017). Deep sequencing for discovery and evolutionary analysis of plant viruses. *Virus Res.* 239, 82–86. doi: 10.1016/j.virusres.2016.11.019

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

Roumagnac, P., Granier, M., Bernardo, P., Deshoux, M., Ferdinand, R., Galzi, S., et al. (2015). Alfalfa leaf curl virus: an aphid-transmitted geminivirus. *J. Virol.* 89, 9683–9688. doi: 10.1128/jvi.00453-15

Sabanadzovic, S., Valverde, R. A., Brown, J. K., Martin, R. R., and Tzanetakis, I. E. (2009). Southern tomato virus: the link between the families *Totiviridae* and *Partitiviridae*. *Virus Res.* 140, 130–137. doi: 10.1016/j.virusres.2008.11.018

Samac, D. A., Rhodes, L. H., and Lamp, W. O. (2015). *Compendium of Alfalfa Diseases and Pests*, 3rd Edn, St. Paul, MI: The American Phytopathological Society.

Samarfard, S., McTaggart, A. R., Sharman, M., Bejerman, N. E., and Dietzgen, R. G. (2020). Viromes of ten alfalfa plants in Australia reveal diverse known viruses and a novel RNA virus. *Pathogens* 9:E214. doi: 10.3390/pathogens9030214

Shaffer, C., Gress, J. C., and Tzanetakis, I. E. (2019). First report of cycas necrotic stunt virus and lychnis mottle virus in peony in the United States. *Plant Dis.* 103:1048. doi: 10.1094/pdis-11-18-2089-pdn

Shapiro, L. R., Salvaudon, L., Mauck, K. E., Pulido, H., De Moraes, C. M., Stephenson, A. G., et al. (2013). Disease interactions in a shared host plant: effects of pre-existing viral infection on cucurbit plant defense responses and resistance to bacterial wilt disease. *PLoS One* 8:e77393. doi: 10.1371/journal.pone.0077393

SIAS, (0000). *Sequence Identity and Similarity. Complutense University of Madrid*. Available at: http://imed.med.ucm.es/Tools/sias.html (accessed April 1, 2020).

Song, L., Jiang, L., Chen, Y., Shu, Y., Bai, Y., and Guo, C. (2016). Deep-sequencing transcriptome analysis of field-grown *Medicago sativa* L. crown buds acclimated to freezing stress. *Funct. Integr. Genom.* 16, 495–511. doi: 10.1007/s10142-016-0500-5

Su, L., Li, Z., Bernardy, M., Wiersma, P. A., Cheng, Z., and Xiang, Y. (2015). The complete nucleotide sequence and genome organization of pea streak virus (genus Carlavirus). *Arch. Virol.* 160, 2651–2654. doi: 10.1007/s00705-015-2467-2462

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

Syller, J., and Grupa, A. (2016). Antagonistic within-host interactions between plant viruses: molecular basis and impact on viral and host fitness. *Mol. Plant Pathol.* 17, 769–782. doi: 10.1111/mpp.12322

Thompson, J. R., Dasgupta, I., Fuchs, M., Iwanami, T., Karasev, A. V., Petrzik, K., et al. (2017). ICTV virus taxonomy profile: *Secoviridae*. *J. Gen. Virol.* 98, 529–531.

Tollenaere, C., Susi, H., and Laine, A. L. (2016). Evolutionary and epidemiological implications of multiple infection in plants. *Trends Plant Sci.* 1, 80–90. doi: 10.1016/j.tplants.2015.10.014

Trucco, V., Breuil, S., Bejerman, N., Lenardon, S., and Giolitti, F. (2016). Bean leafroll virus (BLRV) in argentina: molecular characterization and detection in alfalfa fields. *Eur. J. Plant Pathol.* 146, 207–212. doi: 10.1007/s10658-016-0899-5

Trucco, V., de Breuil, S., Bejerman, N., Lenardon, S., and Giolitti, F. (2014). Complete nucleotide sequence of Alfalfa mosaic virus isolated from alfalfa (*Medicago sativa* L.) in Argentina). *Virus Genes* 48, 562–565. doi: 10.1007/s11262-014-1045-1040

Undersander, D., Cosgrove, D., Cullen, E., Grau, G., Rice, M. E., Renz, M., et al. (2011). *Alfalfa Management Guide*. Madison, WI: Science Society of America Inc.

Undersander, D., Gray, F. A., Kelling, K., and Rice, M. E. (2020). *Alfalfa Analyst*, 3rd Edn, Available at: https://www.americasalfalfa.com/Alfalfa-University-Content/Pests-Pest-Management/Alfalfa-Analyst (assessed May 14, 2020).

Vainio, E. J., Chiba, S., Ghabrial, S. A., Maiss, E., Roossinck, M., Sabanadzovic, S., et al. (2018). ICTV virus taxonomy profile: partitiviridae. *J. Gen. Virol.* 99, 17–18. doi: 10.1099/jgv.0.000985

Van Leur, J. A. G., and Kumari, S. G. (2011). A survey of lucerne in northern New South Wales for viruses of importance to the winter legume industry. *Austr. Plant Pathol.* 40, 180–186. doi: 10.1007/s13313-011-0028-z

Veerisetty, V. (1979). *Description of Plant Viruses. Alfalfa Latent Virus*. Available online at: http://www.dpvweb.net/dpv/showdpv.php?dpvno=211 (accessed April 3, 2020).

Veerisetty, V., and Brakke, M. K. (1977). Alfalfa latent virus, a naturally occurring carlavirus in alfalfa. *Phytopathology* 67, 1202–1206. doi: 10.1094/phyto-67-1202

Vieira, P., and Nemchinov, L. G. (2019). A novel species of RNA virus associated with root lesion nematode *Pratylenchus penetrans*. *J. Gen. Virol.* 100, 704–708. doi: 10.1099/jgv.0.001246

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109, 716–725. doi: 10.1094/PHYTO-07-18-0257-RVW

Vincelli, P., and Smith, R. (2014). *Managing Diseases of Alfalfa. University of Kentucky College of Agriculture, Food and Environment. Plant Pathology Fact Sheet. PPFS-AG-F-09*. Available online at: https://plantpathology.ca.uky.edu/files/ppfs-ag-f-09.pdf (accessed April 3, 2020).

Wren, J. D., Roossinck, M. J., Nelson, R. S., Scheets, K., Palmer, M. W., and Melcher, U. (2006). Plant virus biodiversity and ecology. *PLoS Biol.* 4:e80. doi: 10.1371/journal.pone.000080

Wylie, S. J., Luo, H., Li, H., and Jones, M. G. (2012). Multiple polyadenylated RNA viruses detected in pooled cultivated and wild plant samples. *Arch. Virol.* 157, 271–284. doi: 10.1007/s00705-011-1166-x

Yamashita, A., Sekizuka, T., and Kuroda, M. (2016). VirusTAP: viral genome-targeted assembly pipeline. *Front. Microbiol.* 7:32. doi: 10.3389/fmicb.2016.00032

Yoo, R. H., Zhao, F., Lim, S., Igori, D., Kim, S. M., An, T. J., et al. (2015a). The complete genome sequences of two isolates of cnidium vein yellowing virus, a tentative new member of the family *Secoviridae*. *Arch. Virol.* 160, 2911–2914. doi: 10.1007/s00705-015-2557-1

Yoo, R. H., Zhao, F., Lim, S., Igori, D., Lee, S.-H., and Moon, J. S. (2015b). The complete nucleotide sequence and genome organization of lychnis mottle virus. *Arch. Virol.* 160, 2891–2894. doi: 10.1007/s00705-015-2501-4

Zaumeyer, W. J. (1938). A streak disease of peas and its relation to several strains of alfalfa mosaic virus. *J. Agr. Res.* 56, 747–772.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Zhang, S., Shi, Y., Cheng, N., Du, H., Fan, W., and Wang, C. (2015). De novo characterization of fall dormant and nondormant alfalfa (*Medicago sativa* L.) leaf transcriptome and identification of candidate genes related to fall dormancy. *PLoS One* 10:e0122170. doi: 10.1371/journal.pone.0122170

Zhao, G., Wu, G., Lim, E. S., Droit, L., Krishnamurthy, S., Barouch, D. H., et al. (2017). VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology* 503, 21–30. doi: 10.1016/j.virol.2017.01.005

Zhao, S., Zhang, Y., Gamini, R., Zhang, B., and von Schack, D. (2018). Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Sci. Rep.* 8:4781. doi: 10.1038/s41598-018-23226-23224

Zheng, Y., Gao, S., Padmanabhan, C., Li, R., Galvez, M., Gutierrez, D., et al. (2017). VirusDetect: an automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology* 500, 130–138. doi: 10.1016/j.virol.2016.10.017

Zvereva, A. S., Golyaev, V., Turco, S., Gubaeva, E. G., Rajeswaran, R., Schepetilnikov, M. V., et al. (2016). Viral protein suppresses oxidative burst and salicylic acid-dependent autophagy and facilitates bacterial growth on virus-infected plants. *New Phytol.* 211, 1020–1034. doi: 10.1111/nph.13967

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.

Copyright © 2020 Bejerman, Roumagnac and Nemchinov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

REVIEW
published: 16 September 2020
doi: 10.3389/fmicb.2020.588427

The Plant Negative-Sense RNA Virosphere: Virus Discovery Through New Eyes

Nicolás Bejerman[1,2], Humberto Debat[1,2] and Ralf G. Dietzgen[3]*

[1] Instituto de Patología Vegetal – Centro de Investigaciones Agropecuarias – Instituto Nacional de Tecnología Agropecuaria, Córdoba, Argentina, [2] Consejo Nacional de Investigaciones Científicas y Técnicas, Unidad de Fitopatología y Modelización Agrícola, Buenos Aires, Argentina, [3] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St. Lucia, QLD, Australia

OPEN ACCESS

Edited by:
John Wesley Randles, University of Adelaide, Australia

Reviewed by:
Elvira Fiallo-Olivé, Institute of Subtropical and Mediterranean Horticulture La Mayora, Spain
Yi Xu, Nanjing Agricultural University, China
Thierry Candresse, Institut National de la Recherche Agronomique (INRA), France

***Correspondence:**
Nicolás Bejerman
nicobejerman@gmail.com

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

Received: *28 July 2020*
Accepted: *27 August 2020*
Published: *16 September 2020*

Citation:
Bejerman N, Debat H and Dietzgen RG (2020) The Plant Negative-Sense RNA Virosphere: Virus Discovery Through New Eyes. Front. Microbiol. 11:588427. doi: 10.3389/fmicb.2020.588427

The use of high-throughput sequencing (HTS) for virus diagnostics, as well as the importance of this technology as a valuable tool for discovery of novel viruses has been extensively investigated. In this review, we consider the application of HTS approaches to uncover novel plant viruses with a focus on the negative-sense, single-stranded RNA virosphere. Plant viruses with negative-sense and ambisense RNA (NSR) genomes belong to several taxonomic families, including *Rhabdoviridae, Aspiviridae, Fimoviridae, Tospoviridae,* and *Phenuiviridae.* They include both emergent pathogens that infect a wide range of plant species, and potential endophytes which appear not to induce any visible symptoms. As a consequence of biased sampling based on a narrow focus on crops with disease symptoms, the number of NSR plant viruses identified so far represents only a fraction of this type of viruses present in the virosphere. Detection and molecular characterization of NSR viruses has often been challenging, but the widespread implementation of HTS has facilitated not only the identification but also the characterization of the genomic sequences of at least 70 NSR plant viruses in the last 7 years. Moreover, continuing advances in HTS technologies and bioinformatic pipelines, concomitant with a significant cost reduction has led to its use as a routine method of choice, supporting the foundations of a diverse array of novel applications such as quarantine analysis of traded plant materials and genetic resources, virus detection in insect vectors, analysis of virus communities in individual plants, and assessment of virus evolution through ecogenomics, among others. The insights from these advancements are shedding new light on the extensive diversity of NSR plant viruses and their complex evolution, and provide an essential framework for improved taxonomic classification of plant NSR viruses as part of the realm *Riboviria*. Thus, HTS-based methods for virus discovery, our 'new eyes,' are unraveling in real time the richness and magnitude of the plant RNA virosphere.

Keywords: HTS, virus discovery, plant NSR viruses, virosphere, bioinformactics

INTRODUCTION

Viruses are the most numerous biological entities on Earth, but the number of reported and formally described virus species, the known virosphere, is exiguous. This underestimated global virus landscape has led to a distorted view of virus diversity and function. Some aspects of virus biology including their small size, rapid rate of evolution, or lack of universally conserved viral genetic markers are likely the major cause for the lack of knowledge of most of the viruses present on our planet (Koonin et al., 2015; Zhang et al., 2019). Thus, as we explore the virosphere, it becomes evident that only a tiny proportion (about 1% according to some estimates) has been characterized, with a major bias against the identification of the most divergent genomes (Zhang et al., 2018).

High-throughput sequencing (HTS) technology, also referred to as next generation sequencing, has revolutionized the nucleic acid characterization process since it allows the parallel sequencing of millions of nucleotides in a short period of time at a very high redundancy (depth of sequencing) without any *a priori* knowledge. Thus, when combined with specific bioinformatics tools, HTS provides a powerful, efficient and economical alternative that has enabled not only the untargeted detection of both known and unknown viruses that inhabit a particular organism or environment but also a rapid characterization of their genomes (Massart et al., 2017; Villamor et al., 2019). Thus, the steady increase in the adoption of HTS in the past decade has accelerated virus discovery of both wild and cultivated plant species, which led to the advancement of our knowledge about the diversity of viruses in nature.

High-throughput sequencing is also widely used to advance the molecular characterization not only of those viruses which have a poorly characterized genome, but also of preserved historic virus isolates which had been subjected to studies of their biological properties some decades ago. For instance, sterile stunt disease of maize, characterized by severe stunting and top necrosis of susceptible maize genotypes, was first reported in 1977 in Australia, but the genome of its etiological agent, maize sterile stunt virus was only characterized by HTS and recently reported (Dietzgen and Higgins, 2019). Similarly, sowthistle yellow vein virus (SYVV) was recently rediscovered and its genome characterized after a hiatus of over 30 years following pioneering virus and vector biology research (Stenger et al., 2020).

In spite of the many advantages of HTS, there are some limitations and constraints that should be considered when using this platform for virus discovery. Given the high sensitivity of HTS, this technology may detect contaminant viral sequences or viruses that may not be actually replicating in the sampled plant tissue where they were found (Blawid et al., 2017). In this context it is worth emphasizing the importance of traditional wet bench virology experiments to complement viral HTS data where possible, such as virus isolation and transmission experiments, *in situ* virus particle detection by electron microscopy or immunoassays, to name a few.

Most of the plant viruses described so far, whose nucleotide sequences are available in public databases, were discovered in exemplars of cultivated plant species that showed conspicuous disease symptoms, limiting our view of viral diversity (Wren et al., 2006). However, the steady increase in the last few years of metagenomic studies to characterize the viromes of wild plant species, many of which did not show any visible symptoms, led to the identification of many new viruses. Moreover, several recent reports provide evidence of viruses which appear to be essentially cryptic, identified from cultivated plants (Bernardo et al., 2018; Schoelz and Stewart, 2018; Susi et al., 2019; Ma et al., 2020). Nevertheless, for most of the newly discovered viruses, subsequent studies to characterize their biological properties, such as symptoms in different hosts, and potential vectors, among others, in both cultivated and wild plant species are scarce, and likely will be neglected due to lack of economic significance or unavailability of preserved samples. The huge gap between virus discovery and biological characterization of new viruses, is due to the latter requiring time-consuming research efforts. Therefore, we are advancing into a scenario where the classification of most novel viruses will be based *only* on their genomic sequences which constitutes a paradigm shift in their taxonomical classification (Simmonds et al., 2017; Kuhn et al., 2019), thus representing a major challenge for virologists and the International Committee on Taxonomy of Viruses (ICTV).

Although many pipelines are available for plant virus discovery through HTS, all share a common backbone (Villamor et al., 2019). Various HTS technologies are available commercially, but Illumina short read shotgun sequencing platforms are the most popular choice and most widely used for HTS of viruses because of their high throughput, low error rate and high cost effectiveness among currently available HTS platforms (Villamor et al., 2019). Four main classes of nucleic acids have been targeted as templates for HTS, (*i*) total plant RNA extracts, usually with a ribosomal depletion step (*ii*) virion-associated nucleic acids (VANA) extracted from purified viral particles, (*iii*) double-stranded RNA (dsRNA), enriched through cellulose chromatography or monoclonal antibody pull-down, and (*iv*) small interfering RNAs (siRNAs) (Roossinck et al., 2015; Wu et al., 2015; Blouin et al., 2016). These different strategies are characterized by diverse caveats. For instance, VANA is inefficient in detecting viruses that lack virions, or siRNA-based protocols are not as reliable for complete genome assembly of novel viruses. Sequencing of siRNAs and total RNA are the two approaches most generically applicable to viruses with different genome types and replication strategies, and can be relatively easily integrated into workflows of diagnostic laboratories (Pecman et al., 2017).

Recent technological advances have led to the appreciation of exciting novel aspects of negative-sense and ambisense RNA (NSR) viruses, as recently reviewed by German et al. (2020). Plant NSR viruses are considered emerging viral pathogens (German et al., 2020). In light of the growing interest in these viruses, in this review we delve into the application of HTS approaches to uncover the abundance and

diversity of the negative-sense, single-stranded RNA virosphere associated with plants.

HIGH-THROUGHPUT SEQUENCING TO UNCOVER THE NEGATIVE-SENSE, SINGLE-STRANDED RNA VIROSPHERE OF PLANTS

Hundreds of plant viruses have been characterized by HTS in the last few years (Blawid et al., 2017; Villamor et al., 2019), and most of them have positive-sense RNA genomes. On the other hand, RNA viruses with negative-sense and ambisense genomes discovered through HTS represent only a small fraction of the total number of viruses discovered, thus representing a tiny fraction of the plant virosphere. For example, the characterization of the virome of different *Solanum* species resulted in the identification of viruses belonging to 20 different families, but only one was a NSR virus (Ma et al., 2020). In addition, the characterization of a tomato virome resulted in the identification of 22 viruses belonging to 12 genera, but only three genomes corresponded to NSR viruses (Xu et al., 2017). Moreover, the characterization of a papaya virome resulted in the identification of 52 viruses, but only one was a NSR virus (Alcalá-Briseño et al., 2020). It is worth mentioning that typically, virome studies based on HTS generate a significant amount of sequence contigs that lack detectable homology to both the sampled host and any microorganism. It has been suggested that a fraction of these sequences may correspond to viral 'dark' matter, which may imply that many deeply divergent viruses, or viruses lacking common ancestry or similarity with known virus families, remain to be discovered; this may not happen until better frameworks are implemented to identify viral sequences regardless of their sequence similarity to known viruses (Obbard et al., 2020). Current virome analyses usually rely on sequence similarity searches to identify virus-like sequences through inferred homology. This approach limits the identification of new viruses that can be discovered through traditional empirical search algorithms such as BLAST using identity thresholds of target sequences to genomes, genes, proteins or protein motifs of known viruses (Obbard et al., 2020). As a consequence, in contrast to the more straightforward hypothesis that plant NSR viruses are relatively rare, it is plausible that NSR virus sequences may have been overlooked due to extreme divergence from known viruses, thus providing an alternative reason why so few NSR viruses have been identified when the viromes of different plant hosts are characterized. Nevertheless, as novel bioinformatics tools are developed to increase the sensitivity of similarity search algorithms and more refined *ab initio* probabilistic methods are implemented, such as hidden Markov models that incorporate position-specific information into the alignment process of a group of highly divergent, evolutionarily related sequences and use these profiles to identify virus sequences (Skewes-Cox et al., 2014) or methods that rely on support vector machines (Liao and Noble, 2003), the identification of novel, more divergent viruses will likely become possible. This may result in an increased number of identified NSR viruses, which will allow us to deepen our understanding about the evolution and diversity of NSR viruses. This is clearly illustrated by the discovery of two NSR viruses associated with apple rubbery wood disease. The initial attempt to determine by NGS if any virus was associated with the disease was unsuccessful (Jakovljevic et al., 2016). However, the NGS data was later reanalyzed in depth using a bioinformatics approach focused on viral conserved protein motifs that resulted in the identification and genome assembly of two novel NSR viruses (Rott et al., 2018).

The use of HTS technologies has not only allowed the identification and characterization of novel NSR viruses in several plant hosts (**Table 1**), but has also enabled the completion of genome sequences of NSR viruses for which biological properties and only partial genome fragments were known (**Table 2**). The two nucleic acid classes mainly targeted when using HTS to sequence NSR viruses are total RNA and siRNAs (**Tables 1**, **2**). When Pecman et al. (2017) compared these approaches they found that both can be used to detect and identify a wide array of known plant viruses in the tested samples including orthotospoviruses. However, on this occasion a putative novel cytorhabdovirus genome could only be assembled *de novo* from the sequencing data generated from total RNA and not from the small RNA dataset, due to the low number of short reads in the latter (Pecman et al., 2017), thus the choice of nucleic acid types used in HTS may have an effect on the range of viruses that can be identified. However, a few novel plant rhabdoviruses including cytorhabdoviruses have been identified using small RNA as sequencing template (**Table 1**).

Most plant NSR viruses are transmitted by arthropods in a persistent-circulative and propagative manner, thus they are adapted to infect both arthropod and plant cells. In fact, it has been suggested that plant NSR viruses may have originated from arthropod viruses that evolved to also infect plants (Whitfield et al., 2018; Dolja et al., 2020). Therefore, another potential source to discover novel plant-associated NSR viruses is the characterization of arthropod vector viromes. For instance, Li et al. (2015) performed deep transcriptome sequencing of 70 arthropod species that resulted in the identification of three novel cytorhabdoviruses (Wuhan insect viruses 4–6) associated with the mealy plum aphid (*Hyalopterus pruni*). Wuhan insect viruses 4–6 may be plant viruses based on their clear phylogenetic relationship with plant rhabdoviruses in the genus *Cytorhabdovirus* and that their genomes encode a P3 protein that most resembles plant rhabdovirus cell-to-cell movement proteins.

Since most NSR viruses replicate in both the plant host and the arthropod vector, the true virus host origins cannot be unambiguously discerned from the HTS data. Some of the newly discovered NSR viruses are highly divergent which makes it even more difficult to unequivocally determine if they are plant or plant-associated viruses, and it will require a significant amount of research to confirm their status, which is rarely carried out. This is exemplified by the recently discovered coguviruses and rubodviruses, where the host assignment based on their phylogenetic relationships is only preliminary. Thus, further biological studies will be required to determine if they are plant,

TABLE 1 | Novel NSR viruses discovered using HTS.

Virus family/genus	Virus	Type of target/sequencing technology	References
Aspiviridae/Ophiovirus	Blueberry mosaic associated ophiovirus	Total nucleic acid/Illumina HiSeq	Thekke-Veetil et al. (2014)
Phenuiviridae/Coguvirus	Citrus concave gum-associated virus	Small RNAs/Illumina HiSeq; total RNA/Illumina HiSeq	Navarro et al. (2018a); Wright et al. (2018)
Phenuiviridae/Coguvirus	Citrus virus A	Small RNAs/Illumina Genome HiScan	Navarro et al. (2018b)
Phenuiviridae/Coguvirus	Watermelon crinkle leaf-associated virus 1	Small RNAs and total RNA/Illumina HiSeq	Xin et al. (2017)
Phenuiviridae/Coguvirus	Watermelon crinkle leaf-associated virus 2	Small RNAs and total RNA/Illumina HiSeq	Xin et al. (2017)
Phenuiviridae/Coguvirus	Grapevine associated cogu-like virus 1	Total RNA/Illumina HiSeq	Chiapello et al. (2020)
Phenuiviridae/Coguvirus	Grapevine associated cogu-like virus 2	Total RNA/Illumina HiSeq	Chiapello et al. (2020)
Phenuiviridae/Coguvirus	Grapevine associated cogu-like virus 3	Total RNA/Illumina HiSeq	Chiapello et al. (2020)
Phenuiviridae/Coguvirus	Grapevine associated cogu-like virus 4	Total RNA/Illumina NovaSeq	Bertazzon et al. (2020)
Phenuiviridae/Rubodvirus	Apple rubbery wood virus 1	Double-stranded RNA/Illumina HiSeq; total RNA/Illumina HiSeq	Rott et al. (2018); Wright et al. (2018)
Phenuiviridae/Rubodvirus	Apple rubbery wood virus 2	Double-stranded RNA/Illumina HiSeq; total RNA/Illumina HiSeq	Rott et al. (2018); Wright et al. (2018)
Phenuiviridae/Rubodvirus	Grapevine Muscat rose virus	Total nucleic acid/Illumina HiSeq	Diaz-Lara et al. (2019)
Phenuiviridae/Rubodvirus	Grapevine Garan dmak virus	Total nucleic acid/Illumina HiSeq	Diaz-Lara et al. (2019)
Phenuiviridae/Tenuivirus	Melon chlorotic spot virus	Small RNAs/Illumina HiSeq; total RNA/Illumina MiSeq	Lecoq et al. (2019); Gaafar et al. (2019a)
Phenuiviridae/Tenuivirus	Ramu stunt virus	Total RNA/Illumina HiSeq	Mollov et al. (2016)
Phenuiviridae/Tenuivirus	European wheat striate mosaic virus	Small and total RNA/Illumina HiSeq and NextSeq	Sõmera et al. (2020)
Tospoviridae/Orthotospovirus	Alstroemeria yellow spot virus	Total RNA/Illumina HiSeq	Hassani-Mehraban et al. (2019)
Tospoviridae/Orthotospovirus	Chili yellow ringspot virus	Total RNA/Illumina MiSeq	Zheng et al. (2020)
Fimoviridae/Emaravirus	Blackberry leaf mottle associated virus	Double-stranded RNA/Illumina and 454-Junior (Roche)	Hassan et al. (2017)
Fimoviridae/Emaravirus	Pigeonpea sterility mosaic virus 2	Double-stranded RNA/Illumina HiScan; small RNAs/Illumina GAIIx	Elbeaino et al. (2015); Kumar et al. (2017)
Fimoviridae/Emaravirus	Pistacia virus B	Total RNA/Illumina NextSeq	Buzkan et al. (2019)
Fimoviridae/Emaravirus	Redbud yellow ringspot-associated virus	Double-stranded RNA/Illumina	Di Bello et al. (2016)
Fimoviridae/Emaravirus	Ti ringspot-associated virus	Double-stranded RNA/Illumina HiSeq and MiSeq	Olmedo-Velarde et al. (2019)
Fimoviridae/Emaravirus	Actinidia chlorotic ringspot-associated virus	Small RNAs/Illumina Genome analyzer	Zheng et al. (2017)
Fimoviridae/Emaravirus	Jujube yellow mottle-associated virus	Small RNAs/BGISEQ and total RNA/Illumina HiSeq	Yang et al. (2019)
Fimoviridae/Emaravirus	Blue palo verde broom virus	Total RNA/Illumina HiSeq	Ilyas et al. (2018)
Fimoviridae/Emaravirus	Camellia japonica associated emaravirus 1	Total RNA/Illumina HiSeq	Peracchio et al. (2020)
Fimoviridae/Emaravirus	Camellia japonica associated emaravirus 2	Total RNA/Illumina HiSeq	Peracchio et al. (2020)
Fimoviridae/Emaravirus	Aspen mosaic-associated virus	Total RNA/Illumina HiSeq	von Bargen et al. (2020)
Fimoviridae/Emaravirus	Perilla mosaic virus	Total RNA/Illumina NovaSeq	Kubota et al. (2020)
Fimoviridae/Emaravirus	Lilac chlorotic ringspot-associated virus	Total RNA/Illumina HiSeq	Wang et al. (2020)
Rhabdoviridae/Cytorhabdovirus	Strawberry associated virus 1	Small RNAs/Illumina MiSeq; total RNA/Illumina HiSeq and PCR products/Ion Proton	Ding et al. (2019); Franova et al. (2019a)

(Continued)

TABLE 1 | Continued

Virus family/genus	Virus	Type of target/sequencing technology	References
Rhabdoviridae/Cytorhabdovirus	Tomato yellow mottle-associated virus	Small RNAs/Illumina HiSeq	Xu et al. (2017)
Rhabdoviridae/Cytorhabdovirus	Maize associated rhabdovirus	Total RNA/Illumina HiSeq	Willie and Stewart (2017)
Rhabdoviridae/Cytorhabdovirus	Colocasia bobone disease-associated virus	Total RNA/Illumina HiSeq	Higgins et al. (2016)
Rhabdoviridae/Cytorhabdovirus	Papaya virus E; bean-associated cytorhabdovirus; citrus-associated rhabdovirus	Total RNA/Illumina HiSeq; double stranded RNA/Illumina MiSeq; total RNA/Illumina HiSeq	Medina-Salguero et al. (2019); Alves-Freitas et al. (2019); Zhang et al. (2020)
Rhabdoviridae/Cytorhabdovirus	Cabbage cytorhabdovirus 1	Total RNA/Illumina MiSeq	Pecman et al. (2017)
Rhabdoviridae/Cytorhabdovirus	Rice stripe mosaic virus	Small RNAs/Illumina HiSeq	Yang et al. (2017)
Rhabdoviridae/Cytorhabdovirus	Wuhan insect virus 4	Total RNA/Illumina HiSeq	Li et al. (2015)
Rhabdoviridae/Cytorhabdovirus	Wuhan insect virus 5	Total RNA/Illumina HiSeq	Li et al. (2015)
Rhabdoviridae/Cytorhabdovirus	Wuhan insect virus 6	Total RNA/Illumina HiSeq	Li et al. (2015)
Rhabdoviridae/Cytorhabdovirus	Trifolium pratense virus A	Total RNA/Illumina HiSeq	Franova et al. (2019b)
Rhabdoviridae/Cytorhabdovirus	Trifolium pratense virus B	Total RNA/Illumina HiSeq	Franova et al. (2019b)
Rhabdoviridae/Cytorhabdovirus	Yerba mate chlorosis-associated virus	Small RNAs/Illumina HiSeq	Bejerman et al. (2017)
Rhabdoviridae/Cytorhabdovirus	Yerba mate virus A	Total RNA/Illumina HiSeq	Bejerman et al. (2020)
Rhabdoviridae/Cytorhabdovirus	Persimmon Virus A	Total RNA/Illumina Genome Analyzer	Ito et al. (2013)
Rhabdoviridae/Cytorhabdovirus	Trichosantes associated rhabdovirus 1	Total RNA/Illumina HiSeq	Goh et al. (2020)
Rhabdoviridae/Cytorhabdovirus	Cucurbit cytorhabdovirus 1	Total RNA/Illumina NovaSeq	Orfanidou et al. (2020)
Rhabdoviridae/Betanucleorhabdovirus	Alfalfa-associated nucleorhabdovirus	Total RNA/Illumina MiSeq	Gaafar et al. (2019b)
Rhabdoviridae/Betanucleorhabdovirus	Apple rootstock virus A	Total RNA/Illumina HiSeq	Baek et al. (2019)
Rhabdoviridae/Alphanucleorhabdovirus	Physostegia chlorotic mottle virus	Total RNA/Illumina MiSeq	Menzel et al. (2016); Gaafar et al. (2018)
Rhabdoviridae/Alphanucleorhabdovirus	Morogoro maize-associated virus	Total RNA/Illumina HiSeq	Read et al. (2019)
Rhabdoviridae/Betanucleorhabdovirus	Black currant associated rhabdovirus	Total RNA/Illumina NextSeq	Wu et al. (2018)
Rhabdoviridae/Alphanucleorhabdovirus	Wheat yellow striate virus	Total RNA/Illumina HiSeq	Liu et al. (2018)
Rhabdoviridae/Betanucleorhabdovirus	Green Sichuan pepper nucleorhabdovirus	Small RNAs and total RNA/Illumina HiSeq	Cao et al. (2019)
Rhabdoviridae/Betanucleorhabdovirus	Bird's-foot trefoil-associated virus 1	Total RNA/Illumina HiSeq	Debat and Bejerman (2019)
Rhabdoviridae/Alphanucleorhabdovirus	Peach virus 1	Total RNA/Illumina HiSeq	Zhou et al. (2020)
Rhabdoviridae/Betanucleorhabdovirus	Cardamom vein clearing rhabdovirus 1	Small RNA/Illumina HiSeq	Bhat et al. (2020)
Rhabdoviridae/Dichorhavirus	Clerodendrum chlorotic spot virus	Total RNA/Illumina HiSeq	Ramos-González et al. (2018)
Rhabdoviridae/Dichorhavirus	Citrus leprosis virus N	Total RNA/Illumina HiSeq	Ramos-González et al. (2017)
Rhabdoviridae/Dichorhavirus	Citrus chlorotic spot virus	Total RNA/Illumina HiSeq	Chabi-Jesus et al. (2018)
Rhabdoviridae/Varicosavirus	Red clover-associated varicosavirus	Double stranded RNA/Illumina HiSeq	Koloniuk et al. (2018a)
Rhabdoviridae/Varicosavirus	Alopecurus myosuroides varicosavirus 1	Total RNA/454 (Roche) and Illumina HiSeq	Sabbadin et al. (2017)

TABLE 2 | NSR viruses with partial genome sequence already know, whose genomes were characterized using HTS.

Virus family/genus	Virus	Type of target/sequencing technology	References
Phenuiviridae/Tenuivirus	Rice hoja blanca virus	Small RNAs/Illumina HiSeq	Jimenez et al. (2018)
Tospoviridae/Orthotospovirus	Groundnut chlorotic fan-spot virus	Total RNA/Illumina MiSeq	Chou et al. (2017)
Tospoviridae/Orthotospovirus	Chrysanthemum stem necrosis virus	Viral RNA from ribonucleocapsid/454 (Roche)	Dullemans et al. (2015)
Tospoviridae/Orthotospovirus	Mulberry vein banding associated virus	Small RNAs/Illumina Genome Analyzer	Meng et al. (2015)
Tospoviridae/Orthotospovirus	Alstroemeria necrotic streak virus	Total RNA/Illumina HiSeq	Gallo et al. (2018)
Tospoviridae/Orthotospovirus	Melon severe mosaic virus	Total RNA/Illumina NextSeq	Ciuffo et al. (2017)
Tospoviridae/Orthotospovirus	Zucchini lethal chlorosis virus	Viral RNA from ribonucleocapsid/Illumina HiSeq	Lima et al. (2016)
Tospoviridae/Orthotospovirus	Tomato chlorotic spot virus	Total RNA/Illumina HiSeq	Martínez et al. (2018); Adegbola et al. (2019), Fagundes Silva et al. (2019)
Tospoviridae/Orthotospovirus	Polygonum ringspot virus	Small RNAs	Margaria et al. (2014)
Fimoviridae/Emaravirus	Rose rosette virus	Total nucleic acid/Illumina HiSeq and 454-Junior (Roche)	Di Bello et al. (2015)
Fimoviridae/Emaravirus	Pigeonpea sterility mosaic virus 1	Double-stranded RNA/Illumina HiScan; small RNAs/Illumina GAIIx	Elbeaino et al. (2014); Kumar et al. (2017)
Fimoviridae/Emaravirus	Raspberry leaf blotch virus	Total RNA/Illumina MiSeq	Lu et al. (2015)
Fimoviridae/Emaravirus	European mountain ash ringspot-associated virus	Total RNA/Illumina HiSeq	von Bargen et al. (2019)
Fimoviridae/Emaravirus	High plains wheat mosaic virus	Partially purified virion RNA/Illumina MiSeq	Tatineni et al. (2014)
Rhabdoviridae/Cytorhabdovirus	Alfalfa dwarf virus	Small RNAs/Illumina HiSeq	Bejerman et al. (2015)
Rhabdoviridae/Cytorhabdovirus	Raspberry vein chlorosis virus	Total RNA/Illumina NextSeq	Jones et al. (2019)
Rhabdoviridae/Cytorhabdovirus	Barley yellow striate mosaic virus/maize sterile stunt virus	Small RNAs/Illumina HiSeq; viral enriched RNA/454 (Roche)	Yan et al. (2015); Dietzgen and Higgins (2019)
Rhabdoviridae/Cytorhabdovirus	Strawberry crinkle virus	Total RNA/Illumina HiSeq	Koloniuk et al. (2018b)
Rhabdoviridae/Cytorhabdovirus	Maize yellow striate virus	Viral enriched RNA/Illumina HiSeq	Maurino et al. (2018)
Rhabdoviridae/Betanucleorhabdovirus	Datura yellow vein virus	Total RNA/454 (Roche)	Dietzgen et al. (2015)
Rhabdoviridae/Alphanucleorhabdovirus	Maize mosaic virus	Total RNA/Illumina HiSeq	Martin and Whitfield (2019)
Rhabdoviridae/Alphanucleorhabdovirus	Maize Iranian mosaic virus	Total RNA/Illumina HiSeq	Ghorbani et al. (2018)
Rhabdoviridae/Betanucleorhabdovirus	Sowthistle yellow vein virus	Total RNA/Oxford Nanopore	Stenger et al. (2020)
Rhabdoviridae/Dichorhavirus	Coffee ringspot virus	Total RNA/ION Torrent	Ramalho et al. (2014)

plant-associated, or insect viruses (Navarro et al., 2018b; Diaz-Lara et al., 2019; Chiapello et al., 2020).

NEGATIVE-SENSE, SINGLE-STRANDED RNA VIRUS DIVERSITY AND TAXONOMY

Negative-sense and ambisense RNA viruses belong to the phylum *Negarnaviricota*, which includes members characterized by (*i*) negative-sense or ambisense single-stranded unsegmented or segmented genomes, (*ii*) the presence or absence of a lipid membrane enveloping the capsid, and (*iii*) a diverse host range (Diaz-Lara et al., 2019; Koonin et al., 2020; Kuhn et al., 2020). This phylum contains major groups of pathogenic viruses and our current knowledge of these viruses is strongly biased toward agents with special importance for human and animal health, such as influenza virus (*Orthomyxoviridae*), Zaire ebolavirus (*Filoviridae*) or Crimean-Congo hemorrhagic fever virus (*Nairovirus*). Most of the genera contain NSR viruses that infect vertebrates and only about 10% contain phytoviruses (Käfer et al., 2019) (**Figure 1**).

Most of the NSR viruses are divided into two large lineages based on whether their RNA genomes are unsegmented or segmented (Koonin et al., 2020) (**Figure 1**). The unsegmented and some bi-segmented viruses with negative-sense RNA genomes, belong to the order *Mononegavirales* (Koonin et al., 2020; Kuhn et al., 2020). In contrast, most of the segmented viruses with both negative-sense and ambisense RNA genomes belong to the order *Bunyavirales* (Koonin et al., 2020; Kuhn et al., 2020) (**Figure 1**). Other orders such as *Serpentovirales* (which is the only one of these including NSR phytoviruses), *Muvirales*, *Articulavirales*, among others, have been created to accommodate diverse viruses which have been placed in the major phylogenetic gap between the two large groups of NSR viruses (Wolf et al., 2018).

Twelve genera (*Alphanucleorhabdovirus*, *Betanucleorhabdovirus*, *Coguvirus*, *Cytorhabdovirus*, *Dichorhavirus*, *Emaravirus*, *Gammanucleorhabdovirus*, *Ophiovirus*, *Orthotospovirus*, *Rubodvirus*, *Tenuivirus*, and *Varicosavirus*) belonging to five different families (*Rhabdoviridae, Aspiviridae, Fimoviridae, Tospoviridae*, and *Phenuiviridae*) include species of phytoviruses (**Figure 1**). Two of the genera (*Coguvirus* and *Rubodvirus*) were recently created to accommodate novel species of NSR viruses related to members of the *Phenuiviridae* family (Navarro et al., 2018b; Diaz-Lara et al., 2019). Furthermore, the ongoing discovery of many novel nucleorhabdoviruses and dichorhaviruses by HTS in the last few years resulted in a split of the genus *Nucleorhabdovirus* into three new genera (Freitas-Astúa et al., 2019). Therefore, as the pace of discovery of new NSR plant viruses using HTS is speedily increasing, the creation of new genera and families to accommodate some of these newly discovered viruses will be a common classification task in future years.

Rhabdoviridae

The family *Rhabdoviridae* currently comprises 30 genera for 191 species for viruses infectin vertebrates, invertebrates and plants. Six of these genera include 45 species of phytoviruses: *Cytorhabdovirus, Alphanucleorhabdovirus*, *Betanucleorhabdovirus* and *Gammanucleorhabdovirus* (unsegmented genomes), and *Dichorhavirus* and *Varicosavirus* (bi-segmented genomes) (Walker et al., 2018; Kuhn et al., 2020) (**Figure 1**). The virions of these phytorhabdoviruses have bacilliform or rod-shaped morphology, and those with unsegmented genomes are enveloped. Plant-infecting rhabdovirus genomes are 10–16 kb in size and are composed of 6 to 10 genes (Walker et al., 2018) (**Figure 1**).

The increased application of HTS has seen a significant rise in the number of novel plant rhabdoviruses (**Table 1**), as well as the completion of genomic sequences of those viruses with poorly characterized genomes (**Table 2**).

High-throughput sequencing was used successfully to complete the genome sequences of previously reported plant rhabdoviruses where only partial sequence fragments were available, such as Iranian citrus ringspot-associated virus (Sadeghi et al., 2016), ivy vein banding virus (Petrzik, 2012), and soybean blotchy mosaic virus (Lamprecht et al., 2010). Furthermore, HTS could be a key tool to characterize the genomes of some cyto- and nucleorhabdoviruses which have only been characterized biologically, such as broccoli necrotic yellows virus (Lin and Campbell, 1972) and festuca leaf streak virus (Lundsgaard and Albrechtsen, 1976).

Moreover, partial genome fragments of three putative cytorhabdoviruses and one unassigned rhabdovirus were reported, when the viromes of water lily, common bean, *Lamprocephalus* sp. and kalanchoe were analyzed by HTS (Kreuze, 2014; Verdin et al., 2017; Bernardo et al., 2018; Mwaipopo et al., 2018). However, the obtained sequences are not available in any public database, so it is not possible to know if these viruses are novel or may correspond to already known viruses. Thus, it would be useful to apply HTS to complete the genome characterization of these viruses to increase our understanding of plant rhabdovirus diversity. Furthermore, since a tentative nucleorhabdovirus associated with papaya was recently discovered in a papaya virome (Alcalá-Briseño et al., 2020), it would be useful to employ HTS to characterize the molecular properties of papaya apical necrosis virus, a putative nucleorhabdovirus associated with papaya, which was only characterized biologically almost four decades ago (Lastra and Quintero, 1981), to determine if both viruses are related.

Cytorhabdovirus

The application of HTS has facilitated the discovery of at least 17 novel cytorhabdoviruses during the last 7 years (**Table 1**). Furthermore, the use of HTS has allowed characterization of the genomes of five cytorhabdoviruses, for which only a fragment or only the biological properties were known (**Table 2**). The template mostly used as a source for HTS of these viruses was total RNA, sequenced usually on Illumina platforms (**Tables 1**, **2**).

Alpha-, Beta-, Gammanucleorhabdovirus

The application of HTS has facilitated the discovery of ten novel nucleorhabdoviruses during the last 4 years, four alphanucleorhabdoviruses and six betanucleorhabdoviruses

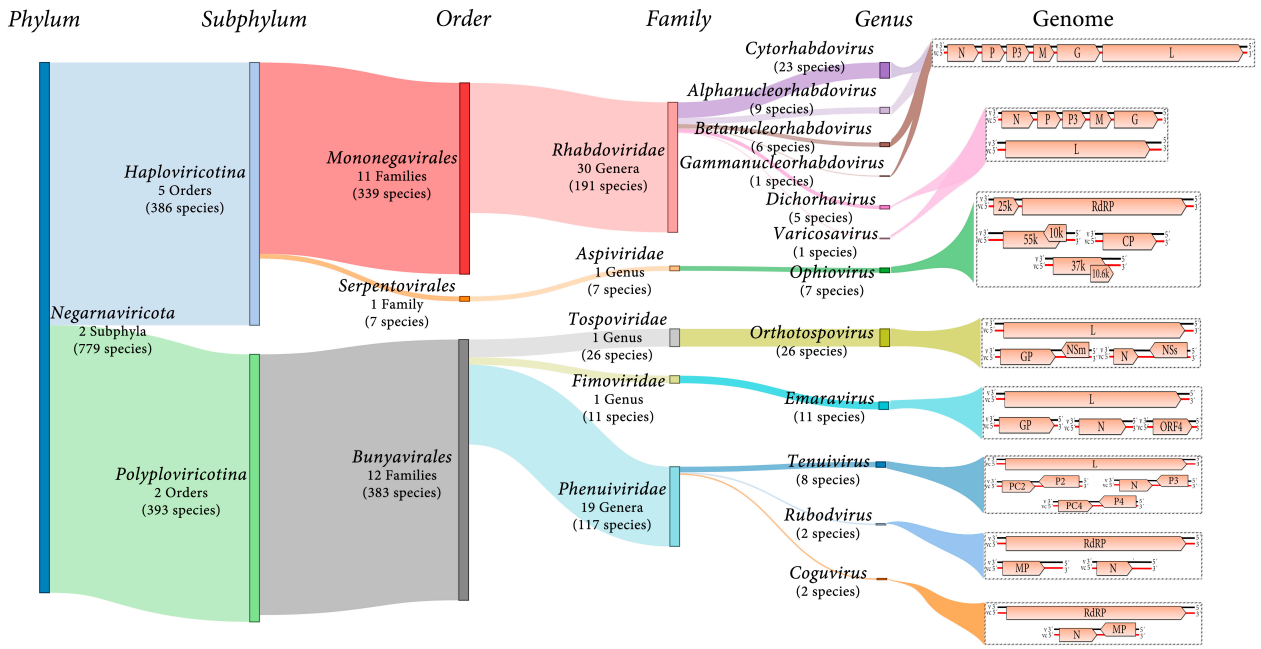


FIGURE 1 | Diagram showing an overview of plant NSR virus taxonomy and schematic genome graphs depicting organization and gene products of representative members of each taxon. The predicted coding sequences are shown in orange arrowed boxes, virion-sense RNA (v) is depicted in black, while virion-complementary sense RNA (vc) is depicted in red. The number of taxonomic categories included in each taxon are indicated. Species number indicate 2019 ICTV-accepted virus species corresponding to each taxonomic rank.

(**Table 1**). Furthermore, the use of HTS has allowed re-sequencing the genomes of two alphanucleorhabdoviruses which had previously been determined using Sanger dideoxy sequencing (maize mosaic virus and maize Iranian mosaic virus). HTS also allowed sequencing of the genome of a betanucleorhabdovirus for which only a genome fragment and the biological properties were known (datura yellow vein virus), and one other for which only the biological properties had been investigated 30 years earlier (sowthistle yellow vein virus) (**Table 2**). The template mostly used as a source for the HTS was total RNA, sequenced usually with Illumina instruments (**Tables 1**, **2**).

Dichorhavirus

The application of HTS has facilitated the discovery of three novel dichorhaviruses during the last 3 years (**Table 1**). Furthermore, the use of HTS has allowed characterization of the genome of a dichorhavirus for which only a fragment and the biological properties were known (coffee ringspot virus) (**Table 2**). The template used as a source for the HTS was total RNA, sequenced on Illumina platforms (**Tables 1**, **2**).

Varicosavirus

The complete genome of only one varicosavirus, lettuce big-vein associated virus, was previously characterized using Sanger dideoxy sequencing (Sasaya et al., 2002, 2004); however, the application of HTS has facilitated the discovery of two novel varicosaviruses during the last 3 years (**Table 1**). The template used as a source for the HTS was total RNA and dsRNA, and Illumina and Roche 454 were the technologies employed in the HTS projects (**Table 1**).

Fimoviridae

This family is composed of only one genus, *Emaravirus*, for viruses that are distantly related to orthotospoviruses, and exclusively comprise members that have plants as their hosts. Emaraviruses have enveloped, spherical virions, with a diameter of 80–100 nm and a segmented, linear, single-stranded genome with generally four to eight RNA segments (Elbeaino et al., 2018), but a recently described novel emaravirus, perilla mosaic virus, has 10 RNA segments (Kubota et al., 2020).

Most of the members of the genus *Emaravirus* have been recently discovered (**Table 1**) or characterized in depth by HTS techniques (**Table 2**). The application of HTS resulted in the discovery of 13 emaraviruses during the last 5 years (**Table 1**). Different templates, such as total RNA, small RNAs and dsRNA were used as sources for the HTS, and Illumina platforms have been mostly used in the HTS projects (**Table 1**).

The use of HTS has revealed the genome sequence of two emaraviruses for which only a genomic fragment and some biological properties were known (high plains wheat mosaic virus and pigeonpea sterility mosaic virus 1) (**Table 2**). Furthermore, the application of HTS has resulted in the detailed characterization of the complete genomes of three emaraviruses. A clearer picture of the genome organization of emaraviruses was obtained by the identification of additional RNA segments. HTS- based research resulted in the identification of three novel genome segments of raspberry leaf blotch virus and rose rosette

virus (Di Bello et al., 2015; Lu et al., 2015), and two novel segments for European mountain ash ringspot-associated virus genome (von Bargen et al., 2019). Different templates, including total nucleic acids, total RNA, dsRNA and RNA from partially purified virions were used as sources for the HTS. Illumina platforms were mostly used in these HTS projects (**Table 2**).

Finally, partial fragments of RNAs1, 2 and 4 of a putative novel emaravirus infecting alfalfa in Australia, named alfalfa ringspot-associated virus were recently identified using HTS (Samarfard et al., 2020). PCR amplification of the conserved termini of the genome segments (Babu et al., 2016) and HTS could be used to characterize the complete genome sequence of this tentative novel emaravirus.

Aspiviridae

This family, formerly named *Ophioviridae*, contains only one genus, *Ophiovirus*, which is exclusively composed of members that have plants as their hosts (**Figure 1**). Ophioviruses have non-enveloped, naked filamentous virions and a segmented, linear open circle, serpentine, single-stranded RNA genome, consisting of three to four segments (García et al., 2017).

One novel ophiovirus, blueberry mosaic associated ophiovirus, has been discovered using HTS during the last 6 years (**Table 1**). Partial genome fragments of three ophioviruses, freesia sneak virus, ranunculus white mottle virus, and tulip mild mottle mosaic virus have been known for some years but their complete genomes remain elusive (García et al., 2017). HTS could be a crucial tool to obtain the complete genomes of these viruses and to expand our understanding of the genomic cues and evolutionary diversity of ophioviruses.

Tospoviridae

This family contains the single genus *Orthotospovirus*, which is exclusively composed of species for viruses that have plants as their hosts (**Figure 1**). Orthotospoviruses have enveloped, spherical virions of 80–120 nm diameter and are transmitted by thrips insects in which they also replicate. The genome of orthotospoviruses is segmented with three linear single-stranded RNA segments named large (L), middle (M), and small (S). The L RNA contains one open reading frame in negative-sense polarity, whereas the other two segments, M and S RNAs, are ambisense and have two open reading frames encoding proteins in opposite orientation (Oliver and Whitfield, 2016).

The application of HTS led to the discovery of two novel orthotospoviruses in 2019 and 2020 (**Table 1**). Total RNA was used as the template and Illumina platforms as the sequencing technology.

The use of HTS has also facilitated completion of the sequence of eight orthotospoviruses for which only one genome segment was available (mostly the S RNA) and their biological properties were known (**Table 2**). Different templates, including total RNA, small RNA and virus-enriched RNA from purified ribonucleocapsids were used as RNA source, and Illumina platforms were mostly used as sequencing technology in the HTS projects (**Table 2**).

At the time of this review, only partial genome sequence of four other reported orthotospoviruses are available: groundnut yellow spot virus (S RNA), lisianthus necrotic ringspot virus (S RNA), pepper necrotic spot virus (S RNA) and tomato necrotic ringspot virus (S and M RNAs) (Satyanarayana et al., 1998; Seepiban et al., 2011; Torres et al., 2012; Shimomoto et al., 2014); HTS could enable completion of the genome of these agronomically important viruses.

Phenuiviridae

Most phenuiviruses have enveloped particles with helical morphology, except tenuiviruses that have non-enveloped filamentous particles, and their genomes are comprised of two to four single-stranded, linear RNA segments. As of 2019, the family *Phenuiviridae* includes 19 ICTV-recognized genera (**Figure 1**). Only the established genus *Tenuivirus* and likely two recently accepted new genera *Coguvirus* and *Rubodvirus* (2019.026M.A.v1.Phenuiviridae_4gen79sp.xlsx) contain species of plant-infecting viruses (Navarro et al., 2018b; Diaz-Lara et al., 2019). Phylogenetic relationships with other phenuiviruses support their classification within this family (Diaz-Lara et al., 2019).

Coguvirus

This genus is composed of two previously unknown viruses that have bi-segmented genomes, which have been discovered using HTS of citrus (Navarro et al., 2018a,b), while two tentative coguviruses that have tri-segmented genomes, have been discovered using HTS in watermelon (Xin et al., 2017) (**Table 1**). Total RNA and small RNAs were used as template sources for HTS on Illumina platforms (**Table 1**). Recently, the complete tri-segmented genomes of four cogu-like viruses were determined when the virome of grapevine was characterized (Bertazzon et al., 2020; Chiapello et al., 2020) (**Table 1**).

Moreover, partial segments of a cogu-like virus were discovered when the virome of ornamental flowers was analyzed by HTS (Wylie et al., 2019). However, it will be necessary to complete the genome characterization and perform further analyses of these viruses to confirm if they are in fact plant-infecting viruses or arthropod-infecting viruses whose RNA was co-purified along with the plant RNA. This additional information would enable further characterization of these putative coguviruses, thus increasing our understanding of the plant-infecting phenuivirus diversity and evolutionary history.

Rubodvirus

This genus is composed of species for two previously unknown viruses that have tri-segmented genomes which have been discovered using HTS in apple (Rott et al., 2018; Wright et al., 2018), while two tentative rubodviruses were discovered using HTS in grapevine (Diaz-Lara et al., 2019) (**Table 1**). DsRNA and total RNA were the templates used as sources for HTS on Illumina platforms (**Table 1**).

Tenuivirus

Tenuiviruses have a segmented genome of four to eight single-stranded RNA segments with negative-sense or ambisense polarity (Gaafar et al., 2019a). The application of HTS led to the discovery of two new tenuiviruses, one infecting melon and

black medic (Gaafar et al., 2019a; Lecoq et al., 2019) and the other from sugarcane associated with Ramu stunt disease (Mollov et al., 2016) (**Table 1**). In addition, a recent HTS study resulted in the molecular characterization of a wheat tenuivirus, the biological properties of which had been previously characterized in the 1960s (Sõmera et al., 2020) (**Table 1**). Total RNA and small RNAs were used as sources for HTS on Illumina platforms (**Table 1**). Moreover, the use of HTS facilitated sequencing of a tenuivirus from rice, for which only the sequence of RNA4 and its biological properties were known previously (**Table 2**). At the time of writing this review, only partial genomic sequence of four other tenuiviruses are known: maize stripe virus (complete sequences of RNAs 2, 3, 4 and 5), maize yellow stripe virus (partial sequences of RNAs 1, 2, 3, 4 and complete sequence of RNA 5), Iranian wheat stripe virus (complete sequences of RNAs 2, 3 and 4), wheat yellow head virus (complete sequence of the nucleoprotein gene) (Huiet et al., 1991; Estabrook et al., 1996; Seifers et al., 2005; Heydarnejad et al., 2006; Mahmoud et al., 2007). HTS would be an essential tool to obtain the complete genomes of these viruses and expand our knowledge of genomic architecture of tenuiviruses.

CONCLUSION AND PERSPECTIVES

The number of NSR plant viruses identified so far represents a negligible fraction of the potential number of NSR viruses present in the virosphere. The discovery and in-depth molecular characterization of these viruses has been challenging given their extensive divergence, their outstanding diversity in terms of genomic architecture, and mostly, the negligible share of plant species studied for virus discovery. However, the widespread implementation of HTS, a cost-effective and efficient technology platform, allowed the discovery and molecular characterization of the genome sequences of at least 70 NSR plant viruses in the last few years. Thus, we predict that the increasing use of HTS, not only for plant samples but also in arthropod vectors, will allow the identification of many novel NSR phytoviruses which will be crucial to unravel the evolutionary landscapes of many NSR virus clades that are poorly characterized today. Nevertheless, as many more novel NSR viruses are being discovered, careful analysis will be essential to confirm their correct ecological context, to determine whether the new viral agents are plant-infecting viruses, fungi- or arthropod-infecting viruses whose RNA was co-purified along with the plant RNA, or just contamination-inadvertently sampled environmental viruses.

As soon as the virome characterization of multiple samples of cultivated and wild plants becomes more routine, more viruses with unique features and novel phylogenetic relationships will likely be discovered. As we assess the viral 'dark matter,' we will start to grasp the evolutionary history of plant viruses, eventually leading us to untangle the diversity of the NSR virosphere and gain increased knowledge about its complex evolutionary and phylogenetic relationships. A rich picture of the plant virome landscape will also provide an essential framework for improved taxonomic classification of plant NSR viruses as part of the realm *Riboviria*. In conclusion, HTS-based methods for virus discovery, our 'new eyes,' are unraveling in real time the richness and magnitude of the plant RNA virosphere.

AUTHOR CONTRIBUTIONS

All authors wrote the manuscript and approved it for publication.

REFERENCES

Adegbola, R., Kemerait, R., Adkins, S., and Naidu, R. (2019). Complete genome segment sequences of tomato chlorotic spot virus from peanut in Haiti. *Microbiol. Resour. Announc.* 8:e00306-19.

Alcalá-Briseño, R. I., Casarrubias-Castillo, K., López-Ley, D., Garrett, K. A., and Silva-Rosales, L. (2020). Network analysis of the papaya orchard virome from two agroecological regions of Chiapas, Mexico. *mSystems* 5:e00423-19.

Alves-Freitas, D. M. T., Pinheiro-Lima, B., Faria, J. C., Lacorte, C., Ribeiro, S. G., and Melo, F. L. (2019). Double-stranded RNA high-throughput sequencing reveals a new cytorhabdovirus in a bean golden mosaic virus-resistant common bean transgenic line. *Viruses* 11:90. doi: 10.3390/v11010090

Babu, B., Washburn, B. K., Poduch, K., Knox, G. W., and Paret, M. L. (2016). Identification and characterization of two novel genomic RNA segments RNA5 and RNA6 in rose rosette virus infecting roses. *Acta Virol.* 60, 156–165. doi: 10.4149/av_2016_02_156

Baek, D., Lim, S., Ju, H., Kim, H., Lee, S., and Moon, J. (2019). The complete sequence of a novel nucleorhabdovirus (apple rootstock virus A) identified in apple rootstock. *Arch. Virol.* 164, 2641–2644. doi: 10.1007/s00705-019-04348-0

Bejerman, N., Acevedo, R., de Breuil, S., Ruiz, O., and Sansberro, P. (2020). Molecular characterization of a novel cytorhabdovirus with a unique genomic organization infecting yerba mate (*Ilex paraguariensis*) in Argentina. *Arch. Virol.* 165, 1475–1479. doi: 10.1007/s00705-020-04609-3

Bejerman, N., de Breuil, S., Debat, H., Miretti, M., Badaracco, A., and Nome, C. (2017). Molecular characterization of yerba mate chlorosis-associated virus, a putative cytorhabdovirus infecting yerba mate (*Ilex paraguariensis*). *Arch. Virol.* 162, 2481–2484. doi: 10.1007/s00705-017-3363-8

Bejerman, N., Giolitti, F., de Breuil, S., Trucco, V., Nome, C., Lenardon, S., et al. (2015). Complete genome sequence and integrated protein localization and interaction map for alfalfa dwarf virus, which combines properties of both cytoplasmic and nuclear plant rhabdoviruses. *Virology* 483, 275–283. doi: 10.1016/j.virol.2015.05.001

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

Bertazzon, N., Chitarra, W., Angelini, E., and Nerva, L. (2020). Two new putative plant viruses from wood metagenomics analysis of an esca diseased vineyard. *Plants* 9:835. doi: 10.3390/plants9070835

Bhat, A. I., Pamitha, N. S., Naveen, K. P., and Biju, C. N. (2020). Identification and characterization of cardamom vein clearing virus, a novel aphid-transmitted nucleorhabdovirus. *Eur. J. Plant Pathol.* 156, 1053–1062. doi: 10.1007/s10658-020-01958-2

Blawid, R., Silva, J. M. F., and Nagata, T. (2017). Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. *Ann. Appl. Biol.* 170, 301–314. doi: 10.1111/aab.12345

Blouin, A. G., Ross, H. A., Hobson-Peters, J., O'Brien, C., Warren, B., and MacDiarmid, R. (2016). A new virus discovered by immunocapture of double-stranded RNA, a rapid method for virus enrichment in metagenomic studies. *Mol. Ecol. Resour.* 16, 1255–1263. doi: 10.1111/1755-0998.12525

Buzkan, N., Chiumenti, M., Massart, S., Sarpkaya, K., Karadağ, S., and Minafra, A. (2019). A new emaravirus discovered in *Pistacia* from Turkey. *Virus Res.* 26, 159–163. doi: 10.1016/j.virusres.2019.01.012

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

Chabi-Jesus, C., Ramos-Gonzalez, P., Tassi, A. D., Guerra-Peraza, O., Kitajima, E. W., Harakava, R., et al. (2018). Identification and characterization of citrus chlorotic spot virus, a new dichorhavirus associated with citrus leprosis-like symptoms. *Plant Dis.* 102, 1509–1519.

Chiapello, M., Rodríguez-Romero, J., Nerva, L., Forgia, M., Chitarra, W., Ayllón, M. A., et al. (2020). Putative new plant viruses associated with *Plasmopara viticola*-infected grapevine samples. *Ann. Appl. Biol.* 176, 180–191. doi: 10.1111/aab.12563

Chou, W. C., Lin, S. S., Yeh, S. D., Li, S. L., Peng, Y. C., Fan, Y. H., et al. (2017). Characterization of the genome of a phylogenetically distinct tospovirus and its interactions with the local lesion-induced host *Chenopodium quinoa* by whole-transcriptome analyses. *PLoS One* 12:e0182425. doi: 10.1371/journal.pone.0182425

Ciuffo, M., Nerva, L., and Turina, M. (2017). Full-length genome sequence of the tospovirus melon severe mosaic virus. *Arch. Virol.* 162, 1419–1422. doi: 10.1007/s00705-017-3237-0

Debat, H. J., and Bejerman, N. (2019). Novel bird's-foot trefoil RNA viruses provide insights into a clade of legume-associated enamoviruses and rhabdoviruses. *Arch. Virol.* 164, 1419–1426. doi: 10.1007/s00705-019-04193-1

Di Bello, P. L., Ho, T., and Tzanetakis, I. E. (2015). The evolution of emaraviruses is becoming more complex: seven segments identified in the causal agent of Rose rosette disease. *Virus Res.* 210, 241–244. doi: 10.1016/j.virusres.2015.08.009

Di Bello, P. L., Laney, A. G., Druciarek, T., Ho, T., and Gergerich, R. C. (2016). A novel emaravirus is associated with redbud yellow ringspot disease. *Virus Res.* 222, 41–47. doi: 10.1016/j.virusres.2016.05.027

Diaz-Lara, A., Navarro, B., Di Serio, F., Stevens, K., Hwang, M. S., Kohl, J., et al. (2019). Two novel negative-sense RNA viruses infecting grapevine are members of a newly proposed genus within the family *Phenuiviridae*. *Viruses* 11:685. doi: 10.3390/v11080685

Dietzgen, R. G., and Higgins, C. (2019). Complete genome sequence of maize sterile stunt virus. *Arch. Virol.* 164, 1221–1223. doi: 10.1007/s00705-019-04164-6

Dietzgen, R. G., Innes, D. J., and Bejerman, N. (2015). Complete genome sequence and intracellular protein localization of Datura yellow vein nucleorhabdovirus. *Virus Res.* 205, 7–11. doi: 10.1016/j.virusres.2015.05.001

Ding, X., Chen, D., Du, Z., Zhang, J., and Wu, Z. (2019). The complete genome sequence of a novel cytorhabdovirus identified in strawberry (*Fragaria ananassa* Duch.). *Arch. Virol.* 164, 3127–3131. doi: 10.1007/s00705-019-04390-y

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

Dullemans, A. M., Verhoeven, J. T., Kormelink, R., and van der Vlugt, R. A. (2015). The complete nucleotide sequence of chrysanthemum stem necrosis virus. *Arch. Virol.* 160, 605–608. doi: 10.1007/s00705-014-2282-1

Elbeaino, T., Digiaro, M., Mielke-Ehret, N., Mühlbach, H.-P., and Martelli, G. P. (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

Estabrook, E. M., Suyenaga, K., Tsai, J. H., and Falk, B. W. (1996). Maize stripe tenuivirus RNA2 transcripts in plant and insect analysis of pvc2, a protein similar to the *Phlebovirus* virion glycoproteins. *Virus Genes* 12, 239–247. doi: 10.1111/j.1365-3059.1995.tb02774.x

Fagundes Silva, J., Silvas de Oliveira, A., Severo de Almeida, M. M., Kormelink, R., Nagata, T., and Resende, R. O. (2019). Tomato chlorotic spot virus (TCSV) putatively incorporated a genomic segment of Groundnut ringspot virus (GRSV) upon a reassortment event. *Viruses* 11, 187–194. doi: 10.3390/v11020187

Franova, J., Pøibylova, J., and Koloniuk, I. (2019a). Molecular and biological characterization of a new strawberry cytorhabdovirus. *Viruses* 11:982. doi: 10.3390/v11110982

Franova, J., Sarkisová, T., Jakešová, H., and Koloniuk, I. (2019b). Molecular and biological properties of two putative new cytorhabdoviruses infecting *Trifolium pratense*. *Plant Pathol.* 11, 1276–1286. doi: 10.1111/ppa.13065

Freitas-Astúa, J., Dietzgen, R. G., Walker, P. J., Blasdell, K. R., and Breyta, S. (2019). *Split the Genus Nucleorhabdovirus, Creating Three New Genera (Alphanucleorhabdovirus, Betanucleorhabdovirus and Gammanucleorhabdovirus) Comprising Sixteen Species, Including 23 Six New Species, in the Family Rhabdoviridae*. London: International Committee on Taxonomy of Viruses.

Gaafar, Y., Abdelgalil, M., Knierim, D., Richert-Pöggeler, K., and Menzel, W. (2018). First report of physostegia chlorotic mottle virus on tomato (*Solanum lycopersicum*) in Germany. *Plant Dis.* 102:255. doi: 10.1094/pdis-05-17-0737-pdn

Gaafar, Y., Richert-Poggeler, K., Sieg-Muller, A., Luddecke, P., and Herz, K. (2019a). A divergent strain of melon chlorotic spot virus isolated from black medic (*Medicago lupulina*) in Austria. *Virol. J.* 16:89.

Gaafar, Y., Richert-Pöggeler, K. R., Maaß, C., Vetten, H. J., and Ziebell, H. (2019b). Characterization of a novel nucleorhabdovirus infecting alfalfa (*Medicago sativa*). *Virol. J.* 16:55. doi: 10.1007/978-1-4020-6754-9_468

Gallo, Y., Toro, L. F., Jaramillo, H., Gutiérrez, P., and Marín, M. (2018). Identificación y caracterización molecular del genoma completo de tres virus en cultivos de lulo (*Solanum quitoense*) de Antioquia (Colombia). *Rev. Colomb. Cienc. Hortíc.* 1, 281–292. doi: 10.17584/rcch.2018v12i2.7692

García, M. L., Dal Bó, E., da Graça, J. V., Gago-Zachert, S., and Hammond, J. (2017). ICTV virus taxonomy profile: *Ophioviridae*. *J. Gen. Virol.* 98, 1161–1162. doi: 10.1099/jgv.0.000836

German, T. L., Lorenzen, M. D., Grubbs, N., and Whitefield, A. E. (2020). New technologies for studying negative-strand RNA viruses in plant and arthropod hosts. *Mol. Plant Microbe Interact.* 33, 382–393. doi: 10.1094/mpmi-10-19-0281-fi

Ghorbani, A., Izadpanah, K., and Dietzgen, R. G. (2018). Completed sequence and corrected annotation of the genome of maize Iranian mosaic virus. *Arch. Virol.* 163, 767–770. doi: 10.1007/s00705-017-3646-0

Goh, C., Park, D., and Hahn, Y. (2020). Identification of *Trichosanthes* associated rhabdovirus 1, a novel member of the genus cytorhabdovirus of the family *Rhabdoviridae*, in the *Trichosanthes kirilowii* transcriptome. *Acta Virol.* 64, 36–43. doi: 10.4149/av_2020_105

Hassan, M., Di Bello, P. L., Keller, K. E., Martin, R. R., Sabanadzovic, S., and Tzanetakis, I. E. (2017). A new, widespread emaravirus discovered in blackberry. *Virus Res.* 235, 1–5. doi: 10.1016/j.virusres.2017.04.006

Hassani-Mehraban, A., Dullemans, A. M., Verhoeven, J. T. J., Roenhorst, J. W., and Peters, D. (2019). Alstroemeria yellow spot virus (AYSV): a new orthotospovirus species within a growing Eurasian clade. *Arch. Virol.* 164, 117–126. doi: 10.1007/s00705-018-4027-z

Heydarnejad, J., Barclay, W. S., Hunter, F. R., Izadpanah, K., and Gooding, M. J. (2006). Molecular characterization of Iranian wheat stripe virus shows its taxonomic position as a distinct species in the genus *Tenuivirus*. *Arch. Virol.* 151, 217–227. doi: 10.1007/s00705-005-0652-4

Higgins, C. M., Bejerman, N., Li, M., James, A. P., and Dietzgen, R. G. (2016). Complete genome sequence of Colocasia bobone disease-associated virus, a putative cytorhabdovirus infecting taro. *Arch. Virol.* 161, 745–748. doi: 10.1007/s00705-015-2713-7

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 RNA-3. *Virology* 182, 47–53. doi: 10.1016/0042-6822(91)90646-s

Ilyas, M., Avelar, A. S., Schuch, U., and Brown, J. K. (2018). First report of an emaravirus associated with witches broom disease and eriophyid mite infestations of the blue palo verde tree in Arizona. *Plant Dis.* 102:1863. doi: 10.1094/pdis-01-18-0124-pdn

Ito, T., Suzaki, K., and Nakano, M. (2013). Genetic characterization of novel putative rhabdovirus and dsRNA virus from Japanese persimmon. *J. Gen. Virol.* 94, 1917–1921. doi: 10.1099/vir.0.054445-0

Jakovljevic, V., Otten, P., Berwarth, C., and Jelkmann, W. (2016). Analysis of the apple rubbery wood disease by next generation sequencing of total RNA. *Eur. J. Plant Pathol.* 148, 637–646. doi: 10.1007/s10658-016-1119-z

Jimenez, J., Carvajal-Yepes, M., Leiva, A. M., Cruz, M., Romero, L. E., Bolaños, C. A., et al. (2018). Complete genome sequence of *Rice hoja blanca tenuivirus* isolated from a susceptible rice cultivar in Colombia. *Genome Announc.* 6:e01490-17.

Jones, S., McGavin, W., and MacFarlane, S. (2019). The complete sequences of two divergent variants of the rhabdovirus raspberry vein chlorosis virus and the design of improved primers for virus detection. *Virus Res.* 265, 162–165. doi: 10.1016/j.virusres.2019.03.004

Käfer, S., Paraskevopoulou, S., Zirkel, F., Wieseke, N., Donath, A., Peterson, M., et al. (2019). Re-assessing the diversity of negative strand RNA viruses in insects. *PLoS Pathog.* 15:e1008224. doi: 10.1371/journal.ppat.1008224

Koloniuk, I., Fránová, J., Sarkisova, T., Pøibylová, J., and Lenz, O. (2018a). Identification and molecular characterization of a novel varicosa-like virus from red clover. *Arch. Virol.* 163, 2213–2218. doi: 10.1007/s00705-018-3838-2

Koloniuk, I., Fránová, J., Sarkisova, T., and Pøibylová, J. (2018b). Complete genome sequences of two divergent isolates of strawberry crinkle virus coinfecting a single strawberry plant. *Arch. Virol.* 163, 2539–2542. doi: 10.1007/s00705-018-3860-4

Koonin, E. V., Dolja, V. V., and Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* 479–480, 2–25. doi: 10.1016/j.virol.2015.02.039

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., et al. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* 84, 1–33.

Kreuze, J. (2014). "siRNA deep sequencing and assembly: piecing together viral infections," in *Detection and Diagnostics of Plant Pathogens, Plant Pathology in the 21st Century 5*, eds M. L. Gullino and P. J. M. Bonants (Dordrecht: Springer Science+Business Media), 21–38. doi: 10.1007/978-94-017-9020-8_2

Kubota, K., Usugi, T., Tomitaka, Y., Shimomoto, Y., and Takeuchi, S. (2020). Perilla mosaic virus is a highly divergent emaravirus transmitted by *Shevtchenkella* Sp. (Acari: Eriophyidae). *Phytopatology* 110, 1352–1361. doi: 10.1094/PHYTO-01-20-0013-R

Kuhn, J. H., Adkins, S., and Alioto, D. (2020). 2020 taxonomic update for phylum *Negarnaviricota* (*Riboviria: Orthornavirae*), including the large orders *Bunyavirales* and *Mononegavirales*. *Arch. Virol.* (in press). doi: 10.1007/s00705-020-04731-2

Kuhn, J. H., Wolf, Y. I., Krupovic, M., Zhang, Y.-Z., Maes, P., Dolja, V. V., et al. (2019). Classify viruses—the gain is worth the pain. *Nature* 566, 318–320. doi: 10.1038/d41586-019-00599-8

Kumar, S., Subbarao, B., and Hallan, V. (2017). Molecular characterization of emaraviruses associated with Pigeonpea sterility mosaic disease. *Sci. Rep.* 7:11831.

Lamprecht, R. L., Kasdorf, G. G. F., Stiller, M., Staples, S. M., Nel, L. H., and Pietersen, G. (2010). Soybean blotchy mosaic virus, a new *Cytorhabdovirus* found in South Africa. *Plant Dis.* 94, 1348–1354. doi: 10.1094/pdis-09-09-0598

Lastra, R., and Quintero, E. (1981). Papaya apical necrosis, a new diseases associated with a rhabdovirus. *Plant Dis.* 65, 439–440. doi: 10.1094/pd-65-439

Lecoq, H., Wipf-Scheibel, C., Verdin, E., and Desbiez, C. (2019). Characterization of the first tenuivirus naturally infecting dicotyledonous plants. *Arch. Virol.* 164, 297–301. doi: 10.1007/s00705-018-4057-6

Li, C. X., Shi, M., Tian, J. H., Lin, X. D., Kang, Y. J., Chen, L. J., et al. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* 4:e05378.

Liao, L., and Noble, W. S. (2003). Combining pairwise sequence similarity and support vector machines for detecting remote protein evolutionary and structural relationships. *J. Comput. Biol.* 10, 857–868.

Lima, R. N., De Oliveira, A. S., Leastro, M. O., Blawid, R., Nagata, T., Resende, R. O., et al. (2016). The complete genome of the tospovirus *Zucchini lethal chlorosis virus*. *Virol. J.* 13:123.

Lin, M. T., and Campbell, R. N. (1972). Characterization of broccoli necrotic yellows virus. *Virology* 48, 30–40. doi: 10.1016/0042-6822(72)90111-0

Liu, Y., Du, Z., Wang, H., Zhang, S., Cao, M., and Wang, X. (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

Lu, Y., McGavin, W., Cock, P. J., Schnettler, E., Yan, F., Chen, J., et al. (2015). Newly identified RNAs of raspberry leaf blotch virus encoding a related group of proteins. *J. Gen. Virol.* 96, 3432–3439. doi: 10.1099/jgv.0.000277

Lundsgaard, T., and Albrechtsen, S. E. (1976). Electron microscopy of rhabdovirus-like particles in *Festuca gigantea* with leaf streak mosaic. *J. Phytopathol.* 87, 12–16. doi: 10.1111/j.1439-0434.1976.tb01715.x

Ma, Y., Marais, A., Lefebvre, M., Faure, C. H., and Candresse, T. (2020). Metagenomic analysis of virome cross-talk between cultivated *Solanum lycopersicum* and wild *Solanum nigrum*. *Virology* 540, 38–44. doi: 10.1016/j.virol.2019.11.009

Mahmoud, A., Royer, M., Granier, M., Ammar, E.-D., Thouvenel, J. C., and Peterschmitt, M. (2007). Evidence for a segmented genome and partial nucleotide sequences of Maize yellow stripe virus, a proposed new *Tenuivirus*. *Arch. Virol.* 152, 1757–1762. doi: 10.1007/s00705-007-0980-7

Margaria, P., Miozzi, L., Ciuffo, M., Pappu, H., and Turina, M. (2014). The complete genome sequence of polygonum ringspot virus. *Arch. Virol.* 159, 3149–3152. doi: 10.1007/s00705-014-2166-4

Martin, K., and Whitfield, A. (2019). Complete genome sequence of *Maize Mosaic Nucleorhabdovirus*. *Microbiol. Resour. Announc.* 8:e00637-19.

Martínez, R. T., de Almeida, M. M. S., Rodriguez, R., de Oliveira, A. S., Melo, F. L., and Resende, R. O. (2018). Identification and genome analysis of tomato chlorotic spot virus and dsRNA viruses from coinfected vegetables in the Dominican Republic by high-throughput sequencing. *Virol. J.* 15:24.

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

Maurino, F., Dumón, A. D., Llauger, G., Alemandri, V., and de Haro, L. (2018). Complete genome sequence of maize yellow striate virus, a new cytorhabdovirus infecting maize and wheat crops in Argentina. *Arch. Virol.* 163, 291–295. doi: 10.1007/s00705-017-3579-7

Medina-Salguero, A. X., Cornejo-Franco, J. F., Grinstead, S., Mollov, D., Mowery, J. D., Flores, F., et al. (2019). Sequencing, genome analysis and prevalence of a cytorhabdovirus discovered in *Carica papaya*. *PLoS One* 14:e0215798. doi: 10.1371/journal.pone.0215798

Meng, J., Liu, P., Zhu, L., Zou, C., Li, J., and Chen, B. (2015). Complete genome sequence of mulberry vein banding associated virus, a new tospovirus infecting mulberry. *PLoS One* 10:e0136196. doi: 10.1371/journal.pone.0136196

Menzel, W., Knierim, D., Richert-Pöggeler, K., and Winter, S. (2016). Charakterisierung eines nucleorhabdovirus aus physostegia. *Julius Kühn Arch.* 454, 283–284.

Mollov, D., Maroon-Lango, C., and Kuniata, L. (2016). Detection by next generation sequencing of a multi-segmented viral genome from sugarcane associated with Ramu stunt disease. *Virus Genes* 52, 152–155. doi: 10.1007/s11262-015-1279-5

Mwaipopo, B., Msolla, N. S., Njau, P., Mark, D., and Mbanzibwa, D. R. (2018). Comprehensive surveys of *Bean common mosaic virus* and *Bean common mosaic necrosis virus* and molecular evidence for occurrence of other *Phaseolus vulgaris* viruses in Tanzania. *Plant Dis.* 102, 2361–2370. doi: 10.1094/pdis-01-18-0198-re

Navarro, B., Minutolo, M., Stradis, A. D., Palmisano, F., Alioto, D., and Di Serio, F. (2018a). The first phlebo-like virus infecting plants: a case study on the adaptation of negative-stranded RNA viruses to new hosts. *Mol. Plant Pathol.* 19, 1075–1089. doi: 10.1111/mpp.12587

Navarro, B., Zicca, S., Minutolo, M., Saponari, M., Alioto, D., and Di Serio, F. (2018b). A negative-stranded RNA virus infecting citrus trees: the second member of a new genus within the order *Bunyavirales*. *Front. Microbiol.* 9:2340. doi: 10.3389/fmicb.2018.02340

Obbard, D. J., Shi, M., Roberts, K. E., Longdon, B., and Dennis, A. B. (2020). A new lineage of segmented RNA viruses infecting animals. *Virus Evol.* 6:vez061.

Oliver, J., and Whitfield, A. (2016). The genus *Tospovirus*: emerging bunyaviruses that threaten food security. *Annu. Rev. Virol.* 3, 101–124. doi: 10.1146/annurev-virology-100114-055036

Olmedo-Velarde, A., Park, A. C., Sugano, J., Uchida, J. Y., and Kawate, M. (2019). Characterization of Ti ringspot-associated virus, a novel Emaravirus associated with an emerging ringspot disease of *Cordyline fruticosa*. *Plant Dis.* 103, 2345–2352. doi: 10.1094/pdis-09-18-1513-re

Orfanidou, C. G., Beta, C., Reynard, J., Tsiolakis, G., Katis, N. I., and Maglioka, V. I. (2020). Identification, molecular characterization and prevalence of a novel cytorhabdovirus infecting zucchini crops in Greece. *Virus Res.* 287:198095. doi: 10.1016/j.virusres.2020.198095

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

Peracchio, C., Fogia, M., Chiapello, M., Vallino, M., Turina, M., and Ciuffo, M. (2020). A complex virome including two distinct emaraviruses associated with virus-like symptoms in *Camellia japonica*. *Virus Res.* 286:197964. doi: 10.1016/j.virusres.2020.197964

Petrzik, K. (2012). Bioinformatic analysis of the L polymerase gene leads to discrimination of new rhabdoviruses. *J. Phytopathol.* 160, 377–381. doi: 10.1111/j.1439-0434.2012.01919.x

Ramalho, T. O., Figueira, A. R., Sotero, A. J., Wang, R., Geraldino Duarte, P. S., Farman, M., et al. (2014). Characterization of coffee ringspot virus-Lavras: a model for an emerging threat to coffee production and quality. *Virology* 464–465, 385–396. doi: 10.1016/j.virol.2014.07.031

Ramos-González, P. L., Chabi-Jesus, C., Banguela-Castillo, A., Tassi, A. D., and Rodrigues, M. D. C. (2018). Unveiling the complete genome sequence of clerodendrum chlorotic spot virus, a putative dichorhavirus infecting ornamental plants. *Arch. Virol.* 163, 2519–2524. doi: 10.1007/s00705-018-3857-z

Ramos-González, P. L., Chabi-Jesus, C., Guerra-Peraza, O., Tassi, A. D., and Kitajima, E. W. (2017). Citrus leprosis virus N: a new dichorhavirus causing citrus leprosis disease. *Phytopathology* 107, 963–976. doi: 10.1094/phyto-02-17-0042-r

Read, D. A., Featherston, J., Rees, D. J. G., Thompson, G. D., and Roberts, R. (2019). Molecular characterization of Morogoro maize-associated virus, a nucleorhabdovirus detected in maize (*Zea mays*) in Tanzania. *Arch. Virol.* 164, 1711–1715. doi: 10.1007/s00705-019-04212-1

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. E., Kesanakurti, P., Berwarth, C., Rast, H., Boyes, I., Phelan, J., et al. (2018). Discovery of negative-sense RNA viruses in trees infected with apple rubbery wood disease by next-generation sequencing. *Plant Dis.* 102, 1254–1263.

Sabbadin, F., Glover, R., Stafford, R., Rozado-Aguirre, Z., and Boonham, N. (2017). Transcriptome sequencing identifies novel persistent viruses in herbicide resistant wild-grasses. *Sci. Rep.* 7:41987.

Sadeghi, M. S., Afsharifar, A., Izadpanah, K., Loconsole, G., De Stradis, A., Martelli, G. P., et al. (2016). Isolation and partial characterization of a novel cytorhabdovirus from citrus trees showing foliar symptoms in Iran. *Plant Dis.* 100, 66–71.

Samarfard, S., McTaggart, A. R., Sharman, M., Bejerman, N. E., and Dietzgen, R. G. (2020). Viromes of ten alfalfa plants in Australia reveal diverse known viruses and a novel RNA virus. *Pathogens* 9:214. doi: 10.3390/pathogens9030214

Sasaya, T., Ishikawa, K., and Koganezawa, H. (2002). The nucleotide sequence of RNA1 of *Lettuce big-vein virus*, genus *Varicosavirus*, reveals its relation to nonsegmented negative-strand RNA viruses. *Virology* 297, 289–297.

Sasaya, T., Kusaba, S., Ishikawa, K., and Koganezawa, H. (2004). Nucleotide sequence of RNA2 of *Lettuce big-vein virus* and evidence for a possible transcription termination/initiation strategy similar to that of rhabdoviruses. *J. Gen. Virol.* 85, 2709–2717.

Satyanarayana, T., Gowda, S., Reddy, K. L., Mitchell, S. E., and Dawson, W. O. (1998). Peanut yellow spot virus is a member of a new serogroup of *Tospovirus* genus based on small (S) RNA sequence and organization. *Arch. Virol.* 143, 353–364.

Schoelz, J. E., and Stewart, L. R. (2018). The role of viruses in the phytobiome. *Annu. Rev. Virol.* 5, 93–111.

Seepiban, C., Gajanandana, O., Attathom, T., and Attathom, S. (2011). Tomato necrotic ringspot virus, a new tospovirus isolated in Thailand. *Arch. Virol.* 156, 263–274.

Seifers, D. L., Harvey, T. L., Martin, T. J., Haber, S., and She, Y.-M. (2005). Association of a virus with wheat displaying yellow head disease symptoms in the Great Plains. *Plant Dis.* 89, 888–895.

Shimomoto, Y., Kobayashi, K., and Okuda, M. (2014). Identification and characterization of Lisianthus necrotic ringspot virus, a novel distinct tospovirus species causing necrotic disease of lisianthus (*Eustoma grandiflorum*). *J. Gen. Plant Pathol.* 80, 169–175.

Simmonds, P., Adams, M. J., Benkō, M., Breitbart, M., Brister, J. R., and Carstens, E. B. (2017). Consensus statement: virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* 15, 161–168.

Skewes-Cox, P., Sharpton, T. J., Pollard, K. S., and DeRisi, J. L. (2014). Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *PLoS One* 9:e105067. doi: 10.1371/journal.pone.0105067

Sõmera, M., Kvarnheden, A., Desbiez, C., Blystad, D., and Sooväli, P. (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.

Stenger, D., Burbank, L., Wang, R., Stewart, A., Mathias, C., and Goodin, M. (2020). Lost and found: rediscovery and genomic characterization of sowthistle yellow vein virus after a 30+ year hiatus. *Virus Res.* 284:197987.

Susi, H., Filloux, D., Frilander, M. J., Roumagnac, P., and Laine, A.-L. (2019). Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. *PeerJ* 7:e6140.

Tatineni, S., McMechan, A. J., Wosula, E. N., Wegulo, S. N., and Graybosch, R. A. (2014). An eriophyid mite-transmitted plant virus contains eight genomic RNA segments with unusual heterogeneity in the nucleocapsid protein. *J. Virol.* 88, 11834–11845.

Thekke-Veetil, T., Ho, T., Keller, K. E., Martin, R. R., and Tzanetakis, I. E. (2014). A new ophiovirus is associated with blueberry mosaic disease. *Virus Res.* 189, 92–96.

Torres, R., Larenas, J., Fribourg, C., and Romero, J. (2012). Pepper necrotic spot virus, a new tospovirus infecting solanaceous crops in Peru. *Arch. Virol.* 157, 609–615.

Verdin, E., Wipf-Scheibel, C., Gognalons, P., Aller, F., Jacquemond, M., and Tepfer, M. (2017). Sequencing viral siRNAs to identify previously undescribed viruses and viroids in a panel of ornamental plant samples structured as a matrix of pools. *Virus Res.* 241, 19–28.

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109, 716–725.

von Bargen, S., Al Kubrusli, R., Gaskin, T., Fürl, S., and Hüttner, F. (2020). Characterisation of a novel emaravirus identified in mosaic-diseased Eurasian aspen (*Populus tremula*). *Ann. Appl. Biol.* 176, 210–222.

von Bargen, S., Dieckmann, H. L., Candresse, T., Mühlbach, H. P., Roßbach, J., and Büttner, C. (2019). Determination of the complete genome of European mountain ash ringspot-associated emaravirus from *Sorbus intermedia* reveals two additional genome segments. *Arch. Virol.* 164, 1937–1941.

Walker, P. J., Blasdell, K. R., Calisher, C. H., Dietzgen, R. G., and Kondo, H. (2018). ICTV virus taxonomy profile: *Rhabdoviridae*. *J. Gen. Virol.* 99, 447–448.

Wang, Y.-Q., Song, Y., Cao, M., Cheng, Q., Wu, J., and Hu, T. (2020). Identification of a novel emaravirus infecting lilac through next generation sequencing. *J. Integr. Agric.* 19, 2064–2071.

Whitfield, A. E., Huot, O. B., Martin, K. M., Kondo, H., and Dietzgen, R. G. (2018). Plant rhabdoviruses: their origins and vector interactions. *Curr. Opin. Virol.* 33, 198–207.

Willie, K., and Stewart, L. R. (2017). Complete genome sequence of a new maize-associated cytorhabdovirus. *Genome Announc.* 5:e00591-17.

Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucia-Sanz, A., Kuhn, J. H., Krupovic, M., et al. (2018). Origins and evolution of the global RNA virome. *mBio* 9:e02329-18.

Wren, J. D., Roossinck, M. J., Nelson, R. S., Sheets, K., Palmer, M. W., and Melcher, U. (2006). Plant virus biodiversity and ecology. *PLoS Biol.* 4:e80. doi: 10.1371/journal.pbio.0040080

Wright, A. A., Szostek, S. A., Beaver-Kanuya, E., and Harper, S. J. (2018). Diversity of three bunya-like viruses infecting apple. *Arch. Virol.* 163, 3339–3343.

Wu, L. P., Yang, T., Liu, H. W., Postman, J., and Li, R. (2018). Molecular characterization of a novel rhabdovirus infecting blackcurrant identified by high-throughput sequencing. *Arch. Virol.* 162, 2493–2494.

Wu, Q., Ding, S. W., Zhang, Y., and Zhu, S. (2015). Identification of viruses and viroids by next-generation sequencing and homology- dependent and homology- independent algorithms. *Annu. Rev. Phytopathol.* 53, 425–444.

Wylie, S. J., Tran, T. T., Nguyen, D. Q., Koh, S.-H., and Chakraborty, A. (2019). A virome from ornamental flowers in an Australian rural town cites. *Arch. Virol.* 164, 2255–2263.

Xin, M., Cao, M., Liu, W., Ren, Y., Zhou, X., and Wang, X. (2017). Two negative-strand RNA viruses identified in watermelon represent a novel clade in the order *Bunyavirales*. *Front. Microbiol.* 8:1514. doi: 10.3389/fmicb.2017.01514

Xu, C., Sun, X., Taylor, A., Jiao, C., Xu, Y., Cai, X., et al. (2017). Diversity, distribution, and evolution of tomato viruses in china uncovered by small RNA sequencing. *J. Virol.* 91:e00173-17.

Yan, T., Zhu, J. R., Di, D., Gao, Q., and Zhang, Y. (2015). Characterization of the complete genome of barley yellow striate mosaic virus reveals a nested gene encoding a small hydrophobic protein. *Virology* 478, 112–122.

Yang, C., Zhang, S., Han, T., Fu, J., Di Serio, F., and Cao, M. (2019). Identification and characterization of a novel emaravirus associated with jujube (*Ziziphus jujuba* Mill.) yellow mottle disease. *Front. Microbiol.* 25:1417. doi: 10.3389/fmicb.2019.01417

Yang, X., Huang, J. L., Liu, C. H., Chen, B., Zhang, T., and Zhou, G. H. (2017). Rice stripe mosaic virus, a novel cytorhabdovirus infecting rice via leafhopper transmission. *Front. Microbiol.* 7:2140. doi: 10.3389/fmicb.2016.02140

Zhang, S., Huang, A., Zhou, X., Li, Z., and Dietzgen, R. G. (2020). Natural defect of a plant rhabdovirus glycoprotein gene: a case study of virus-plant co-evolution. *Phytopathology* doi: 10.1094/PHYTO-05-20-0191-FI [Epub ahead of print].

Zhang, Y.-Z., Chen, Y.-M., Wang, W., Qin, X.-C., and Holmes, E. C. (2019). Expanding the RNA virosphere by unbiased metagenomics. *Annu. Rev. Virol.* 6, 119–139.

Zhang, Y. Z., Shi, M., and Holmes, E. C. (2018). Using metagenomics to characterize an expanding virosphere. *Cell* 172, 1168–1172.

Zheng, K., Chen, T.-H., Wu, K., Kang, Y.-C., and Yeh, S.-D. (2020). Characterization of a new orthotospovirus from chilli pepper in Yunnan Province, China. *Plant Dis.* 104, 1175–1182.

Zheng, Y., Navarro, B., Wang, G., Wang, Y., and Yang, Z. (2017). Actinidia chlorotic ringspot-associated virus: a novel emaravirus infecting kiwifruit plants. *Mol. Plant Pathol.* 18, 569–581.

Zhou, J., Cao, K., Zhang, Z., Wang, L., and Li, S. (2020). Identification and characterization of a novel rhabdovirus infecting peach in China. *Virus Res.* 280:197905. doi: 10.1016/j.virusres.2020.197905

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.

Copyright © 2020 Bejerman, Debat and Dietzgen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 December 2020
doi: 10.3389/fmicb.2020.564310

Artificially Edited Alleles of the Eukaryotic Translation Initiation Factor 4E1 Gene Differentially Reduce Susceptibility to Cucumber Mosaic Virus and Potato Virus Y in Tomato

Hiroki Atarashi[1†], Wikum Harshana Jayasinghe[2,3†], Joon Kwon[2], Hangil Kim[2], Yosuke Taninaka[2†], Manabu Igarashi[4,5], Kotaro Ito[1], Tetsuya Yamada[2], Chikara Masuta[2] and Kenji S. Nakahara[2]**

[1]Research and Development Division, Kikkoman Corporation, Noda, Chiba, Japan, [2]Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan, [3]Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka, [4]Division of Global Epidemiology, Research Center for Zoonosis Control, Hokkaido University, Sapporo, Japan, [5]Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education, Hokkaido University, Sapporo, Japan

OPEN ACCESS

Edited by:
Ahmed Hadidi
Agricultural Research Service (USDA), United States

Reviewed by:
Magdy Mahfouz, King Abdullah University of Science and Technology, Saudi Arabia
Fabrizio Cillo, Italian National Research Council, Italy
Peter Palukaitis, Seoul Women's University, South Korea

***Correspondence:**
Kenji S. Nakahara
knakahar@res.agr.hokudai.ac.jp
Chikara Masuta
masuta@res.agr.hokudai.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: *21 May 2020*
Accepted: *11 November 2020*
Published: *10 December 2020*

Citation:
Atarashi H, Jayasinghe WH, Kwon J, Kim H, Taninaka Y, Igarashi M, Ito K, Yamada T, Masuta C and Nakahara KS (2020) Artificially Edited Alleles of the Eukaryotic Translation Initiation Factor 4E1 Gene Differentially Reduce Susceptibility to Cucumber Mosaic Virus and Potato Virus Y in Tomato. Front. Microbiol. 11:564310. doi: 10.3389/fmicb.2020.564310

Eukaryotic translation initiation factors, including eIF4E, are susceptibility factors for viral infection in host plants. Mutation and double-stranded RNA-mediated silencing of tomato *eIF4E* genes can confer resistance to viruses, particularly members of the *Potyvirus* genus. Here, we artificially mutated the *eIF4E1* gene on chromosome 3 of a commercial cultivar of tomato (*Solanum lycopersicum* L.) by using CRISPR/Cas9. We obtained three alleles, comprising two deletions of three and nine nucleotides (3DEL and 9DEL) and a single nucleotide insertion (1INS), near regions that encode amino acid residues important for binding to the mRNA 5' cap structure and to eIF4G. Plants homozygous for these alleles were termed 3DEL, 9DEL, and 1INS plants, respectively. In accordance with previous studies, inoculation tests with potato virus Y (PVY; type member of the genus *Potyvirus*) yielded a significant reduction in susceptibility to the N strain (PVY^N), but not to the ordinary strain (PVY^O), in 1INS plants. 9DEL among three artificial alleles had a deleterious effect on infection by cucumber mosaic virus (CMV, type member of the genus *Cucumovirus*). When CMV was mechanically inoculated into tomato plants and viral coat accumulation was measured in the non-inoculated upper leaves, the level of viral coat protein was significantly lower in the 9DEL plants than in the parental cultivar. Tissue blotting of microperforated inoculated leaves of the 9DEL plants revealed significantly fewer infection foci compared with those of the parental cultivar, suggesting that 9DEL negatively affects the initial steps of infection with CMV in a mechanically inoculated leaf. In laboratory tests, viral aphid transmission from an infected susceptible plant to 9DEL plants was reduced compared with the parental control. Although many pathogen resistance genes have been discovered in tomato and its wild relatives, no CMV resistance genes have been used in practice. RNA silencing of *eIF4E* expression has previously been reported to not affect susceptibility to CMV in tomato. Our findings suggest that artificial gene editing can introduce additional resistance to that achieved with mutagenesis breeding, and that edited *eIF4E* alleles confer an alternative way to manage CMV in tomato fields.

Keywords: recessive resistance, clustered regularly interspaced short palindromic repeats/Cas9, tomato, potato virus Y, cucumber mosaic virus, eIF4E

INTRODUCTION

Tomato, which originated in Latin America, belongs to the Solanaceae family. This family consists of approximately 100 genera and 2,500 species, including other agronomically important plants, such as potato (*Solanum tuberosum* L.), pepper (*Capsicum*), and tobacco (*Nicotiana tabacum* L.; Olmstead et al., 2008). Tomato has a rich nutrient profile. As one of the major globally grown crops, annual global production exceeds 160 million tons, accounting for more than 10% of global vegetable production.[1]Tomato is susceptible to more than 200 diseases, which are caused by a great variety of pathogens including viruses, bacteria, fungi, and nematodes (Lukyanenko, 1991). Because the use of agricultural chemicals is often ineffective, especially for plant viruses, and expensive for growers, the breeding of resistant crops is important for sustainable crop production.

Cucumber mosaic virus (CMV) has been ranked among the top five scientifically and economically important plant viruses (Scholthof et al., 2011). In the field, this virus is readily transmitted in a stylet-borne, nonpersistent manner by more than 80 species of aphids; transmission may occasionally be seed-borne. CMV can negatively affect plant growth and fruit yield, with associated mosaic or lethal necrotic symptoms. Although the losses are difficult to quantify because the viral incidence changes annually in different locations, tomato yield losses caused by CMV infection have been reported to reach 25–50% of the total yield in China (Tien and Wu, 1991) and 80% of the total yield in Italy and Spain (Gallitelli et al., 1991; Jordá et al., 1992).

Although many genes that convey resistance to pathogens have been discovered in tomato and its wild relatives (Foolad, 2007), the few genes that have been shown to promote resistance to CMV have not been practically used in tomato breeding (Stamova and Chetelat, 2000). Consequently, many trials of ways to improve resistance have been performed. These include investigations where CMV gene was introduced into the host by genetic engineering, and antiviral resistance was achieved by RNA silencing (Anderson et al., 1992; Palukaitis et al., 1992; Yie et al., 1992) and the practical development of an attenuated CMV strain containing satellite RNA (Sayama et al., 1993).

Studies of resistance in plant breeding over the past 2 decades have identified not only dominant but also recessive resistance genes (Gallois et al., 2018). Plant viruses are obligate parasites. Their replications depend on host plant because viral genomes are smaller than those of other plant pathogens except viroids and contain only a dozen or fewer genes (Sanfacon, 2017). A plant that harbors a loss-of-function mutation in an opportune host factor but can complete the viral replication becomes resistant to the virus through inheritance of a recessive trait (Hashimoto et al., 2016; Bastet et al., 2017). Examples of recessive resistance have been demonstrated for mutant alleles of genes of the eukaryotic translation initiation factor (eIF) family. Loss-of-function studies on the *eIF4E* gene isoform in *Arabidopsis thaliana* have shown that resistance is conferred against members of the genus *Potyvirus* (Lellis et al., 2002; Sato et al., 2005). Several similar cases of eIF4E-associated resistance against other viruses in various host species from different families have been reported, including barley yellow mosaic virus in rice (*Oryza sativa*; Albar et al., 2006), CMV and turnip crinkle virus in *A. thaliana* (Yoshii et al., 2004), and melon necrotic spot virus in melon (*Cucumis melo*) (Nieto et al., 2006).

Natural antiviral alleles of *eIF4E1* have been identified as *pot1* and *pot1*[2] on chromosome 3 in tomato relatives (Ruffel et al., 2005; Lebaron et al., 2016). A knockout mutant of *eIF4E2* on chromosome 2 confers resistance to pepper veinal mottle virus (Piron et al., 2010; Moury et al., 2020). Interestingly, *pot1* shows a wider spectrum of resistance to members of the *Potyvirus* genus than does a knockout mutant of eIF4E1; in addition, it shows a similar spectrum to that of eIF4E1 and eIF4E2 double-knockout, without the growth defect seen in the double-knockout plant (Gauffier et al., 2016). RNA silencing of eIF4E expression also leads to a wider spectrum of resistance to members of the genus *Potyvirus*, but not to other viruses, including CMV (Mazier et al., 2011). Mutagenesis breeding has been carried out widely by using physical and chemical mutagens to randomly mutagenize a large population and then select mutants that show the target traits (Sikora et al., 2011; Oladosu et al., 2016). Recently genome editing techniques based on the bacterial immune system (Clustered Regularly Interspaced Short Palindromic Repeats; CRISPR/Cas9 system; Jinek et al., 2012) have become very useful tools for efficiently obtaining loss-of-function mutants (Bortesi and Fischer, 2015; Jaganathan et al., 2018). In the breeding of crops for resistance, CRISPR/Cas9-mediated mutagenesis has been shown to be a practical method for obtaining mutations in eIFs that confer viral resistance in cucumber (*Cucumis sativus L.*) and rice (*Oryza sativa*; Chandrasekaran et al., 2016; Macovei et al., 2018). Here, by applying CRISPR/Cas9 to edit the *eIF4E1* gene on chromosome 3 in a commercial tomato cultivar, we obtained additional alleles of *eIF4E1*. Tomato plants carrying these alleles showed differential resistance not only to potato virus Y (PVY) but also to CMV.

MATERIALS AND METHODS

Binary Plasmid Construction

This study used two vectors, pUC19-AtU6oligo and pZK_gYSA_FFCas9 (Mikami et al., 2015; Kanazashi et al., 2018), to create the binary vector pZK_AtU6gRNA_FFCas9_NPTII, following the method described by Osmani et al. (2019). Guide RNA (gRNA) recognition sites in tomato were found using CRISPRdirect.[2] The sense and antisense oligonucleotide sequences (5'-attgagggtaaatctgataccagc-3' and 5'-aaacgctggtatcagatttaccct-3', respectively), including the nucleotide sequence of a target site in the *eIF4E1* gene on chromosome 3 (underlined in **Figure 1A**), were synthesized and annealed to a pair of oligonucleotides for the gRNA to *eIF4E1*. The annealed oligonucleotides were

[1]http://www.fao.org/home/en/

[2]https://crispr.dbcls.jp/

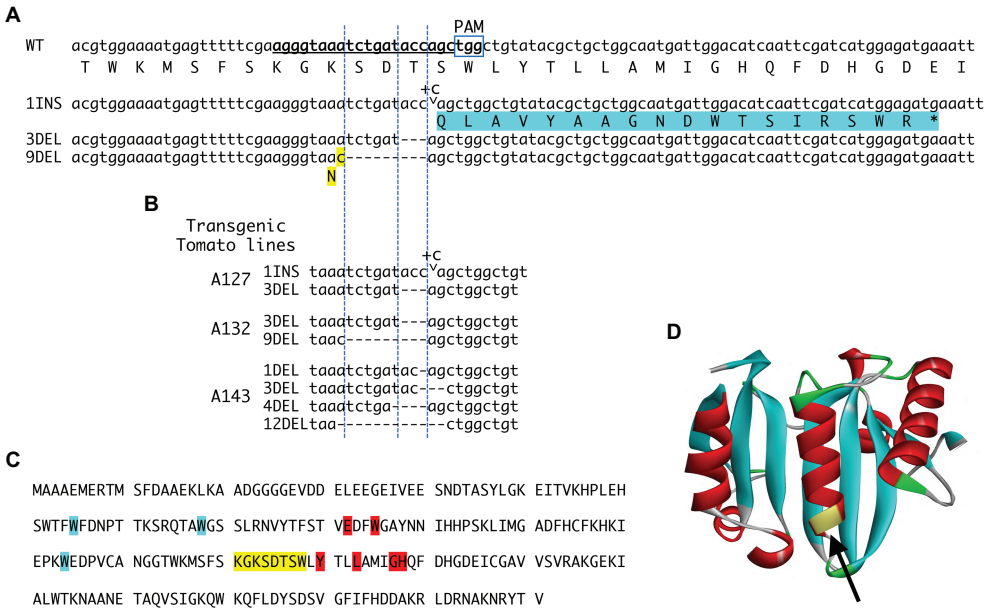


FIGURE 1 | eIF4E1 alleles artificially created *via* CRISPR/Cas9. **(A)** Wild-type (WT) sequence and sequences of the new alleles. Bold underlined letters indicate the guide RNA (gRNA) target sequence. The protospacer adjacent motif (PAM) is marked. Three alleles were obtained: a one-cytosine insertion (1INS), a three-nucleotide deletion (3DEL), and a nine-nucleotide deletion (9DEL). The 1INS (marked by the arrowhead and + C) induces a − 1 frameshift immediately after the 144th codon, assigning threonine in the translation of *eIF4E1*; this results in premature truncation of the frameshift-mediated extending peptide (highlighted in blue). Protein translated from the 3DEL allele has a deletion of threonine encoded by the 144th codon, and the protein translated from the 9DEL allele has a single amino acid substitution (lysine to asparagine at the X codon; highlighted in yellow) and a deletion of three amino acid residues, serine, aspartic acid and threonine. **(B)** Mutations in the eIF4E1 gene in three transgenic tomato plants harboring the CRISPR/Cas9 construct were investigated by amplicon sequencing. In addition to the above three alleles **(A)**, alleles with 1-, 4-, and 12-nucleotide deletions (1DEL, 4DEL, and 12DEL) were detected. **(C)** The amino acid sequence of eIF4E1. The amino acids corresponding to the gRNA nucleotide sequence are highlighted in yellow and are in close proximity to the amino acids that are critical for binding to the mRNA 5′ cap structure (highlighted in blue) and the eIF4G protein (highlighted in red; German-Retana et al., 2008; Miras et al., 2017). **(D)** Three-dimensional structure of the tomato eIF4E1 protein. The region indicated with an arrowhead corresponds to that targeted by the gRNA sequence.

cloned into the BbsI sites of pUC19_AtU6oligo. Then, the cloned plasmid was digested with restriction enzyme I-SceI, and the gRNA expression cassette was re-cloned into pZD_AtU6gR NA _FFCas9_NPTII. *Agrobacterium tumefaciens* LBA4404 (Takara Bio, Shiga, Japan) was then transformed with the binary vector.

Plant Transformation and Transgenic Plant Propagation

Kikkoman Corporation's inbred line S8, a big-fruited tomato, was used for artificial editing by CRISPR/Cas9. Plants were transformed as described by Sun et al. (2006). In brief, diced leaves were inoculated with *A. tumefaciens* LBA4404 transformed with the constructed binary plasmid vector. Callus was generated, and shoots were formed. The shoots were transferred to and grown on a selective regeneration medium containing 1.5 mg/L zeatin, 50 mg/L kanamycin, and 200 mg/L carbenicillin in a growth chamber (16 h light/8 h dark). Then the largest shoots were transferred to a rooting medium containing 50 mg/L kanamycin, 100 mg/L carbenicillin, and 0.1 mg/L 1-naphthylacetic acid. Lines with roots were transferred for culture in soil pots under white fluorescent light (14 h light/10 h dark). T1, T2, and T3 generations of the transgenic lines were produced through self-pollination of each parent generation.

Genotyping of Transgenes and Mutations

Tomato genomic DNA was extracted from transgenic and non-transgenic regenerated plants by using a MonoFas Plant DNA Extraction Kit (GL Sciences, Tokyo, Japan). The presence of the transgene including the gRNA region was checked by polymerase chain reaction (PCR) analysis using a pair of primers (5′-tgggaatctgaaagaagagaagca-3′ and 5′-aaacgctggtatcagatttaccct-3′) and KOD-FX enzyme solution (Toyobo, Osaka, Japan). PCR conditions were incubation at 95°C for 2 min, followed by 30 cycles of 98°C for 10 s, 60°C for 25 s, and 68°C for 15 s, with a final step of 68°C for 3 min. Subsequently, the transgenic lines were genotyped for mutations at the *eIF4E* target sites by using a cleaved amplified polymorphic sequences (CAPS) assay as follows. PCRs were performed under the above conditions, but using a different pair of primers (5′- atccatcacccaagcaagttaatt-3′ and 5′-gtccacaaagctatttttctccc-3′), and the PCR products were digested with the restriction enzyme PvuII. The digested products were separated using 1.8% agarose gel electrophoresis. The indel mutations in T1 and T2 progeny seedlings were confirmed by sequencing the target region. PCR fragments from the CAPS assay were cloned into the pTA2 vector (Toyobo), and nucleotide sequences of several clones were determined.

Three-Dimensional Structure Modeling

The three-dimensional structure of tomato eIF4E1 was constructed by homology modeling based on the crystal structure of *Pisum sativum* eIF4E (Protein Data Bank ID: 2WMC) by using Discovery Studio 2017 software (Biovia).

Mechanical Inoculation and Detection of the Viruses

Cucumber mosaic virus strain yellow (CMV-Y) and PVY strains N and O (PVY^N and PVY^O) (Kwon et al., 2020) were initially propagated on *Nicotiana benthamiana*, and infected upper leaves were used as inoculum for mechanical inoculation. Tomato seedlings at the second or third true leaf stage were mechanically inoculated with crude inoculum, which was viral-infected leaf tissue ground in 0.1 M phosphate buffer (pH 7.0). Following the time-course observations of symptom expression, viral coat protein (CP) accumulation in non-inoculated upper leaves at 25 or 35 days post inoculation (dpi) was investigated with a double antibody sandwich-enzyme linked immunosorbent assay (ELISA) using anti-CP polyclonal antibody (Japan Plant Protection Association). Reverse transcription PCR (RT-PCR) analysis to detect CMV-Y genomic RNA was performed with a pair of primers, 5'-gtacagagttcagggttgagcg-3' and 5'-agcaatactgccaactcagctcc-3', as described in our previous study (Kwon et al., 2020).

Blotting of Microperforated Leaves Inoculated With CMV-Y

Tissue blotting was performed as described previously (Kaneko et al., 2004; Murakami et al., 2016); anti-CMV CP polyclonal antibody was used. Sizes of infection foci and infected area were measured using ImageJ (Wayne Rasband, National Institutes of Health). Focus size was measured for foci whose area ranged from 0.1 to 30 mm^2. Focus number was calculated for foci whose size ranged from 0.001 to 30 mm^2.

Aphid Inoculation With CMV-O and Evaluation of Transmission Efficiency

Myzus persicae was grown on healthy *Brassica rapa*, and the ordinary strain of CMV (CMV-O) was used for the aphid transmission assay, since CMV-Y is known to have defects in transmission *via* aphids. To assess the infection frequency of *M. persicae*, apterous aphids that had been starved for 3 h were allowed a 2–5 min acquisition access period on CMV-O-infected *N. tabacum* and tomato plants at 10–30 dpi. Ten aphids were collected and transferred to each genome-edited and wild-type (WT) S8 tomato plant with the use of a fine dry paintbrush. After 1 day on the tomato plant, the aphids were then killed by pesticide. CMV infection was confirmed by the PCR analysis described below and by the assessment of visual symptoms, yellowing, mosaic, and ringspot of leaves. Viral RNA was extracted by TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, United States), and first-strand cDNA was synthesized using AMV reverse transcriptase (Promega, Madison, WI, United States). PCR was performed using cDNA corresponding to 50 μg of RNA extract, 0.5 μM each primer (5'-gtacagagttcagggttgagcg-3', 5'-agcaatactgccaactcagctcc-3'), and Ex Taq polymerase (Takara Bio). The PCR was carried out under the conditions of 2 min at 95°C, followed by 35 cycles of 20 s at 95°C, 30 s at 55°C, and 30 s at 72°C, with a final step of 5 min at 72°C. The PCR products were fractionated using 1.8% agarose gel electrophoresis.

RESULTS

Development of Artificial Mutated eIF4E Alleles in a Commercial Tomato Line

Tomato has two *eIF4E* genes, *eIF4E1* and *eIF4E2*, on chromosomes 3 and 2, respectively; and an isoform *eIF(iso)4E* on chromosome 9. Here we transformed a big-fruited tomato line S8, with *A. tumefaciens* harboring the binary vector for CRISPR-mediated editing of *eIF4E1*. The region targeted by the gRNA (underlined, **Figure 1A**) corresponds to the yellow-highlighted amino acid sequence in **Figure 1C** and to the part of the three-dimensional structure indicated by the arrow in **Figure 1D**. The targeted region is near the amino acid residues critical for binding to the mRNA cap (7-methylguanosine) and eIF4G (**Figure 1C**, highlighted in blue and red, respectively) (German-Retana et al., 2008; Miras et al., 2017). Thus, mutations in the targeted region may affect binding to these entities. The CAPS assay for the targeted sites identified 20 transformants with artificially edited *eIF4E1* alleles. T1 progenies were derived from self-pollination of three out of 20 T0 regenerated plants, and generated mutations in the targeted *eIF4E* loci in their genomes were investigated (**Figure 1B**). DNA fragments including the targeted sites were amplified by PCR and cloned. As a result of nucleotide sequencing of these clones, six artificial *eIF4E1* alleles were detected in T1 plants (**Figure 1B**), including insertion of one nucleotide (1INS) and deletion of three and nine nucleotides (3DEL and 9DEL, respectively). The allele 1INS causes a − 1 frameshift immediately after the 144th codon assigning threonine in the translation of eIF4E1 and results in premature truncation of the frameshift-mediated extending peptide. The alleles 3DEL and 9DEL encode the proteins that have a deletion of threonine at the 144th codon and serine, aspartic acid, and threonine, following an amino acid substitution from lysine to asparagine (N in **Figure 1A**, highlighted in yellow). All cloned PCR fragments (a total of 16 clones) for which nucleotide sequences were determined corresponded to one of the above-mentioned artificial alleles but not the WT allele, suggesting that *eIF4E1* was bi-allelically edited by CRISPR/Cas9 in the T0 generation. Furthermore, null-segregant progenies, without transgene, were obtained from the T1 or T2 generation. The mutant plants showed almost the same phenotype and traits as WT tomato in enclosed greenhouse cultivation (data not shown).

Distinct Susceptibility to PVY^N and PVY^O in Homozygotes of the Artificial *eIF4E1* Alleles

Since the eIF4E protein is a host susceptibility factor that plays a pivotal role in potyviral infection, its recessive resistance

alleles are considered to lack the function that supports infection. Therefore, plants become resistant to the virus only when the natural resistance alleles (including the *pot1* allele of tomato *eIF4E1*) are homozygous (Hashimoto et al., 2016; Bastet et al., 2017). We inoculated tomato plants that were homozygous for one of the above three artificially edited alleles with PVY^N or PVY^O. These strains cause systemic veinal necrosis and mottle, respectively, in tobacco, and both systemically and asymptomatically infect WT tomato plants. The results of ELISA with anti-PVY CP antibody showed lower accumulation of PVY^N CP in the non-inoculated upper leaves in plants homozygous for the artificial *eIF4E1* alleles than in WT plants. This difference was statistically significant for the 1INS and 3DEL alleles, but not the 9DEL allele, and was particularly marked for the 1INS allele, which is considered to confer loss-of-function through frame-shifting and the introduction of a premature stop codon during translation (**Figure 2**). In contrast, PVY^O CP accumulated comparably in the artificial allele homozygotes and parental WT cultivar, suggesting that functional eIF4E1 protein is required for infection with PVY^N but not PVY^O.

Reduced Susceptibility to CMV in Tomato Carrying the Artificial *eIF4E1* Alleles

The loss-of-function mutant of *Arabidopsis* eIF4E, *cum1*, is less susceptible to CMV than WT *Arabidopsis* (Yoshii et al., 2004). Therefore, we next examined whether CMV infection was controlled in tomato plants with the artificial *eIF4E1* alleles. For each of the three edited alleles, a dozen T2 plants homozygous for the allele were mechanically inoculated with CMV-Y. Most inoculated plants showed symptoms at 30 dpi, with or without the artificial eIF4E1 allele. Investigation of CMV CP accumulation in the non-inoculated upper leaves by ELISA using anti-CMV CP antibody showed that the levels of CP accumulation in plants carrying 9DEL were significantly lower than those in the parental WT plants; no significant difference was found for the other alleles. This suggests that tomato *eIF4E1* is involved in CMV infection and that 9DEL negatively affects CMV infection in tomato (**Figure 2**).

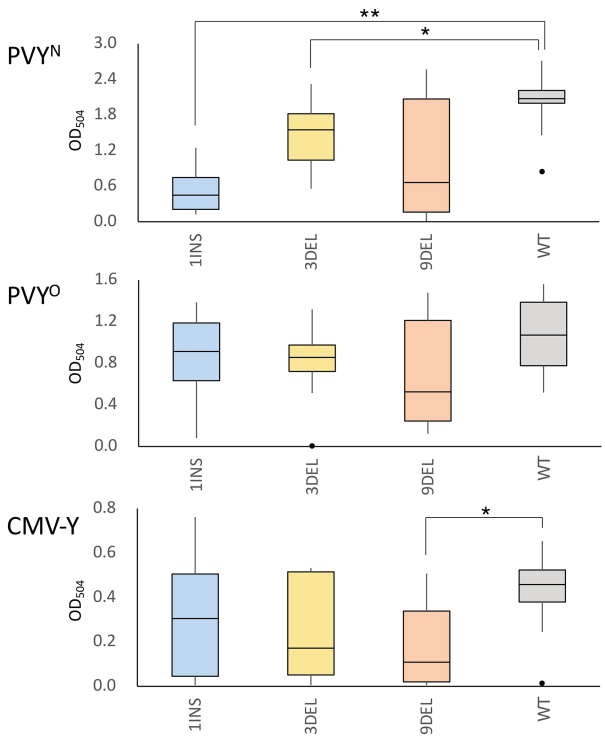


FIGURE 2 | Accumulation of potato virus Y (PVY) and cucumber mosaic virus (CMV) coat protein (CP) in the non-inoculated upper leaves of tomato plants carrying the artificial *eIF4E1* alleles. PVY^N was mechanically inoculated into plants homozygous for the *eIF4E1* alleles (1INS, $n = 8$; 3DEL, $n = 12$; 9DEL, $n = 7$) or the parental line S8 (WT; $n = 14$). Similarly, plants were inoculated with PVY^O or CMV-Y (PVY^O experiment: 1INS, $n = 10$; 3DEL, $n = 10$; 9DEL, $n = 7$; and WT, $n = 8$; CMV-Y experiment: each allele, $n = 10$, WT, $n = 10$). Accumulation of PVY and CMV CP was investigated using enzyme linked immunosorbent assay (ELISA) with anti-PVY or anti-CMV CP polyclonal antibodies at 25 and 35 days post inoculation (dpi), respectively. Data are presented as boxplots. $^*p < 0.05$; $^{**}p < 0.01$ (Wilcoxon's signed-rank test).

Resistance of 9DEL Tomato Plants to Mechanically Inoculated CMV

We then investigated which steps are affected by 9DEL during the infection cycle of CMV (**Figure 3**). By tissue blotting of microperforated inoculated leaves using anti-CMV CP antibody, we found that there was a significantly smaller infection area per inoculated true leaf in 9DEL plants than in WT plants (**Figures 3A**,**B** right graphs); this difference was mainly due to a significantly lower infection focus number per leaf in 9DEL plants because there was no significant difference in focus size (**Figures 3A**,**B** left and middle graphs). All inoculated parental cultivar plants showed symptoms, including severe yellowing in the upper leaves by 11 dpi, whereas the inoculated 9DEL plants showed no or few symptoms in the upper leaves at this time (**Figure 3C**), and one-third (7/20) of the 9DEL plants did not show any symptoms by 35 dpi (**Figure 3D**). By conducting ELISA to detect CMV CP, we demonstrated that the symptomless inoculated 9DEL plants contained an undetectable level of CMV CP (comparable or less than that in the control mock-infected plant) in the upper leaves at 35 dpi (**Figure 3E**). When the symptomless inoculated 9DEL plants were subjected to RT-PCR analysis, CMV genomic RNA was detected in six-sevenths of plants (**Figure 3F**), indicating their latent systemic infection with CMV. These results indicate that 9DEL allows systemic infection with CMV although it inhibits the establishment of CMV infection in the mechanically inoculated leaves.

Resistance to Infection by Aphid Transmission of CMV in Tomato Carrying 9DEL

To determine whether the 9DEL allele is useful for CMV management in practice, we tested the efficiency of aphid transmission of CMV from infected *N. tabacum* to 9DEL tomato plants (**Figure 4**). When we monitored the number of plants

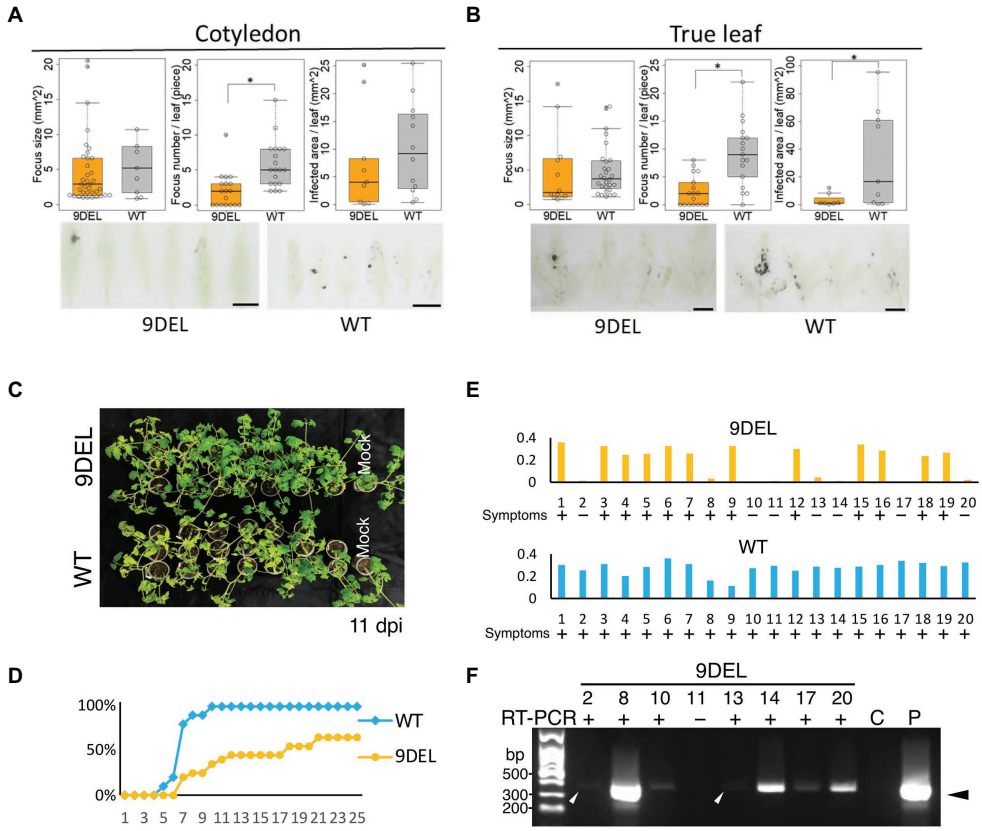


FIGURE 3 | Detailed investigation of resistance to CMV-Y in 9DEL plants. Infection foci in inoculated cotyledons **(A)** and true leaves **(B)** were detected by microperforated leaf blotting with anti-CMV CP antibody. **(A)** The focus size [9DEL, $n = 9$; parental cultivar (WT), $n = 37$], focus number per leaf (9DEL, $n = 18$; WT, $n = 19$), and infected area per leaf (9DEL, $n = 9$; WT, $n = 12$xx) were compared between cotyledons of 9DEL plants and WT. **(B)** The focus size (9DEL, $n = 12$; WT, $n = 28$), focus number per leaf (9DEL, $n = 17$; WT, $n = 17$), and infected area per leaf (9DEL, $n = 7$; WT, $n = 9$) were compared between true leaves of 9DELplants and WT. Data are presented as boxplots and individual data points. $^*p < 0.05$ (Student's t-test). **(C)** Images of inoculated and mock-inoculated plants at 11 dpi are shown. Symptoms are yellowing of upper leaves. **(D)** Time-course of percentage of inoculated plants displaying symptoms ($n = 20$). **(E)** Results of ELISA of CMV CP accumulation in the upper leaves at 34 dpi. The data were corrected by subtracting the level in a corresponding healthy plant from the raw data. Data for 20 individual plants are displayed for each genotype. **(F)** Reverse transcription-polymerase chain reaction (RT-PCR) analysis of CMV genomic RNA in the upper leaves of the eight symptomless 9DEL plants. Control samples were prepared from healthy (C) and infected (P) tomato leaves. Arrowheads indicate the position of the expected band by RT-PCR.

with mosaics and abnormal leaves over time, we obtained similar results in two separate experiments (**Figure 4A**): i.e., the aphid transmission of CMV was reduced to around half of the level observed in WT tomato (**Figure 4A**). In addition, RT-PCR results revealed that the non-symptomatic plants were not actually infected by CMV (**Figure 4B**). These results indicate that the 9DEL mutation resulted in immunity as opposed to tolerance (**Figures 3C-F**). We then compared aphid transmission from an infected 9DEL plant vs. infected WT plant to a WT plant. Infected 9DEL and WT plants that showed symptoms were used as the aphid source. Aphid transmission from the infected 9DEL and WT plants showed similar efficiency to each other (**Figure 4C**).

DISCUSSION

Disease-related yield loss is a critical issue for farmers. Breeding disease-resistant crops with maintained or improved yields and reduced pesticide dependence is important for sustainable agriculture. Although tomatoes have been bred for resistance against viruses, no such work has been undertaken for CMV because no genetic resource is available (see "Introduction"). Here, we used artificial gene editing of eIF4E to introduce additional resistance against PVY and CMV in a commercial tomato cultivar.

In most cases, natural mutants and genetically modified alleles of eIF4E in tomato and other crops confer resistance to members of the genus *Potyvirus*, such as PVY (see Introduction). In general, plants possess multiple genes for eIF4E and its isoforms, and *Potyvirus* members do not necessarily use a single eIF4E for their infection (Bastet et al., 2017); the particular eIF4E gene(s) required for infection with a virus differ between host plant species. We initially tested the effect of our artificially edited *eIF4E1* alleles in tomato on susceptibility to two strains of PVY, PVY^N, and PVY^O. The 1INS and 3DEL alleles significantly reduced the susceptibility to PVY^N but not to PVY^O (**Figure 2**). These results can be interpreted on the

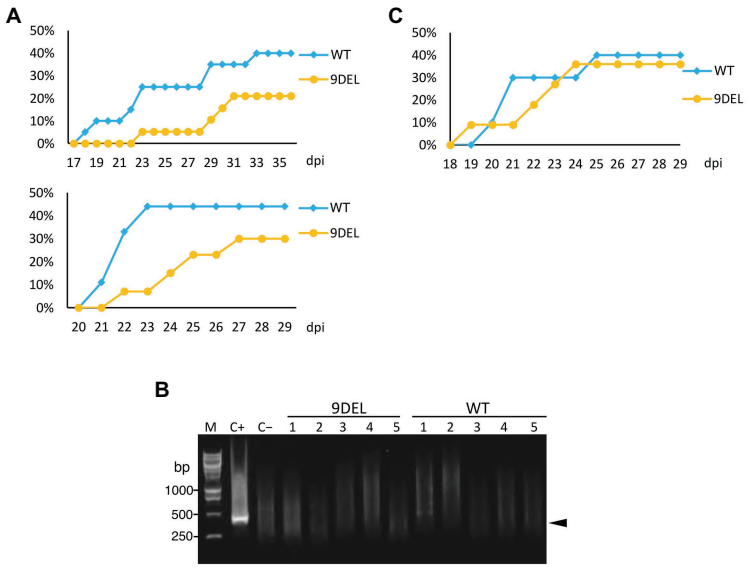


FIGURE 4 | Aphid transmission of CMV in 9DEL plants compared with the parental line **(A)** CMV-O was transmitted from an infected tobacco plant to 9DEL plants (n = 19) or the parental line (WT). The percentage of plants with symptomatic leaves over time is shown for two independent experiments. **(B)** RT-PCR was conducted to detect CMV genomic RNA in five of each of the 9DEL and WT plants that did not show symptoms. C+, control plant showing symptoms; C–, healthy tomato plant. M, DNA size markers. **(C)** CMV-O was transmitted from an infected 9DEL or WT plant to healthy WT plants (n = 10). The percentage of plants with mosaics and abnormal leaves over time is shown.

basis of previous studies in which eIF4E1 and eIF4E2 were individually or doubly knocked out or down in tomato (Mazier et al., 2011; Gauffier et al., 2016; Bastet et al., 2017; Moury et al., 2020). The previous studies indicated that PVY primarily uses eIF4E1, but most PVY strains can alternatively use eIF4E2 to promote infection when eIF4E1 is knocked out or down. Our results suggest that PVY^O is one of the major PVY strains that can alternatively use eIF4E2, whereas PVY^N is one that exclusively uses eIF4E1 to promote infection. Therefore, 1INS, a frameshift mutation, appears to make eIF4E1 non-functional for PVY infection. Viruses of the family *Potyviridae* have a viral protein genome-linked (VPg) covalently attached at the 5' end of their RNA genomes instead of the 7-methylguanosine cap structure that is found at the 5' end of some eukaryotic and viral mRNAs. VPg has affinity to host eIF4E and its isoform not only for translation of viral proteins but also for viral RNA stabilization and evading antiviral RNA silencing (Saha and Makinen, 2020); the degree of affinity is correlated with viral infectivity (Leonard et al., 2000). Therefore, significant reduction of susceptibility to PVY^N in tomato carrying 3DEL implies the possibility that the deletion mutation of 3DEL affects the affinity to PVY^N VPg. Docking models between eIF4E1s translated from the artificially edited alleles with the Cap structure (**Supplementary Figure S1A**) and VPg (**Supplementary Figure S1B**) suggest that deletions and substitution of both 3DEL and 9DEL disrupt eIF4E1 binding to both Cap and VPg. Our finding that the *eIF4E1* region deleted in 9DEL encodes amino acid residues important for binding to cap structures implies that the reduced susceptibility to CMV in tomato carrying 9DEL could be attributed to a defect eIF4E1 binding to the 5' caps of CMV RNAs. These results also suggest that we can design non-transgenic virus-resistant plants by using the genome-editing technology, on the basis of the *in silico* estimation of the interactions between host factors and viral proteins.

Previously, RNA silencing of both *eIF4E1* and *eIF4E2* conferred broad resistance against members of the genus *Potyvirus* but not against other viruses, such as tomato spotted wilt virus, alfalfa mosaic virus, tobacco mosaic virus, and CMV (Mazier et al., 2011). However, in the current study, editing of a single *eIF4E* gene, *eIF4E1*, to produce the 9DEL allele, negatively affected CMV multiplication and infection efficiency by aphid transmission. This apparent discrepancy may be attributed to a dominant effect of the mutant eIF4E1 protein expressed by this allele. The natural mutant allele of the *eIF4E1* gene, *pot1*, was identified in a wild relative of tomato (Ruffel et al., 2005). The *eIF4E1* knockout allele, which expresses a truncated protein lacking cap-binding activity, was obtained by ethyl methanesulfonate mutagenesis (Piron et al., 2010). Plants homozygous for both *pot1* and the knockout allele show resistance to members of the genus *Potyvirus*. Recent studies compared the resistance spectra among plants carrying *pot1*, the *eIF4E1* knockout allele, or both *eIF4E1* and *eIF4E2* double-knockout alleles (Gauffier et al., 2016; Bastet et al., 2017). The *pot1* homozygous and double-knockout plants of *eIF4E1* and *eIF4E2* showed resistance to a higher number of members of the genus *Potyvirus* compared with *eIF4E1* knockout homozygous plants, even though the *pot1* product has four amino acid substitutions and seems to retain the essential functions of eIF4E. Their analyses suggest that the *pot1* product is not only functionless for *Potyvirus* infection but also dominant-negatively suppresses eIF4E2 expression and thus confers a similar resistance spectrum as the double knockout. The 9DEL allele in this study may have

a similar dominant-negative effect on the genes required for CMV multiplication and transmission. The 9DEL allele reduced CMV susceptibility more than the 1INS allele, which we consider to be a knockout mutant because the deduced truncated translation product lacks the critical domain to bind to eIF4G (**Figure 1B**).

Almost all mechanically inoculated tomato plants carrying the artificially generated *eIF4E1* alleles were infected with CMV, but 9DEL plants accumulated a lower amount of CMV CP in their upper leaves than that in WT plants (**Figure 2**). In our detailed investigation of the differences in resistance to CMV between 9DEL and WT plants (**Figure 3**), one-third of inoculated 9DEL plants did not develop any symptoms when the accumulation of CMV CP in the upper leaves was low (**Figure 3E**). Although most of the symptomless plants still allowed systemic infection with CMV (**Figure 3F**), there were significantly fewer foci in the inoculated leaves of the 9DEL plants that the WT plants (**Figures 3A,B**), suggesting that 9DEL negatively affects the initial establishment of CMV infection. Incomplete resistance to CMV was previously reported in *A. thaliana*. When the eIF4E knockout mutant *cum1* of *A. thaliana* was inoculated with CMV, the mutant inhibited the expression of CMV movement protein but still accumulated a similar amount of the other viral proteins and RNAs to those of WT (Yoshii et al., 2004). Here, the expression of the movement protein and CMV movement may have been affected in the tomato plants carrying the edited alleles, although recent studies indicate that eIF4E is involved in multiple steps of various types of viral infection or propagation (Montero et al., 2015). CMV was still detected by RT-PCR in most sample RNAs from mechanically inoculated plants without symptoms (**Figure 3F**), but CMV RNAs were not detected in the 9DEL plants that did not show any symptoms in the aphid transmission trial (**Figure 4B**). On the basis of these results, we assume that the efficacy of the resistance provided by the 9DEL allele depends on the amount of invading CMV virions in the initially inoculated cells, and that when the invading CMV amount is small, the 9DEL allele completely inhibits systemic infection of CMV.

This study is the first report to demonstrate the possibility of introducing resistance to CMV by genome editing. Unlike conventional random mutagenesis, recently developed genome editing techniques are capable of efficiently mutagenizing a specific target gene. Direct genome editing of commercial crop cultivars reduces the time required for altering their traits. Although recessive resistance or natural mutants of host-susceptibility factors such as eIF4E and eIF4G are less utilized for breeding than dominant resistance mutants, some susceptibility factors including eIF4E are thought to be required for most virus infections, implying the possibility that one recessive resistance gene confers resistance to multiple different viruses at once. Considering the emergence of resistance-breaking races or strains of pathogens (Tsuda et al., 1998; Enya et al., 2009), we need to continuously explore and develop resistance genes by taking advantage of recent genetic engineering techniques, including genome editing approaches. The stacking of incomplete-resistance genes might result in sustainable resistance. Even the 9DEL plants were not able to completely inhibit CMV transmission by aphids in the laboratory trial in this study (**Figure 4**). Considering that exposure to CMV in natural fields is generally lower than in a laboratory trial (Hord et al., 2001; Kayode et al., 2014; Jalender et al., 2017), the 9DEL allele is likely to be of practical use for breeding cultivars with resistance to CMV.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

HA, CM, and KN conceived and designed the experiments. HA, KN, and TY created CRISPR-edited tomato plants. HA, WJ, YT, JK, HK, and KN collected the samples and conducted analyses of the inoculated plants. HK, MI, and CM performed 3D and docking-modeling. HA, KI, CM, and KN discussed the results and drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Tetsuo Maoka and Masaki Endo (National Agriculture and Food Research Organization, NARO) for providing PVY-N and plasmid vectors, respectively, and for the CRISPR/Cas9 experiments. We thank Teruo Sano, Misao Umehara, Zhila Osmani, Kento Mori, and Kami Murakami for technical assistance with the experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.564310/full#supplementary-material

REFERENCES

Albar, L., Bangratz-Reyser, M., Hébrard, E., Ndjiondjop, M. N., Jones, M., and Ghesquière, A. (2006). Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to *Rice yellow mottle virus*. *Plant J.* 47, 417–426. doi: 10.1111/j.1365-313X.2006.02792.x

Anderson, J. M., Palukaitis, P., and Zaitlin, M. (1992). A defective replicase gene induces resistance to cucumber mosaic virus in transgenic tobacco plants. *Proc. Natl. Acad. Sci. U. S. A.* 89, 8759–8763. doi: 10.1073/pnas.89.18.8759

Bastet, A., Robaglia, C., and Gallois, J. L. (2017). eIF4E resistance: natural variation should guide gene editing. *Trends Plant Sci.* 22, 411–419. doi: 10.1016/j.tplants.2017.01.008

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi: 10.1016/j.biotechadv.2014.12.006

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi: 10.1111/mpp.12375

Enya, J., Ikeda, K., Takeuchi, T., Horikoshi, N., Higashi, T., Sakai, T., et al. (2009). The first occurrence of leaf mold of tomato caused by races 4.9 and 4.9.11 of *Passalora fulva* (syn. *Fulvia fulva*) in Japan. *J. Gen. Plant Pathol.* 75, 76–79. doi: 10.1007/s10327-008-0134-0

Foolad, M. R. (2007). Genome mapping and molecular breeding of tomato. *Int. J. Plant Genom.* 2007:64358. doi: 10.1155/2007/64358

Gallitelli, D., Vovlas, C., Martelli, G., Montasser, M. S., Tousignant, M. E., and Kaper, J. M. (1991). Satellite-mediated protection of tomato against cucumber mosaic virus: II. Field test under natural epidemic conditions in southern Italy. *Plant Dis.* 75, 93–95. doi: 10.1094/PD-75-0093

Gallois, J. L., Moury, B., and German-Retana, S. (2018). Role of the genetic background in resistance to plant viruses. *Int. J. Mol. Sci.* 19:2856. doi: 10.3390/ijms19102856

Gauffier, C., Lebaron, C., Moretti, A., Constant, C., Moquet, F., Bonnet, G., et al. (2016). A TILLING approach to generate broad-spectrum resistance to potyviruses in tomato is hampered by eIF4E gene redundancy. *Plant J.* 85, 717–729. doi: 10.1111/tpj.13136

German-Retana, S., Walter, J., Doublet, B., Roudet-Tavert, G., Nicaise, V., Lecampion, C., et al. (2008). Mutational analysis of plant cap-binding protein eIF4E reveals key amino acids involved in biochemical functions and potyvirus infection. *J. Virol.* 82, 7601–7612. doi: 10.1128/JVI.00209-08

Hashimoto, M., Neriya, Y., Yamaji, Y., and Namba, S. (2016). Recessive resistance to plant viruses: potential resistance genes beyond translation initiation factors. *Front. Microbiol.* 7:1695. doi: 10.3389/fmicb.2016.01695

Hord, M. J., Garcia, A., Villalobos, H., Rivera, C., Macaya, G., and Roossinck, M. J. (2001). Field survey of *Cucumber mosaic virus* subgroups I and II in crop plants in Costa Rica. *Plant Dis.* 85, 952–954. doi: 10.1094/PDIS.2001.85.9.952

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for crop improvement: an update review. *Front. Plant Sci.* 9:985. doi: 10.3389/fpls.2018.00985

Jalender, P., Bhat, B. N., Anitha, K., and Vijayalakshmi, K. (2017). Survey for the incidence of cucumber mosaic virus in tomato growing areas of telangana and Andhra Pradesh. *Int. J. Pure App. Biosci.* 5, 2058–2063. doi: 10.18782/2320-7051.5763

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Jordá, C., Alfaro, A., Aranda, M. A., Moriones, E., and García-Arenal, F. (1992). Epidemic of cucumber mosaic virus plus satellite RNA in tomatoes in eastern Spain. *Plant Dis.* 76, 363–366. doi: 10.1094/PD-76-0363

Kanazashi, Y., Hirose, A., Takahashi, I., Mikami, M., Endo, M., Hirose, S., et al. (2018). Simultaneous site-directed mutagenesis of duplicated loci in soybean by using a single guide RNA. *Plant Cell Rep.* 37, 553–563. doi: 10.1007/s00299-018-2251-3

Kaneko, Y. H., Inukai, T., Suehiro, N., Natsuaki, T., and Masuta, C. (2004). Fine genetic mapping of the TuNI locus causing systemic veinal necrosis by turnip mosaic virus infection in *Arabidopsis thaliana*. *Theor. Appl. Genet.* 110, 33–40. doi: 10.1007/s00122-004-1824-4

Kayode, A. B., Odu, B. O., Ako-Nai, K. A., and Alabi, O. J. (2014). Occurrence of *Cucumber mosaic virus* subgroups IA and IB isolates in tomatoes in Nigeria. *Plant Dis.* 98:1750. doi: 10.1094/PDIS-08-14-0844-PDN

Kwon, J., Kasai, A., Maoka, T., Masuta, C., Sano, T., and Nakahara, K. S. (2020). RNA silencing-related genes contribute to tolerance of infection with potato virus X and Y in a susceptible tomato plant. *Virol. J.* 17:149. doi: 10.1186/s12985-020-01414-x

Lebaron, C., Rosado, A., Sauvage, C., Gauffier, C., German-Retana, S., Moury, B., et al. (2016). A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to potato virus Y in tomato albeit with a low durability. *J. Gen. Virol.* 97, 3063–3072. doi: 10.1099/jgv.0.000609

Lellis, A. D., Kasschau, K. D., Whitham, S. A., and Carrington, J. C. (2002). Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF(iso)4E during potyvirus infection. *Curr. Biol.* 12, 1046–1051. doi: 10.1016/S0960-9822(02)00898-9

Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M. G., and Laliberte, J. F. (2000). Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *J. Virol.* 74, 7730–7737. doi: 10.1128/jvi.74.17.7730-7737.2000

Lukyanenko, A. N. (1991). "Disease resistance in tomato" in *Genetic improvement of tomato. Monographs on theoretical and applied genetics. Vol. 14*. ed. G. Kalloo (Berlin, Heidelberg: Springer).

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Cermak, T., et al. (2018). Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol. J.* 16, 1918–1927. doi: 10.1111/pbi.12927

Mazier, M., Flamain, F., Nicolai, M., Sarnette, V., and Caranta, C. (2011). Knock-down of both eIF4E1 and eIF4E2 genes confers broad-spectrum resistance against potyviruses in tomato. *PLoS One* 6:e29595. doi: 10.1371/journal.pone.0029595

Mikami, M., Toki, S., and Endo, M. (2015). Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. *Plant Mol. Biol.* 88, 561–572. doi: 10.1007/s11103-015-0342-x

Miras, M., Truniger, V., Silva, C., Verdaguer, N., Aranda, M. A., and Querol-Audi, J. (2017). Structure of eIF4E in complex with an eIF4G peptide supports a universal bipartite binding mode for protein translation. *Plant Physiol.* 174, 1476–1491. doi: 10.1104/pp.17.00193

Montero, H., Garcia-Roman, R., and Mora, S. I. (2015). eIF4E as a control target for viruses. *Viruses* 7, 739–750. doi: 10.3390/v7020739

Moury, B., Lebaron, C., Szadkowski, M., Ben Khalifa, M., Girardot, G., Bolou Bi, B. A., et al. (2020). Knock-out mutation of eukaryotic initiation factor 4E2 (eIF4E2) confers resistance to pepper veinal mottle virus in tomato. *Virology* 539, 11–17. doi: 10.1016/j.virol.2019.09.015

Murakami, T., Tayama, R., and Nakahara, K. S. (2016). Microperforated leaf blotting on polyvinylidene difluoride and nylon membranes to analyze spatial distribution of endogenous and viral gene expression in plant leaves. *J. Gen. Plant Pathol.* 82, 254–260. doi: 10.1007/s10327-016-0671-x

Nieto, C., Morales, M., Orjeda, G., Clepet, C., Monfort, A., Sturbois, B., et al. (2006). An eIF4E allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. *Plant J.* 48, 452–462. doi: 10.1111/j.1365-313X.2006.02885.x

Oladosu, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A., et al. (2016). Principle and application of plant mutagenesis in crop improvement: a review. *Biotechnol. Biotechnol. Equip.* 30, 1–16. doi: 10.1080/13102818.2015.1087333

Olmstead, R. G., Bohs, L., Migid, H. A., Santiago-Valentin, E., Garcia, V. F., and Collier, S. M. (2008). A molecular phylogeny of the Solanaceae. *Taxon* 57, 1159–1181. doi: 10.1002/tax.574010

Osmani, Z., Jin, S., Mikami, M., Endo, M., Atarashi, H., Fujino, K., et al. (2019). CRISPR/Cas9-mediated editing of genes encoding rgs-CaM-like proteins in transgenic potato plants. *Methods Mol. Biol.* 2028, 153–165. doi: 10.1007/978-1-4939-9635-3_9

Palukaitis, P., Roossinck, M. J., Dietzgen, R. G., and Francki, R. I. (1992). Cucumber mosaic virus. *Adv. Virus Res.* 41, 281–348. doi: 10.1016/s0065-3527(08)60039-1

Piron, F., Nicolai, M., Minoia, S., Piednoir, E., Moretti, A., Salgues, A., et al. (2010). An induced mutation in tomato eIF4E leads to immunity to two potyviruses. *PLoS One* 5:e11313. doi: 10.1371/journal.pone.0011313

Ruffel, S., Gallois, J. L., Lesage, M. L., and Caranta, C. (2005). The recessive potyvirus resistance gene *pot-1* is the tomato orthologue of the pepper *pvr2-eIF4E* gene. *Mol. Gen. Genom.* 274, 346–353. doi: 10.1007/s00438-005-0003-x

Saha, S., and Makinen, K. (2020). Insights into the functions of eIF4E-biding motif of VPg in potato virus a infection. *Viruses* 12:197. doi: 10.3390/v12020197

Sanfacon, H. (2017). Grand challenge in plant virology: understanding the impact of plant viruses in model plants, in agricultural crops, and in complex ecosystems. *Front. Microbiol.* 8:860. doi: 10.3389/fmicb.2017.00860

Sato, M., Nakahara, K., Yoshii, M., Ishikawa, M., and Uyeda, I. (2005). Selective involvement of members of the eukaryotic initiation factor 4E family in the infection of *Arabidopsis thaliana* by potyviruses. *FEBS Lett.* 579, 1167–1171. doi: 10.1016/j.febslet.2004.12.086

Sayama, H., Sato, T., Kominato, M., Natsuaki, T., and Kaper, J. M. (1993). Field testing of a satellite-containing attenuated strain of cucumber mosaic virus for tomato protection in Japan. *Phytopathology* 83, 405–410. doi: 10.1094/Phyto-83-405

Scholthof, K. B., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., et al. (2011). Top 10 plant viruses in molecular plant pathology. *Mol. Plant Pathol.* 12, 938–954. doi: 10.1111/j.1364-3703.2011.00752.x

Sikora, P., Chawade, A., Larsson, M., Olsson, J., and Olsson, O. (2011). Mutagenesis as a tool in plant genetics, functional genomics, and breeding. *Int. J. Plant Genom.* 2011:314829. doi: 10.1155/2011/314829

Stamova, B. S., and Chetelat, R. T. (2000). Inheritance and genetic mapping of cucumber mosaic virus resistance introgressed from *Lycopersicon chilense* into tomato. *Theor. Appl. Genet.* 101, 527–537. doi: 10.1007/s001220051512

Sun, H. J., Uchii, S., Watanabe, S., and Ezura, H. (2006). A highly efficient transformation protocol for micro-tom, a model cultivar for tomato functional genomics. *Plant Cell Physiol.* 47, 426–431. doi: 10.1093/pcp/pci251

Tien, P., and Wu, G. S. (1991). Satellite RNA for the biocontrol of plant disease. *Adv. Virus Res.* 39, 321–339. doi: 10.1016/s0065-3527(08)60799-x

Tsuda, S., Kirita, M., and Watanabe, Y. (1998). Characterization of a pepper mild mottle tobamovirus strain capable of overcoming the L3 gene-mediated resistance, distinct from the resistance-breaking Italian isolate. *Mol. Plant-Microbe Interact.* 11, 327–331. doi: 10.1094/MPMI.1998.11.4.327

Yie, Y., Zhao, F., Zhao, S. Z., Liu, Y. Z., Liu, Y. L., and Tien, P. (1992). High resistance to cucumber mosaic virus conferred by satellite RNA and coat protein in transgenic commercial tobacco cultivar G-140. *Mol. Plant-Microbe Interact.* 5, 460–465. doi: 10.1094/mpmi-5-460

Yoshii, M., Nishikiori, M., Tomita, K., Yoshioka, N., Kozuka, R., Naito, S., et al. (2004). The Arabidopsis cucumovirus multiplication 1 and 2 loci encode translation initiation factors 4E and 4G. *J. Virol.* 78, 6102–6111. doi: 10.1128/JVI.78.12.6102-6111.2004

Conflict of Interest: HA and KI were employed by the company Kikkoman Corporation.

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.

Copyright © 2020 Atarashi, Jayasinghe, Kwon, Kim, Taninaka, Igarashi, Ito, Yamada, Masuta and Nakahara. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 December 2020
doi: 10.3389/fmicb.2020.602039

Next-Generation Sequencing Combined With Conventional Sanger Sequencing Reveals High Molecular Diversity in *Actinidia* Virus 1 Populations From Kiwifruit Grown in China

*Shaohua Wen[1,2], Guoping Wang[1], Zuokun Yang[1], Yanxiang Wang[1], Min Rao[1], Qian Lu[1] and Ni Hong[1,2]**

[1] Key Lab of Plant Pathology of Hubei Province, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [2] Key Laboratory of Horticultural Crop (Fruit Trees) Biology and Germplasm Creation of the Ministry of Agriculture, Wuhan, China

OPEN ACCESS

Edited by:
John Wesley Randles,
University of Adelaide, Australia

Reviewed by:
Michael Norman Pearson,
The University of Auckland,
New Zealand
Xuefeng Wang,
Chinese Academy of Agricultural Sciences, China

***Correspondence:**
Ni Hong
whni@mail.hzau.edu.cn

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

Received: 02 September 2020
Accepted: 11 November 2020
Published: 16 December 2020

Citation:
Wen S, Wang G, Yang Z, Wang Y, Rao M, Lu Q and Hong N (2020) Next-Generation Sequencing Combined With Conventional Sanger Sequencing Reveals High Molecular Diversity in Actinidia Virus 1 Populations From Kiwifruit Grown in China. Front. Microbiol. 11:602039. doi: 10.3389/fmicb.2020.602039

Kiwifruit (*Actinidia* spp.) is native to China. Viral disease–like symptoms are common on kiwifruit plants. In this study, six libraries prepared from total RNA of leaf samples from 69 kiwifruit plants were subjected to next-generation sequencing (NGS). *Actinidia* virus 1 (AcV-1), a tentative species in the family Closteroviridae, was discovered in the six libraries. Two full-length and two near-full genome sequences of AcV-1 variants were determined by Sanger sequencing. The genome structure of these Chinese AcV-1 variants was identical to that of isolate K75 and consisted of 12 open reading frames (ORFs). Analyses of these sequences together with the NGS-derived contig sequences revealed high molecular diversity in AcV-1 populations, with the highest sequence variation occurring at ORF1a, ORF2, and ORF3, and the available variants clustered into three phylogenetic clades. For the first time, our study revealed different domain compositions in the viral ORF1a and molecular recombination events among AcV-1 variants. Specific reverse transcriptase–polymerase chain reaction assays disclosed the presence of AcV-1 in plants of four kiwifruit species and unknown *Actinidia* spp. in seven provinces and one city.

Keywords: next-generation sequencing, *Actinidia* spp., *Actinidia* virus 1, molecular diversity, recombination, RT-PCR

INTRODUCTION

The family Closteroviridae contains many plant viruses causing different diseases of economically important crops (Martelli et al., 2012). Viruses in the family Closteroviridae are characterized by having 650–2,200 nm, flexuous filamentous virions consisting of a positive-sense, single-stranded (+ss) RNA with size ranging from 13 to 19.3 kb (Agranovsky et al., 1995; Rubio et al., 2013). According to the International Committee on Taxonomy of Viruses (ICTV) report posted on August 2018 (ICTV Master Species List, 2018 v1), the family Closteroviridae comprises more than

50 recognized species, which are classified into four genera, *Closterovirus*, *Ampelovirus*, *Crinivirus*, and *Velarivirus*, and seven unassigned members, based on their vectors and phylogenetic relationships (Karasev, 2000; Dolja et al., 2006). Members of the family Closteroviridae share a common genome organization, containing a replication-related module encoded by open reading frame (ORF) 1a and 1b, and a five-gene module encoding a small hydrophobic protein, a homolog of the plant heat shock protein HSP70 (HSP70h), a ∼60 kDa protein, a major coat protein (CP), and a minor coat protein (CPm) (Martelli et al., 2012). A notable exception is that viruses in the genus *Ampelovirus* have smaller genomes, which do not possess CPm and have unique ORFs. Previous studies showed that viruses in the family Closteroviridae have great genetic variability, and molecular evolution might be necessary for adaptation to new environments (Holland et al., 1982; Kreuze et al., 2002; Cuellar et al., 2008; Bertazzon et al., 2010; García-Arenal and Fraile, 2011; Esteves et al., 2012). Homologous recombination has been described for CP and P20 genes of citrus tristeza virus (CTV) (Rubio et al., 2001, 2013; Bar-Joseph and Mawassi, 2013), a member in the genus *Closterovirus*, for CP genes of ampeloviruses grapevine leafroll-associated virus 3 (GLRaV-3) (Turturo et al., 2005; Farooq et al., 2012), GLRaV-1 (Fan et al., 2015), and GLRaV-11 (Boulila, 2010), for HSP70h gene of plum bark necrosis stem pitting-associated virus (PBNSPaV) (Qu et al., 2014), and for ORF1a of GLRaV-4 (Adiputra et al., 2019).

Kiwifruit (*Actinidia* spp.) is an economically important fruit crop cultivated worldwide. Since the first report of apple stem grooving virus (ASGV) (Clover et al., 2003) naturally infecting kiwifruit, currently at least 19 viruses infecting kiwifruit have been reported (Pearson et al., 2011; Chavan et al., 2012, 2013; Blouin et al., 2013, 2018; Wang et al., 2016, 2020; Zheng et al., 2017; Veerakone et al., 2018; Wang D. et al., 2018; Wang Y. et al., 2018; Zhao et al., 2019, 2020). Among these viruses, *Actinidia* virus 1 (AcV-1), a tentative member of the family Closteroviridae, was first characterized from *Actinidia chinensis* in New Zealand (Blouin et al., 2018). The reported genome of AcV-1 isolate K75 (GenBank accession no. KX857665) consists of an 18,848 nt + ss RNA containing at least 12 ORFs. ORF1a encodes a multifunctional protein with two papain-like leader protease (L-Pro) domains (L1 and L2), one methyltransferase (Mtr) domain, and one helicase (Hel) domain. ORF 1b, expressed by a + 1 ribosomal frameshift, encodes an RNA-dependent RNA polymerase (RdRp). ORF2 and ORF3 encode two hypothetical proteins with unknown functions and predicted molecular masses of 13.6 and 25.4 kDa, respectively. ORFs 4–8 consist of a five-gene module, in which, instead of CPm in the module of most viruses in the family Closteroviridae, ORF7 encodes a 30 kDa protein (p30) with a conserved domain of thaumatin-like proteins. ORFs 9–11 encode three small proteins, whose functions are currently unknown. The virus was reported in China in 2018, the second country with AcV-1 infecting kiwifruit (Peng et al., 2018; Wen et al., 2018).

Until now, only one complete genome sequence of AcV-1 isolate K75 has been documented (Blouin et al., 2018), and a few sequences of the viral CP and HSP70h genes are available in GenBank. These sequences are valuable for designing molecular diagnostic methods. The relationship between the infection with the virus and occurrence of kiwifruit disease is still unknown, and the viral populations remain poorly understood. In this study, for the first time, the genome-wide genetic diversity of AcV-1 variants from kiwifruit plants grown in China was determined by using next-generation sequencing (NGS) combined with conventional Sanger sequencing for reverse transcriptase–polymerase chain reaction (RT-PCR) products. This study provides useful information for understanding the molecular evolution within AcV-1 populations and developing reliable molecular detection methods of the virus, which is necessary for epidemiological investigation of the viral disease.

MATERIALS AND METHODS

Plant Materials

During the growing season of kiwifruit plants in 2016–2019, leaf samples of 245 kiwifruit plants, including 104 plants of *A. chinensis*, 39 *Actinidia delicious* plants, nine *Actinidia arguta* plants, 36 *Actinidia eriantha* plants, one *Actinidia rufa* plant, and 56 plants of unknown species, were collected in Hubei, Yunnan, Jiangxi, Shandong, Zhejiang, Henan, Fujian, and Shanxi provinces and Chongqing and Shanghai cities, covering the major kiwifruit production areas in China. Of these plants, 231 plants showed virus disease–like symptoms, including uneven chlorosis between veins, mosaic, ringspot, chlorotic spot, and/or deformation, whereas 14 plants were asymptomatic. Leaves from each plant were pooled as one sample named with geographic origin followed by a plant number. The presence of six known viruses AcV-1, *Actinidia* virus A (AcVA), *Actinidia* virus B (AcVB), ASGV, *Actinidia* chlorotic ringspot-associated emaravirus (AcCRaV), and citrus leaf blotch virus (CLBV), in these samples were tested by RT-PCR using specific primers reported previously (James, 1999; Harper et al., 2008; Blouin et al., 2012; Zheng et al., 2017; **Supplementary Table S1**). A leaf sample of a virus-free seedling of *A. chinensis* was used as a negative control.

According to the origins and symptoms of individually collected leaf samples as mentioned above, six samples were prepared for RNA-seq analyses (**Supplementary Table S2**). One leaf sample (ID: Z2) from an *A. delicious* plant showing severe chlorotic spots, leaf mottle, and deformation, grown at a field in Henan province, and one sample (ID: JX5) from an *A. chinensis* plant showing leaf yellowing from Jiangxi province were individually used for RNA-seq analyses. Additionally, in order to have a full view of kiwifruit virome, four pooled samples JS, ZZ, CQ, and JX were also subjected to RNA-seq analyses. The sample JS was a mixture of four leaf samples (JS27, JS29, JS30, and JS45) from Hubei province. The sample ZZ consisted of leaf samples from 30 kiwifruit plants (ZZ1–ZZ11, ZZ13–ZZ31) grown at a field in Henan province. The sample CQ consisted of leaf samples from 28 kiwifruit plants (CQ1–CQ28) grown at a field in Chongqing city. The sample JX consisted of leaf samples from five kiwifruit plants (JX1–JX4 and JX6) grown at a field in Jiangxi province. Most of these RNA-seq–analyzed samples

(except for two asymptomatic samples JX6 and CQ18) showed viral disease–like symptoms as listed in **Supplementary Table S4**.

Additionally, one leaf sample (ID: WH4) from an *A. arguta* plant showing severe leaf mottle and deformation, grown at a field in Hubei province, was used for Sanger sequencing of AcV-1 genome.

RNA-Seq

Total RNAs were extracted from the six samples prepared for RNA-seq analyses as mentioned above, and ribosomal RNAs (rRNAs) were removed using an Epicenter Ribo-Zero rRNA removal kit (Epicenter, Madison, WI, United States). The rRNA-depleted RNA extracts were used for cDNA library construction with a TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA, United States) and sequenced on an Illumina HiSeq XTen sequencing machine with a paired-end 150 bp setup (Biomarker Biology Technology Ltd., Company, Beijing, China).

Raw reads were trimmed of adaptor sequences and filtered for low-quality reads with more than 5% ambiguous bases (Ns) or with more than 10% of the bases below Q20 quality as described previously (Liu et al., 2020). Qualified reads were *de novo* assembled into contigs using the CLC Genomics Workbench 11.0 (Qiagen, Valencia, CA, United States) platform. Contigs were subsequently screened for sequence identity against the NCBI databases[1] using BlastX and BlastN programs.

Determination of Complete Genomic Sequences of AcV-1 Isolates

The complete genomic sequence of one AcV-1 variant (JS27) from sample JS27 and the near-full-length genomic sequence of one AcV-1 variant (Z2) from sample Z2 were individually determined by RT-PCR using primers (**Supplementary Table S3**) designed on the basis of the contig sequences derived from the RNA-seq and/or the sequence of the reported isolate K75. The complete genomic sequence of AcV-1 variant WH4-2 and the near-full-length genomic sequence of AcV-1 variant WH4-1 from sample WH4 were individually determined by RT-PCR using primers (**Supplementary Table S3**) designed on the basis of available AcV-1 sequences. Most of the adjacent fragments overlapped by more than 100 bp. The 5′ and 3′ terminal sequences of the viral genomic RNA were determined by a RACE strategy using a commercial kit (Invitrogen GeneRacer Kit, United States) according to the manufacturer's instructions and specific reverse primers (**Supplementary Table S3**). RT-PCR solutions and conditions were like those reported previously (Zheng et al., 2017), except that annealing temperature and extension time varied, depending on the primer sets used in each reaction and the sizes of PCR products. Sequences of RT-PCR fragments from each sample were assembled into a continuous sequence using DNAMAN6.0 (Lynnon, Quebec, Canada).

RT-PCR Detection of AcV-1

The primer set cp-F/cp-R (5′-TGAGCTRGGRATAGATGTTGC-3′/5′-TCTCTCAGGGTTMGGATGAGT-3′) (Wen et al., 2018) designed on the basis of the conserved sequence of the CP gene (positions 16,607–16,627 nt/16,962–16,982 nt of isolate K75 genome) of AcV-1 was used for the specific detection of AcV-1 in kiwifruit leaf samples. Total RNA was extracted from the collected leaf materials (each about 0.5 g) using a CTAB method (Fan et al., 2015). The RNA was diluted in 40 μL RNase-free water. The 20-μL cDNA synthesis solution contained 9 μL RNA template, 1 μL p(N)6 random primer, 0.5 μL Moloney murine leukemia virus (M-MLV) reverse transcriptase, 0.5 μL recombinant RNase inhibitor (Takara, Dalian, China), 4 μL 5 × M-MLV buffer, 1 μL 2.5 mM dNTP, and 4 μL RNase-free water. A 25 μL PCR reaction solution contained 2.0 μL of the cDNA template, 2.5 μL of 10 × reaction buffer, 1.5 μL of 2.5 mM dNTP, 0.5 μL each of 10 mM primers, 0.25 μL r*Taq* (5 U/μl) (TakaRa, Dalian, China), and 17.75 μL ddH_2O. The cycling parameters were set as follows: denaturation at 95°C for 5 min, followed by 35 cycles of denaturation (30 s at 95°C), annealing (30 s at 54°C), and extension (45 s at 72°C). The PCR products were separated on 1.2% agarose gel, stained with ethidium bromide, and visualized under UV light.

Cloning and Sequencing of RT-PCR Products

PCR products were gel purified, ligated into the pMD18-T vector (Takara, Dalian, China), and subsequently transformed into cells of *Escherichia coli* strain Top10. Positive clones were identified by PCR using the *E. coli* cultures as templates. At least three positive clones of each PCR product were sequenced at Shanghai Sangon Biological Engineering & Technology and Service Co., Ltd., Shanghai, China.

Sequence Analyses

ORFs were predicted using the Open Reading Frame Finder at https://www.ncbi.nlm.nih.gov/orffinder/. Conserved domains in predicted proteins were identified by using the Conserved Domain Database (CDD)[2] websites in NCBI. Multiple alignments and identity analyses of nucleotide and amino acid sequences were performed using the ClustalW2 program[3]. Phylogenetic trees were constructed using the neighbor-joining method with 1,000 bootstrap in MEGA 7.0.

Recombination Analysis

Genomic sequences of AcV-1 variants were examined for potential recombination events using the Recombination Detection Program (RDP) version RDP4.94 with default parameter settings (Martin et al., 2015). The RDP software includes the RDP, GENECONV, BOOTSCAN, MaxChi, Chimaera, and SISCAN methods. AcV-1 genomic sequences were aligned in MEGA v.7.0.1 software and exported to the RDP program to generate evidence of recombination. At a Bonferroni-corrected p-value cutoff ≤ 0.05, recombinant sites detected by four or more algorithms in the RDP were considered as recombination events.

[1] http://www.ncbi.nlm.nih.gov/

[2] https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

[3] http://www.ebi.ac.uk/Tools/msa/clustalw2/

RESULTS

Viruses Identified by RNA-Seq Analysis

In total, 67,640,546, 92,477,928, 73,387,306, 82,730,664, 79,782,996, and 73,419,336 clean reads were obtained and 177,848, 315,308, 165,312, 246,566, 221,290, and 174,235 unique contigs were generated from RNA-seq samples JS, ZZ, Z2, CQ, JX, and JX5, respectively (**Supplementary Table S2**). BlastX searches against the NCBI NR database revealed that 254, 154, 137, 620, 73, and 18 contigs (accounting for 0.01–0.25% of total contigs) from the six samples respectively matched with viral proteins (**Supplementary Table S2**). AcV-1 was recovered from all these RNA-seq samples, and 121 contigs were assigned to this virus, accounting for 9.63% of the total viral contigs (**Table 1**). Meanwhile, five other well-documented kiwifruit viruses AcCRaV, AcVA, AcVB, CLBV, and *Actinidia* seed-borne latent virus (ASbLV) (Veerakone et al., 2018) were identified in sample JS. Four other viruses, including ASGV, *Actinidia* emaravirus 2 (AcEV-2) (Wang et al., 2020), and two unreported kiwifruit-infecting viruses, belonging to families Betaflexiviridae and Secoviridae, were identified in samples ZZ and Z2. In sample CQ, eight viruses, including AcV-1, AcCRaV, ASbLV, AcVA, AcVB, CLBV, ASGV, and an unreported virus in the family Secoviridae, were identified. In sample JX, five known viruses, including AcV-1, ASbLV, AcVA, AcVB, and CLBV, were identified. In sample JX5, only one known virus AcV-1 was identified. Additionally, a virus potentially belonging to the family Bromoviridae was identified in samples JX and JX5. Here, only AcV-1 sequences were considered for further analyses.

AcV-1 Sequences Identified by RNA-Seq Analysis

From sample JS, 16 contigs matched the genomic sequence of AcV-1. One contig of 18,779 nt was near the full length of AcV-1 genome sequence, while the remaining contigs were 395–1,467 nt long. These contigs shared 58.0–90.0% nt sequence identity with corresponding sequences of isolate K75 (**Figure 1**). Comparison of these short contigs and the 18,779 nt contig revealed 59.6–92.9% nt sequence identity with each other.

From sample ZZ, 18 AcV-1 contigs with sizes ranging from 224 to 4,132 nt covered 71.5% of the viral genome and displayed 62.1–88.2% nt sequence identity with corresponding sequences of AcV-1 isolate K75 (**Figure 1**).

From sample Z2, 26 contigs ranging from 248 to 6,041 nt in length matched the genomic sequence of AcV-1, with a total coverage of 97.9%. These contig sequences shared 53.1–88.2% nt identity with corresponding sequences of isolate K75 (**Figure 1**).

From sample CQ, 37 contigs matched the genomic sequence of AcV-1. One contig (named CQ-c13707) was 13,707 nt in length and covered the ORF1a-ORF4 (position at 248–14,118 nt) of the viral genomic sequence. The remaining contigs were 212–7,590 nt in length. These contigs shared 55.2–94.3% nt sequence identity with corresponding sequences of isolate K75 (**Figure 1**).

From sample JX, 14 contigs matched the genomic sequence of AcV-1. One large contig (named JX-c18869) had a size of 18,869 nt, nearly covering the full-length of AcV-1 genome

TABLE 1 | Viruses and their sequences identified from six kiwifruit samples by using RNA-seq analysis.

Virus[a]	JS		ZZ		Z2		CQ		JX		JX5	
	Reads	Contigs	Reads	Contigs	Reads	Contigs	Reads	Contigs	Reads	Contigs	Reads	Contigs
Actinidia virus 1	34,773	16	1,136	18	3,153	26	9,586	37	3,014	14	7,896	10
Actinidia chlorotic ringspot-associated emaravirus	27,982	34	11,820	5	72	5	75,516	33				
Actinidia seed-borne latent virus	46,066	12	12	2	27	2	70,643	28	24,626	2		
Actinidia virus A	9,149	13	3,347	32	11,986	13	104,048	57	966	12		
Actinidia virus B	38,338	16	706	9	18,297	4	317,494	27	55	4		
Apple stem grooving virus			4,947	11	19	5	48,357	20				
Citrus leaf blotch virus	107,584	32	102,543	3	72	8	54,632	3				
Actinidia emaravirus 2			1,034	15	28,572	4						
Bromoviridae									192,019	24	189	5
Betaflexiviridae			19	3	31	4						
Secoviridae			2,333	1	234	2	277,845	14				

[a]*Three virus species, which are firstly identified in kiwifruit samples, are indicated by their family names.*

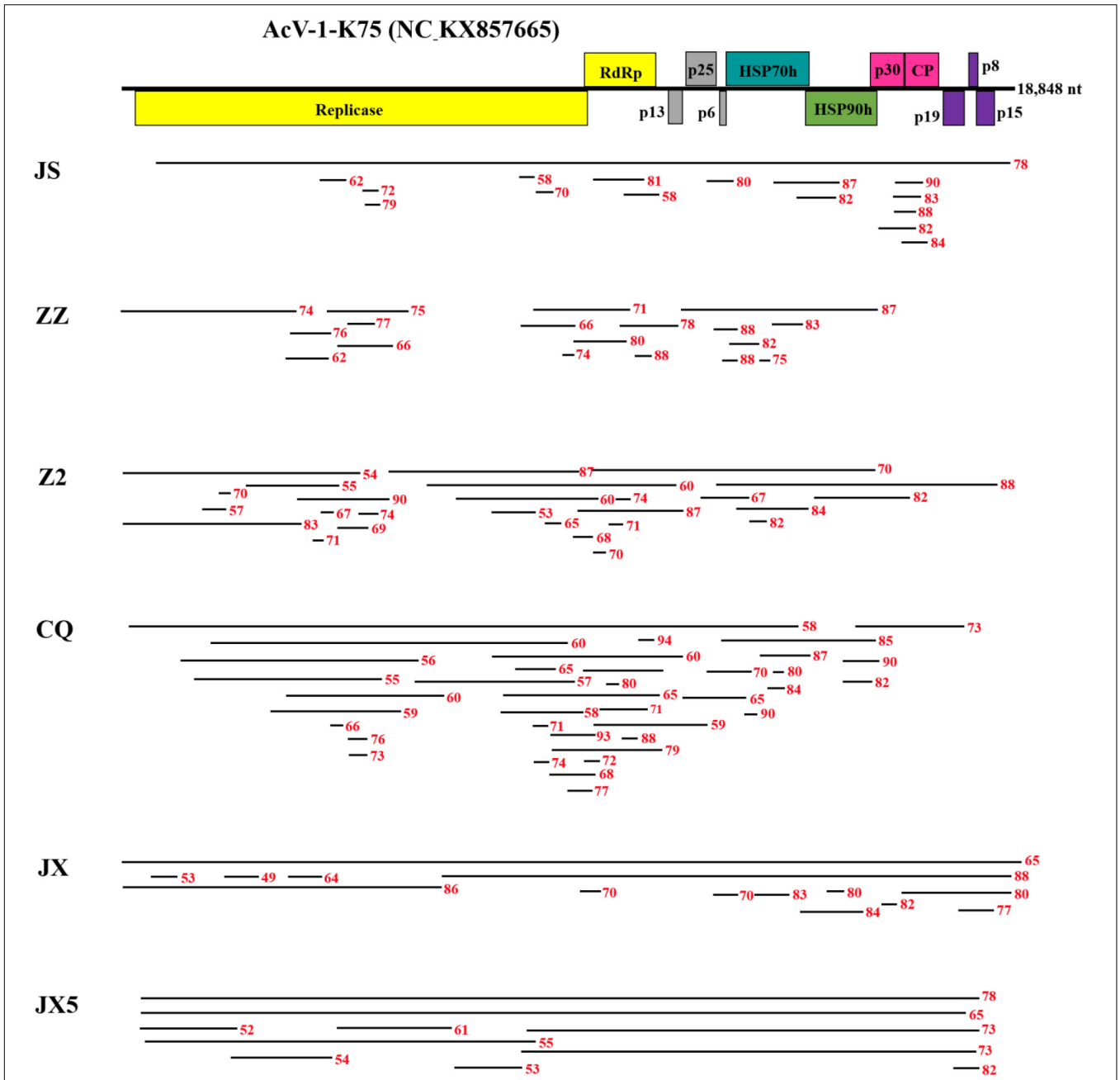


FIGURE 1 | Contigs with significant nucleotide sequence identity to *Actinidia* virus 1 (AcV-1) isolate K75. A schematic diagram of the genome organization of AcV-1 isolate K75 was used as reference showing the contig positions. Six samples JS, ZZ, Z2, CQ, JX, and JX5 were named with the abbreviation of their geographic origins. Black lines represent AcV-1 contigs and their position at the genome of isolate K75. Red numbers indicate the nucleotide identity between each contig and the corresponding sequence of AcV-1 isolate K75.

sequence, and another contig (named JX-c12007) was 12,007 nt (position at 6,772–18,873 nt) covering about 64% of the viral genomic sequence. The remaining contigs were 311–6,776 nt in length. All these contigs shared 49.0–88.3% nt sequence identity with corresponding sequences of isolate K75 (**Figure 1**).

From sample JX5, 10 contigs matched the genomic sequence of AcV-1. Two contigs (named JX5-c18818 and JX5-c18410) were 18,818 and 18,410 nt in length, respectively, and covered nearly the complete genome of AcV-1. Meanwhile, another two contigs (named JX5-c10087 and JX5-c10084) matched approximately in the same position (8,710–18,800 nt) of the viral genome. The four contigs shared 63.3–86.4% nt sequence identity with each other and 65.2–78.1% nt sequence identity with corresponding sequences of isolate K75. The remaining six contigs were 667–8,857 nt and shared 54.4–78.3% nt sequence identity with corresponding sequences of AcV-1 isolate K75 (**Figure 1**).

For the following analyses, each contig derived from RNA-seq analyses was named as a sample name followed by the contig length, and each sequence was considered as a molecular variant.

Genomic Characteristics and Sequence Diversity of Chinese AcV-1 Variants

A complete genomic sequence (JS27, 18,896 nt, GenBank accession no. MT936297) of AcV-1 from the sample JS27 was reconstructed by Sanger sequencing of RT-PCR and 3′ and 5′ RACE products using primers designed on the basis of sequence of a large contig (18,779 nt) derived from RNA-seq. In addition, a near-full-length genomic sequence (Z2, 18,685 nt, excluding about 300 nt at the 3′ terminus) (GenBank accession no. MT936306) of AcV-1 infecting an *A. delicious* plant (ID: Z2) was reconstructed by RT-PCR using primers designed on the basis of sequences in the RNA-seq contigs. The two reconstructed genomic sequences showed greater than 98.4% identity at nt level with corresponding contig sequences generated from RNA assembly, indicating that the assembled contig sequences from RNA-seq data were reliable. Meanwhile, a complete genomic sequence of 18,851 nt (WH4-2) (GenBank accession no. MT936305) and a near-full-length genomic sequence (WH4-1, 16,871 nt, excluding about 2,000 nt at the 5′ terminus) (GenBank accession no. MT936304) of AcV-1 infecting an *A. arguta* plant (ID: WH4) were determined by Sanger sequencing of RT-PCR and 3′ and 5′ RACE products using primers designed on the basis of all available sequences.

These sequences together with near-full-length genomic sequences of three contigs (JX-c18869, JX5-c18818, and JX5-c18410) (GenBank accession nos. MT936303, MT936301, and MT936300), four contigs (CQ-c13707, JX-c12007, JX5-c10087, and JX5-c10084) (GenBank accession nos. MT936296, MT936302, MT936299, and MT936298) with sizes > 10,000 nt obtained from RNA-seq, and the genomic sequence of K75 were compared with each other (**Table 2**). The genome-wide nt sequence identity among the Chinese AcV-1 variants was about 60–87%. All these sequences showed 68.3–79.3% nt identity with that of isolate K75 reported from New Zealand. Moreover, high diversity was present in the AcV-1 sequences derived from the same sample. The two variants WH4-1 and WH4-2 from sample WH4 showed 62.6% nt sequence identity with each other, and the four contigs JX5-c18818, JX5-c18410, JX5-c10087, and JX5-c10084 from sample JX5 showed 63.3–86.5% nt sequence identity with each other. The genome structure of these Chinese AcV-1 variants was identical to that of K75 and consisted of 12 ORFs. There was variation in the length of the 5′ untranslated region (UTR) of genomic RNA among the Chinese AcV-1 variants and isolate K75. The 5′-UTR lengths of three AcV-1 variants Z2, JS27, and WH4-2 were 416, 290, and 288 nt, with 39.7, 42.9, and 78.3% nt sequence identities between Z2 and WH4-2, Z2 and JS27, and WH4-2 and JS27. These sequences shared 64.3–79.3% identity with 304-nt 5′-UTR of K75. The 3′-UTR lengths of AcV-1 variants JS27, WH4-2, and WH4-1 were 253, 252, and 252 nt, which were about 152 nt shorter than the 404-nt 3′-UTR of K75. The 3′-UTR of these Chinese AcV-1 variants shared 74.1–92.1% nt sequence identity with each other and 80.0–88.1% nt sequence identity with that of K75.

Sequence alignment for each ORF (**Table 2**) encoded in the above genomic and contig sequences showed that sizes of ORF5, ORF6, ORF8, and ORF9 of these Chinese AcV-1 variants were the same as the corresponding ORFs of K75, except for the ORF6 in variant JS27. The ORF1b size of these Chinese AcV-1 variants was the same (1,527 nt), but 6-nt shorter than that of K75. The sizes of other ORFs were variable among these sequences. The ORF1a sizes showed the highest variation, ranging from 9,543 to 9,666 nt. ORF7 and ORF10 had three different sizes, and ORFs 2, 3, 4, and 11 had two sizes. Moreover, ORF11 sequences determined in this study were 165 nt or 171 nt larger than the ORF11 (405 nt) of K75 due to 6-nt and 164-nt deletions located near the 5′ and 3′ termini of ORF11 of K75, respectively (**Supplementary Figure S1**).

High nucleotide and amino acid sequence diversity occurred in each ORF of the Chinese AcV-1 populations (**Table 2**). Generally, ORF1a, ORF2, and ORF3 near the 5′ terminus of AcV-1 genome were more variable. ORF1a of Chinese AcV-1 variants shared 59.5–86.1% nt sequence identity with each other and 60.0–73.2% nt sequence identity with the ORF1a of K75. RdRp encoded in ORF1b of all AcV-1 variants had an AcV-1 specific 26-amino acid (aa) insert located immediately after the GDD motif, when it was compared to RdRp encoded by other viruses in the family Closteroviridae. Multiple alignments showed that the 26-aa insertion sequences were highly variable among AcV-1 variants (**Figure 2A**) and formed two phylogenetic clades (**Figure 2B**). The ORF2 sequences of AcV-1 variants determined in this study were even more variable and showed the highest divergence up to 47.5% at nt sequence level and 58.9% at aa sequence level compared to isolate K75. Nucleotide variations distributed across the entire ORF2 (**Supplementary Figure S2A**) and amino acid variations were mainly present at the C-terminal half of the encoded protein (p13) (**Supplementary Figure S2B**). Similarly, the ORF3 of these Chinese AcV-1 variants shared 56.2–86.4% nt and 50.5–85.7% aa sequence identity with that of K75. ORFs 5, 6, 8, and 9 were relatively conserved among AcV-1 variants by showing greater than 80% nt sequence identity with the corresponding sequences of K75 (**Table 2**).

Phylogenetic analyses for the complete aa sequences of proteins RdRp, HSP70h, and CP, for which aa sequences are specified by ICTV for genus demarcation within the family Closteroviridae, showed that all Chinese AcV-1 variants together with isolate K75 clustered into a clade, distantly related to persimmon virus B (PeVB), an unassigned closterovirus species (Ito et al., 2015) (**Figure 3**). In the AcV-1 clade, Chinese AcV-1 variants formed three subclades represented by JX5-c18410, Z2, and K75. In the HSP90h-, p30-, and p19-based trees, AcV-1 variants also clustered into three clades. It was found that the positions of some AcV-1 variants changed in these phylogenetic trees. For example, in the RdRp- and HSP70h-based trees, WH4-1 was in the subclade represented by JX5-c18410, but clustered in the subclade represented by K75 in the HSP90h-based tree and in the subclade represented by Z2 in the CP-, p30-, and p19-based trees.

TABLE 2 | Nucleotide and amino acid sequence comparison of the genome, ORFs, and UTRs of 11 *Actinidia* virus 1 (AcV-1) variants determined in this study with the corresponding sequences of isolate K75.

Vatiants	Genome		5′UTR		ORF1a			ORF1b (RdRp)			ORF2 (p13)			ORF3 (p25)			ORF4 (p6)			ORF5 (HSP70h)		
	nt	%	nt	%	nt	nt%	aa%	nt	nt%	aa%	nt	nt%	aa%	nt	nt%	aa%	nt	nt%	aa%	nt	nt%	aa%
K75	18,848	—	304	—	9,558			1,533			339			675			156			1,755		
Z2	18,685	68.3	416	64.3	9,597	61.6	61.1	1,527	72.2	80.9	366	56.6	48.2	675	58.4	51.3	156	76.9	86.3	1,755	81.0	88.0
JS27	18,896	79.3	290	77.2	9,600	72.1	74.2	1,527	83.3	92.9	339	85.6	74.1	675	86.4	85.3	156	89.1	94.1	1,755	88.0	92.1
WH4-2	18,851	79.2	288	79.3	9,561	73.2	75.0	1,527	82.2	91.9	339	79.7	74.1	675	85.6	85.7	168	89.1	90.2	1,755	88.7	93.0
WH4-1	16,871	70.6	—	—	—	—	—	1,527	70.4	79.7	366	54.6	48.2	675	58.4	51.8	156	82.1	86.3	1,755	86.1	91.3
CQ-c13707	13,707	63.6	—	—	9,666	60.3	59.4	1,527	71.3	79.9	366	52.5	41.1	768	56.2	55.8	156	82.1	88.2	—	—	—
JX-c18869	18,869	68.9	—	—	9,609	60.1	59.1	1,527	71.7	81.5	366	57.5	44.6	675	56.9	50.5	156	82.7	86.3	1,755	84.1	90.2
JX-c12007	12,007	87.3	—	—	—	—	—	1,527	90.4	94.5	339	80.8	72.3	675	84.6	84.8	156	89.1	92.2	1,755	88.4	93.5
JX5-c18818	18,818	79.2	—	—	9,534	72.7	74.9	1,527	81.3	93.5	339	79.4	71.4	675	83.4	81.7	168	90.4	94.1	1,755	88.2	94.2
JX5-c18410	18,410	68.7	—	—	9,633	60.0	59.4	1,527	71.5	80.1	366	55.2	44.6	675	57.2	50.5	156	82.7	88.2	1,755	84.6	90.1
JX5-c10087	10,087	75.1	—	—	—	—	—	1,527	72.4	81.7	366	56.0	46.4	675	58.8	50.9	156	76.3	86.3	1,755	81.3	88.0
JX5-c10084	10,084	75.8	—	—	—	—	—	1,527	71.2	79.9	366	52.5	41.7	768	56.4	56.3	156	82.1	88.2	1,755	83.8	90.1

Variants	ORF6 (HSP90h)			ORF7 (p30)			ORF8 (CP)			ORF9 (p19)			ORF10 (p8)			ORF11			3′UTR	
	nt	nt%	aa%	nt	nt%	aa%	nt	nt%	aa%	nt	nt%	aa%	nt	nt%	aa%	nt	nt%	aa%	nt	nt%
K75	1,524		—	741		–	732			474			213			405			404	
Z2	1,524	83.0	87.2	741	76.3	69.9	732	83.5	89.7	474	85.0	86.6	213	81.2	72.9	—	—	—	—	—
JS27	1,521	89.7	93.1	744	86.4	81.4	732	90.0	95.9	474	89.9	94.3	216	89.8	84.5	576	62.6	85.1	253	87.8
WH4-2	1,524	88.7	92.7	741	86.5	80.5	732	88.3	93.4	474	88.6	93.6	216	88.4	81.7	570	82.5	82.1	252	88.1
WH4-1	1,524	88.9	93.1	747	81.3	78.5	732	83.1	89.3	474	87.1	88.5	207	77.9	71.4	576	78.8	82.1	252	80.0
CQ-c13707	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
JX-c18869	1,524	83.7	88.8	741	80.6	76.4	732	90.0	90.5	474	83.5	86.0	216	81.0	74.6	576	80.7	81.3	—	—
JX-c12007	1,524	89.6	93.7	738	87.8	83.7	732	88.1	93.8	474	88.6	91.1	213	91.1	87.1	570	86.7	88.1	—	—
JX5-c18818	1,524	89.4	92.9	741	86.6	82.9	732	89.9	94.7	474	87.1	91.7	213	85.5	77.1	570	86.4	85.1	—	—
JX5-c18410	1,524	83.7	88.6	741	80.0	76.0	732	86.5	90.1	474	81.7	84.7	216	79.6	75.7	—	—	—	—	—
JX5-c10087	1,524	83.3	87.0	741	76.4	71.1	732	83.2	89.7	474	85.9	88.5	213	82.2	72.9	570	81.0	79.1	—	—
JX5-c10084	1,524	83.6	89.0	741	82.1	77.6	732	86.2	92.6	474	82.1	86.6	216	80.1	69.0	570	81.2	79.9	—	—

"—" indicates partial or unavailable sequences of the virus genome. AcV-1 variants with full-length genome sequences are in bold font.

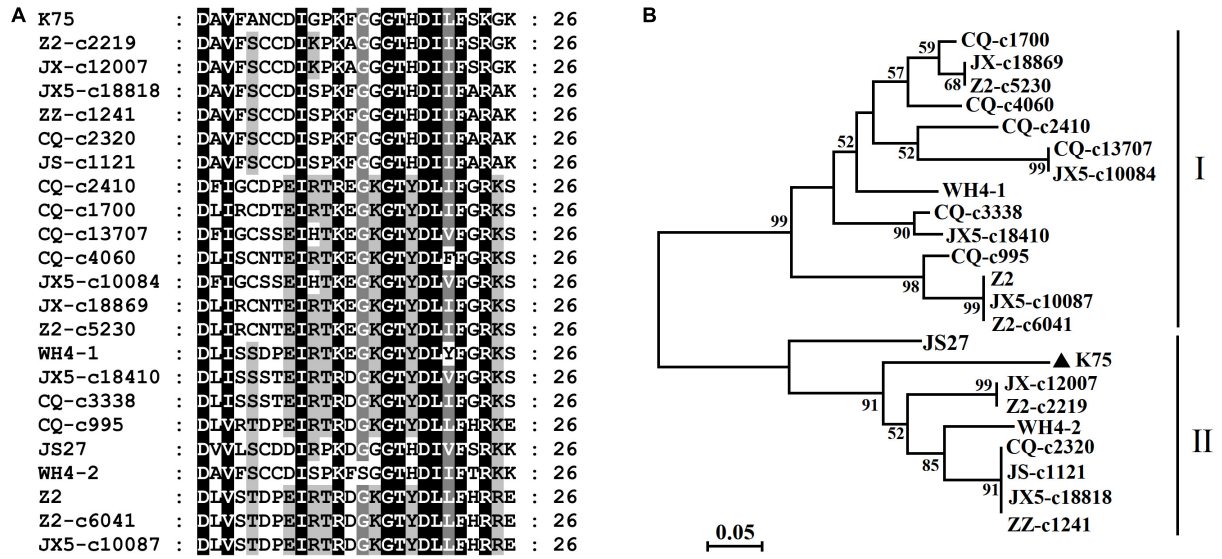


FIGURE 2 | Multiple alignment of the 26-aa insert in RdRp of *Actinidia* virus 1 (AcV-1) variants **(A)** and a neighbor-joining phylogenetic tree inferred from the 26–amino acid (aa) insert sequences **(B)**.

Identification of Domains of AcV-1 Proteins

Putative domains of the proteins predicted from AcV-1 genomic sequences (including assembled contig sequences) were identified by using CDD at https://www.ncbi.nlm.nih.gov/Structure/cdd. It was found that motifs of papain-like leader protease (L-Pro), methyltransferase (MTR, PF01660), and helicase (HEL, PF01443), necessary for virus replication, were present in the multifunctional protein encoded in ORF1a of all AcV-1 variants (**Figure 4**). However, the number of L-Pro domains varied among AcV-1 variants. Notably, the ORF1a of Z2, JX5-c8857, CQ-c7590, and ZZ-c3649 had a single L-Pro domain (L1 or L2), and the ORF1a of other AcV-1 variants had two L-Pro domains (L1 and L2). Sequence analyses revealed the presence of a sec-independent translocase domain (cl35116) in variant Z2 and an ATP-dependent RNA helicase RhiB (cl35267) in variants CQ-c13707 and JX5-c2170. These domains were not identified and documented in the reported AcV-1 isolate K75. Further sequence alignments showed that the sequences covering the two domains were highly variable in other AcV-1 variants (**Supplementary Figures S3A,B**). Additionally, according to the CDD prediction, ORF1a of the AcV-1 variants JS27, CQ-c5050, CQ-c3990, and JX-c6776 harbored a domain (56 or 57 aa) of 2OG-Fe (II) oxygenase superfamily (cl21496) (Aravind and Koonin, 2001; Erwin et al., 2008) upstream of the MET domain. Further alignment showed that all AcV-1 variants had the core sequence of AlkB domain (pfam 03171) (Maree et al., 2008; Ghanem-Sabanadzovic et al., 2012; Donda et al., 2017), belonging to 2OG-Fe (II) oxygenase superfamily, as in some closteroviruses (**Supplementary Figure S3C**).

A thaumatin-like domain was identified in protein p30 encoded by ORF7 of all available AcV-1 variants. The thaumatin-like domain of all available AcV-1 variants was 109 or 115 aa (**Supplementary Figure S4**) and shared 69.9–96.3% aa sequence identity with each other.

Recombination Analysis

The incongruent positions of some variants in phylogenetic trees based on the sequences of different proteins indicated possible recombination events occurring in the Chinese AcV-1 variants. Recombination analysis performed on an alignment of 11 AcV-1 genomic sequences revealed that 11 recombination events were present in eight variants (**Table 3**). All the detected recombination regions located at the 3′ half of the viral genome. Three very close recombination events at positions 13,029–14,361, 14,881–16,595, and 16,614–18,703 nt were detected in WH4-1 with JX5-c18410 as a possible major parent in two recombination events (events 2 and 3) and as a possible minor parent in one recombination event (event 1). The variant WH4-2 from the same sample as WH4-1 had two recombination events (events 4 and 5) with parents different from those of WH4-1. It was found that three recombination events (events 4, 8, and 10) in variants WH4-2, Z2, and JX5-c10087 had a common major parent JX5-c18818. Additionally, two variants JX-c12007 and JS27, for which the start sites of recombination were not detected, had the same parents WH4-2 and K75. AcV-1 variants JX5-c18410, JX5-c10087, and JX5-c10084 were from the same sample (ID: JX5), but had different parents.

RT-PCR Detection of AcV-1 in Kiwifruit Plants Grown in China

To disclose the infection status of AcV-1, leaf samples collected from 245 kiwifruit plants grown in 10 provinces and cities in China were subjected to RT-PCR detection of AcV-1 using the primer set cp-F/cp-R. Results showed that 75 samples, including 70 symptomatic samples and five asymptomatic samples, were

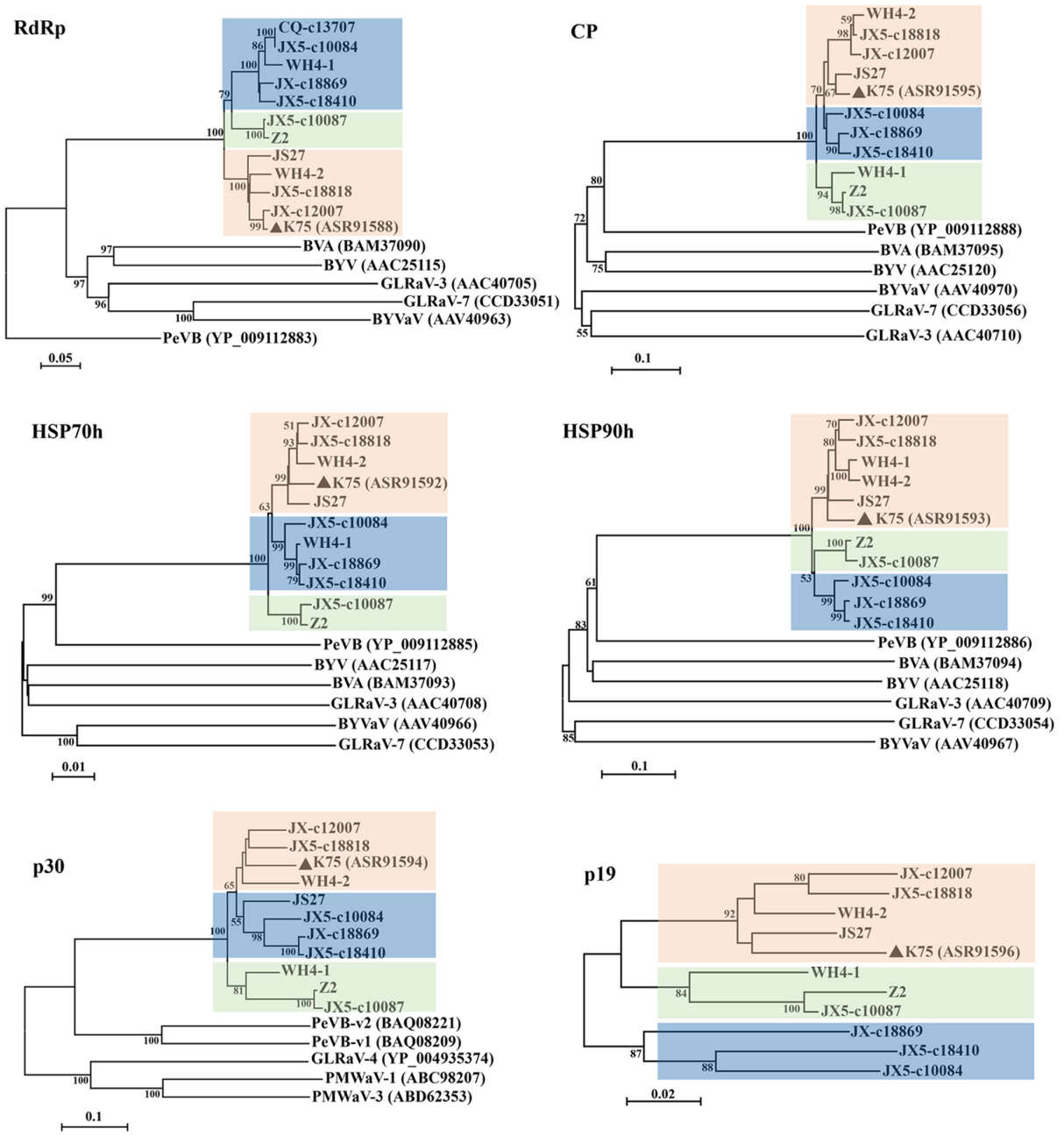


FIGURE 3 | Phylogenetic trees constructed using amino acid (aa) sequences of putative RdRp, CP, HSP70h, HSP90h, p30, and p19 of *Actinidia* virus 1 (AcV-1) variants. The genomic sequence of AcV-1 isolate K75, two complete (JS27 and WH4-2) and two near-complete (Z2 and WH4-1) genomic sequences determined in this study, and seven RNA-seq derived contigs with sizes > 10,000 nt, were included in the analysis. The contig CQ-c13707 covering the ORF1a-ORF4 (position 248–14,118 nt) of the viral genome was only included in the RdRp-based tree. Bootstrap values (1,000 replicates) > 50% were shown at branch nodes. The reported AcV-1 isolate K75 was marked by a black triangle. GenBank accession numbers of viruses referred for phylogenetic analyses were indicated following the abbreviation of each virus name. PeVB, Persimmon virus B (AB923924); BVA, blueberry virus A (AB733585); BYV, beet yellows virus (AF056575); BYVaV, blackberry yellow vein-associated virus (AY776334); GLRaV-3, grapevine leafroll-associated virus 3 (AF037268); GLRaV-7, grapevine leafroll-associated virus 7 (HE588185).

positive for AcV-1 (**Table 4**), accounting for 30.6% of total samples. The virus was detected in four analyzed species, including *A. chinensis*, *A. delicious*, *A. arguta*, and *A. eriantha*, and in some other kiwifruit plants with species unknown. Of these AcV-1–positive samples, two samples ZZ4 and CQ11 were negative for other detected viruses, and 73 samples were coinfected by AcV-1 with one or more other viruses, including AcVA, AcVB, ASGV, CLBV, and AcCRaV (**Supplementary Table S4**), and/or unreported kiwifruit-infecting viruses (data not shown). Except for samples from Yunnan Province and Shanghai

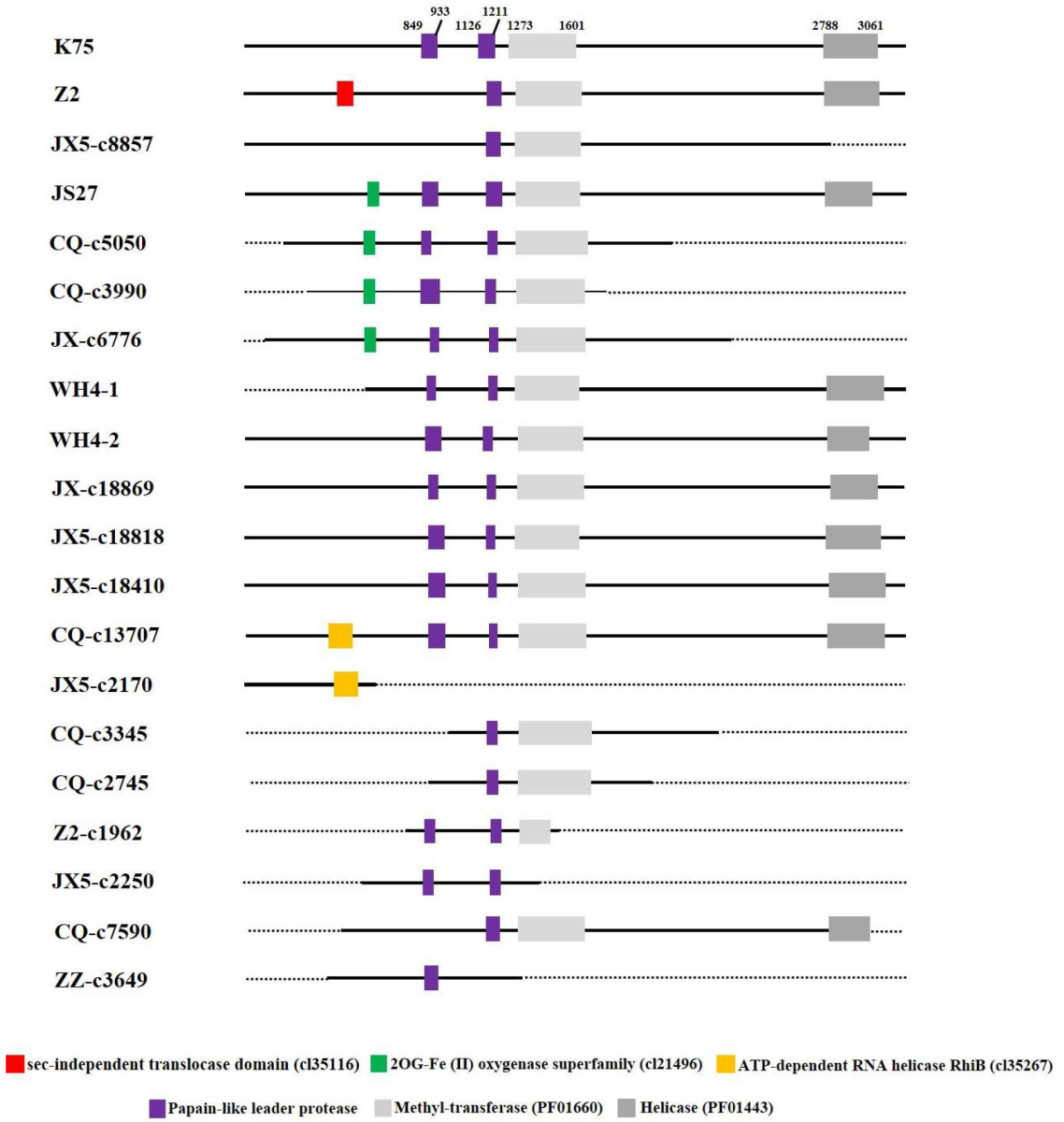


FIGURE 4 | Schematic diagram of domains contained in ORF1a of *Actinidia* virus 1 (AcV-1) variants. The starting and ending sites of each domain in the ORF1a of AcV-1 isolate K75 are indicated by black numbers. The sec-independent translocase domain (cl35116), domain belonging to the 2OG-Fe (II) oxygenase superfamily (cl21496), and ATP-dependent RNA helicase RhiB (cl35267) are marked as red, green, and orange boxes, respectively. MET and HEL are shown as light gray and dark gray boxes, respectively. Black solid lines represent available sequences, and black dotted lines indicate unavailable sequences.

city, where all collected samples were negative for AcV-1 in this test, the virus was detected in samples from seven provinces and one city, with relative high infection rates for samples from Hubei and Shandong provinces and Chongqing city (**Table 4**).

The 376 bp PCR products obtained using the primer set cp-F/cp-R from 75 AcV-1–positive samples were sequenced. These sequences together with the corresponding sequences derived from RNA-seq and 13 sequences of CP gene referred from GenBank database were used for phylogenetic analysis. The 376-bp sequences of all Chinese AcV-1 variants showed 78.5–100% nt identity with the equivalent sequence of isolate K75 and clustered into three distinct clades, respectively, represented by the variants K75, WH4-1, and Z2, except for one variant YT5, which was distant from the three clades (**Supplementary Figure S5**). All eight AcV-1 variants, previously reported from Sichuan province (Peng et al., 2018), clustered in the same clade (clade I). It was found that variants from each of samples Z2 and JX5 were distributed in different clades, confirming the sequence variation within the AcV-1 populations.

TABLE 3 | Putative recombination events identified by using RDP4 in *Actinidia* virus 1 (AcV-1) genomic sequences.

Event	Recombinant	Parent		Break point		Av. *P*-value						
		Major	Minor	Start	End	R	G	B	M	C	S	P
1	WH4-1	JX5-c10084	JX5-c18410	13,029	14,361	3.519E-16	1.179E-17	3.952E-16	2.231E-11	4.889E-09	4.236E-20	2.300 E-05
2	WH4-1	JX5-c18410	WH4-2	14,881	16,595	1.200E-127	6.732E-146	2.192E-139	9.358E-25	5.854E-15	2.408E-35	2.198E-13
3	WH4-1	JX5-c18410	Z2	16,614	18,703	3.609E-51	—	1.973E-58	4.951E-22	4.348E-10	1.262E-18	—
4	WH4-2	JX5-c18818	K75	11,310	11,534	7.772E-09	—	6.056E-10	2.336E-05	2.050E-06	—	2.777E-07
5	WH4-2	JX-c12007	JX5-c10084	18,618	18,834	6.659E-03	—	—	2.484E-05	3.751E-03	1.461E-06	1.059E-03
6	JS27	K75	WH4-2	—	11,078	3.448E-20	—	7.854E-10	2.855E-19	3.342E-25	3.110E-92	2.442E-14
7	JX-c12007	WH4-2	K75	—	11,518	5.338E-51	3.235E-42	2.153E-43	4.727E-24	2.318E-10	1.618E-75	8.976E-09
8	Z2	JX5-c18818	WH4-1	11,866	13,102	1.596E-16	1.544E-02	1.265E-12	7.382E-08	2.315E-09	3.237E-16	—
9	JX5-c18410	JX-c18869	JX5-c10084	17,834	18,245	1.278E-06	—	7.447E-07	4.517E-07	5.658E-08	—	1.697E-06
10	JX5-c10087	JX5-c18818	Unknown	13,129	18,237	—	—	—	2.539E-09	9.841E-05	9.409E-06	1.580E-05
11	JX5-c10084	WH4-1	Unknown	11,468	12,506	3.903E-04	—	3.443E-03	2.431E-03	1.946E-04	1.581E-10	—

"—" indicates that no recombination event was predicted.

TABLE 4 | Incidence of *Actinidia* virus 1 (AcV-1) in the kiwifruit samples collected from eight provinces and two cities in China.

Origin	No. of Samples	Infected/tested (%)	
		Symptomatic[a]	Asymptomatic
Hubei	71	27/63 (42.9)	3/8
Jiangxi	14	1/13 (7.7)	1/1
Shandong	28	16/28 (57.1)	0/0
Zhejiang	31	4/29 (13.8)	0/2
Henan	32	4/32 (12.5)	0/0
Fujian	7	1/6 (16.7)	0/1
Shanxi	4	3/4 (75)	0/0
Chongqing	36	14/34 (41.2)	1/2
Yunnan	20	0/20	0/0
Shanghai	2	0/2	0/0
Total	245	70/231 (30.3)	5/14 (35.7)

[a]*The symptoms of the tested samples include uneven chlorosis, mosaic, ringspot, chlorotic spot, and deformation.*

DISCUSSION

NGS combined with bioinformatics analysis has become a routine technology for the rapid discovery and characterization of known or novel plant viruses (Adams et al., 2009; Kreuze et al., 2009; Barba et al., 2014; Wu Q. et al., 2015; Hadidi et al., 2016; Cretazzo and Velasco, 2017; Villamor et al., 2019; Zhao et al., 2020; Wang et al., 2020; Zhang et al., 2020). NGS-based metagenomic analyses have revealed the natural biodiversity of plant viruses (Roossinck, 2011; Roossinck et al., 2015). In this study, the RNA-seq analyses of two samples individually collected from two kiwifruit plants and four pooled kiwifruit samples revealed the presence of the eight kiwifruit-infecting viruses AcV-1, AcCRaV, ASbLV, AcVA, AcVB, CLBV, AcEV-2, and ASGV, and three viruses related to the families Bromoviridae, Betaflexiviridae, and Secoviridae. These data indicate that NGS technology is useful for identifying viruses infecting kiwifruit plants.

AcV-1 was identified from all six samples analyzed by RNA-seq. Except for the sample ZZ, from which AcV-1 contigs with size of 224–4,132 nt covering 71.5% of the viral genome were recovered, large contig sequences (>5,000 nt) of AcV-1 genome were obtained from the remaining five RNA-seq samples, indicating that the RNA-seq analyses were efficient for the genomic and molecular characterization of the virus. The RNA-seq sample ZZ was a pool of leaf samples from 30 kiwifruit plants, of which only two were positive for AcV-1. The relatively low AcV-1 titer in the mixed sample ZZ and molecular diversity of the virus in the two sampled plants might contribute to the absence of large contig sequences from the sample. The Sanger sequencing for the complete genome of AcV-1 variants JS27 and Z2 also confirmed the reliability of NGS sequence assembly for the viral genome. RNA-seq analyses also revealed mixed infection of AcV-1 variants in its natural host. Two contigs covering nearly the complete genome of AcV-1 were assembled from the sample JX5 and shared about 63% nt sequence identity, and the remaining contigs from the sample shared less than 85% nt sequence identity

with each other. The constitution of AcV-1 contigs from sample Z2 was even more complicated. From the sample, 26 AcV-1 contigs were derived, but could not be assembled as contiguous sequences, indicating high sequence heterogeneity of these AcV-1 contigs, which was also the case in the pooled sample CQ. Consistent with RNA-seq analyses, Sanger sequencing for RT-PCR products revealed at least two divergent AcV-1 variants (WH4-1 and WH4-2) in the sample WH4. Like viruses infecting other perennial woody plants, viruses can be transmitted among kiwifruit plants through vegetative propagation. Vectors play important roles in the natural transmission of viruses in the family Closteroviridae, which might increase the virus diversity in a single host plant (Iglesias et al., 2008; Roy and Brlansky, 2009; Rubio et al., 2013).

The complete genomic sequences of variants JS27 and WH4-2 determined here were 18,896 and 18,851 nt, respectively, close to the 18,848 nt of reported isolate K75 (Blouin et al., 2018). However, the 3′-UTR of JS27, WH4-2, and a partially sequenced variant WH4-1 was about 150 nt shorter than that of K75. Analyses of the two complete genomic sequences of AcV-1 variants JS27 and WH4-2 and nine near-complete genomic sequences of AcV-1 variants Z2, WH4-1, CQ-c13707, JX-c18869, JX-c12007, JX5-c18818, JX5-c18410, JX5-c10087, and JX5-c10084 derived from different kiwifruit plants and geographical areas in China revealed high genetic divergence across the viral genome, especially in the 5′-UTR and the ORFs (ORF1a, ORF2, and ORF3) located in the 5′ half of the viral genome. The sequence variation might impede primer design to amplify the viral 5′ terminal sequence of variant WH4-1. The 5′ UTR could play a regulatory role in virus replication (Mawassi et al., 1996; Liveratos et al., 2004; Mongkolsiriwattana et al., 2016; Jarugula et al., 2018). The 5′ UTR of GLRaV-3 contains critical elements required for virus replication (Jarugula et al., 2010; Jarugula et al., 2018; Adiputra et al., 2019). Similarly, the 5′ UTR of CTV is also highly variable (Lopez et al., 1998), but can form stable SL (stem-loop) structures, which are essential for viral replication and assembly (Tatineni et al., 2002; Gowda et al., 2003).

The ORF1a of viruses in the family Closteroviridae encodes a multifunctional protein (Rubio et al., 2013). The size of ORF1a differed among all available AcV-1 variants with divergence up to 40% at nt sequence level. Previously, the ORF1a size variation was also found in GLRaV-2 (Meng et al., 2005) and GLRaV-5 (Thompson et al., 2012). Overall, the organization of ORF1a of the AcV-1 genome is similar to that of members in the family Closteroviridae (Agranovsky et al., 1995; Karasev, 2000; Dolja et al., 2006). In this study, we found that AcV-1 variants might have one or two L-Pro domains (L1 and L2). Some viruses in the family Closteroviridae possess a tandem of papain-like cysteine proteases with two divergent functional domains L1 and L2, which might have evolved via gene duplication (Karasev, 2000; Peng et al., 2001; Dolja et al., 2006; Liu et al., 2009). The absence of domain L1 in some AcV-1 variants indicated that the domain L2 might be indispensable, which was different from the case that domain L1 plays a crucial role in the establishment of infection and accumulation of some viruses in the family Closteroviridae (Liu et al., 2009; Kang et al., 2018). Except for motifs L-Pro, MTR, and HEL, necessary for viral replication, the ORF1a of some AcV-1 variants harbors a sec-independent translocase domain, or an ATP-dependent RNA helicase RhiB, or an 2OG-Fe (II) oxygenase domain, suggesting that these domains might have been acquired relatively recently via horizontal gene transfer (Bratlie and Drablos, 2005). The AlkB domain is present in the polyproteins of some viruses in the family Closteroviridae, as well as in alpha-like plant viruses belonging to genera *Allexivirus*, *Carlavirus*, *Foveavirus*, *Potexvirus*, *Trichovirus*, and *Vitivirus* (Erwin et al., 2008). The AlkB domain usually locates between MET and HEL domains, whereas the AlkB domain of AcV-1 variants identified in this study locates upstream of the MET domain. Certain cellular AlkBs are involved in RNA repair via methylation reversal (Aravind and Koonin, 2001). However, the function of AlkB in plant virus infection is unknown. The variation of L-Pro numbers and other domains in the multifunctional protein might show not only the diversity of the AcV-1 populations, but also different biological features (Liu et al., 2009; Atallah et al., 2016; Kang et al., 2018). The p30 encoded in AcV-1 ORF7 has a thaumatin-like domain like that in the p29 of PeVB-v1 (GenBank accession no. AB923924) and the p28 of PeVB-v2 (GenBank accession no. AB923925). Although the location of p30 in the AcV-1 genome is similar to the CPm of some closteroviruses, the p30 did not show aa sequence identity with the CPm encoded by other viruses in the family Closteroviridae. The function of the protein remains to be unknown.

The ORF-by-ORF comparisons also showed that the sequences of ORF2 and ORF3 were highly variable among the AcV-1 variants, with the highest divergence up to 58 and 49% at aa sequence level, respectively. Similar diversities are observed among divergent variants of some closteroviruses, such as GLRaV-3 (Maree et al., 2008), GLRaV-4 (Adiputra et al., 2019), and PeVB (Ito et al., 2015). However, other ORFs were much more conserved in AcV-1 populations. Among the viral variants, the aa divergence values of 10–20% for CP, HSP70h and RdRp are less than the species discriminating threshold of 25% approved by ICTV. Therefore, we contend that all AcV-1 sequences represent divergent variants of the virus, which often occurred in a single kiwifruit plant as illustrated by RNA-seq analyses. Phylogenetic analyses for the complete aa sequences of proteins RdRp, HSP70h and CP showed that all Chinese AcV-1 variants together with isolate K75 clustered into a clade, distantly related to PeVB. In the AcV-1 clade, Chinese AcV-1 variants formed three subclades. The results further support that the sequences determined from kiwifruit plants grown in China belong to AcV-1 variants, which are highly divergent in molecular composition.

Recombination in RNA viruses has been extensively documented as a powerful driving force for generating new variants (Karasev, 2000; Chare and Holmes, 2006; Farooq et al., 2012; Katsiani et al., 2015; James and Phelan, 2016; Ruiz et al., 2018). Frequent homologous recombination events have been reported for viruses in the family Closteroviridae (Karasev, 2000; Turturo et al., 2005; Sztuba-Solińska et al., 2011; Rubio et al., 2013). The incongruent positions of some AcV-1 variants in the phylograms based on proteins coded near the 3′ half of AcV-1 genome suggested occurrences of recombination events.

Although the parents for these recombinants were identified, some parents were also potential recombinants. The exact parents for these recombinants need to be clarified as more genomic sequences available from a wide host source. The data will help to understand whether an ancient recombination event occurred before the recombinants spread worldwide, as previously suggested for CTV (Roy et al., 2005; Rubio et al., 2013; Wu G. et al., 2015). Most of the recombination events occurred in the region following the viral ORF1b. In CTV, the genes encoded in the region play important roles in the determination of pathogenicity, movement, and host range (Folimonova, 2020). The frequent recombination events in the region of AcV-1 genome might be necessary to meet possible selection pressure (Roy and Brlansky, 2009).

Finally, RT-PCR detection showed that AcV-1 occurred in four *Actinidia* species and some kiwifruit plants of unknown species grown in seven provinces and one city in China. Previous investigation showed that mixed infection of several viruses in a single kiwifruit plant was very common (Zheng et al., 2014, 2017; Blouin et al., 2018; Veerakone et al., 2018; Wang et al., 2020). Most of the plants tested in this study showed viral disease–like symptoms, including uneven chlorosis between veins, mosaic, ringspots, chlorotic spots, and/or deformation, which were indicated previously to be associated with some virus infections. However, some asymptomatic plants were also positive for the virus. Although two AcV-1–infected samples ZZ4 and CQ11 exhibited interveinal chlorosis and were negative for other kiwifruit viruses known in China, we could not conclude that the disease symptom was caused by the virus infection due to the potential infection of unknown viruses. Therefore, the association of AcV-1 infection with diseases of kiwifruit plants is unclear at present. The biological properties of AcV-1 need to be determined in future studies.

All nucleotide sequences obtained in this study supported high molecular diversity in AcV-1 populations from kiwifruit plants grown in China. This represents the first report for the complete genome sequences of AcV-1 in China and molecular diversity of AcV-1. The reported AcV-1 sequences would be helpful for further detailed taxonomic study and assignment of efficient molecular diagnostic techniques of AcV-1 to improve the sanitary status of kiwifruit planting materials in China.

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

NH and GW supervised the project, conceived and designed the experiments. SW, ZY, MR, QL, and YW performed the experiments and analyzed the data. SW and NH wrote the manuscript, prepared the tables and figures, revised and approved final version of the manuscript. This is the first submission of the manuscript and it is not being considered for publication elsewhere in part or in whole. All authors approved the submission of this manuscript.

FUNDING

This work was financially supported by the Key National Project (Grant No. 2019YFD1001800); the program for Key International S&T Cooperation Proposal (Grant No. 2017YFE0110900); the earmarked fund for the China Agriculture Research System (Grant No. CARS-28-15). The funders did not influence the study design, data collection and analysis, decision to publish, or the preparation of this manuscript.

ACKNOWLEDGMENTS

We thank Dr. Ahmed Hadidi for his linguistic assistance and suggestions during the preparation of this manuscript. We thank the team of kiwifruit research at Zhengzhou Fruit Research Institute, Academy of Agricultural Sciences, Zhengzhou, China, and Dr. Hong-Wei Yi at Fruit Research Institute, Chongqing Academy of Agricultural Sciences, Chongqing, China, for kind helps during sampling.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.602039/full#supplementary-material

Supplementary Figure 1 | Multiple alignment of nucleotide sequences of ORF11 of Chinese Actinidia virus 1 (AcV-1) variants and reported isolate K75.

Supplementary Figure 2 | Multiple alignment of nucleotide **(A)** and amino acid **(B)** sequences of ORF2 of Actinidia virus 1 (AcV-1) variants.

Supplementary Figure 3 | Amino acid sequence alignments of sec-independent translocase domain **(A)**, ATP-dependent RNA helicase RhiB domain **(B)** and AlkB domain belonging to the 2OG-Fe (II) oxygenase superfamily **(C)** of Actinidia virus 1 (AcV-1) variants or viruses. Black triangle indicates the AcV-1 variants containing predicated domains. A solid line above aligned sequences marks the core of the AlkB domain and arrows indicate the conserved residues in the AlkB. Accession numbers of sequences used in the analysis include: grapevine leafroll-associated virus 1 (GLRaV-1) isolates WA-CH (KU674796), WA-PN (KU674797) and Canada (NC_016509), GLRaV-3 (GU983863), GLRaV-4 (FJ467503), GLRaV-5 (JX559640), GLRaV-6 (NC_016417), GLRaV-9 (AY297819), little cherry virus 2 (LChV-2; AF531505).

Supplementary Figure 4 | Multiple sequence alignment of thaumatin-like domain in ORF7 (p30) of the Actinidia virus 1 (AcV-1) variants.

Supplementary Figure 5 | Neighbor joining (NJ) phylogenetic tree generated from the nucleotide sequences of partial CP of Actinidia virus 1 (AcV-1) variants. Bootstrap values (1,000 replicates) > 50% are shown at branch nodes. The reported AcV-1 isolate K75 is marked by a black triangle. The sequences referred from GenBank are identified by their GenBank accession numbers. The scale bar is 0.01 substitutions per site.

REFERENCES

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

Adiputra, J., Jarugula, S., and Naidu, R. A. (2019). Intra-species recombination among strains of the ampelovirus grapevine leafroll-associated virus 4. *Virol. J.* 16, 139. doi: 10.1186/s12985-019-1243-4

Agranovsky, A. A., Lesemann, D. E., Maiss, E., Hull, R., and Atabekov, J. G. (1995). "Rattlesnake" structure of a filamentous plant RNA virus built of two capsid proteins. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2470–2473. doi: 10.1073/pnas.92.7.2470

Aravind, L., and Koonin, E. V. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* 2:research0007.1. doi: 10.1186/gb-2001-2-3-research0007

Atallah, O. O., Kang, S., El-Mohtar, C. A., Shilts, T., María, B., and Folimonova, S. Y. (2016). A 5'-proximal region of the citrus tristeza virus genome encoding two leader proteases is involved in virus superinfection exclusion. *Virology* 489, 108–115. doi: 10.1016/j.virol.2015.12.008

Barba, M., Czosnek, H., and Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106–136. doi: 10.3390/v6010106

Bar-Joseph, M., and Mawassi, M. (2013). The defective RNAs of *Closteroviridae*. *Front. Microbiol.* 4:132. doi: 10.3389/fmicb.2013.00001

Bertazzon, N., Borgo, M., Vanin, S., and Angelini, E. (2010). Genetic variability and pathological properties of grapevine leafroll-associated virus 2 isolates. *Eur. J. Plant Pathol.* 127, 185–197. doi: 10.1007/s10658-010-9583-3

Blouin, A. G., Biccheri, R., Khalifa, M. E., Pearson, M. N., Poggi Pollini, C., Hamiaux, C., et al. (2018). Characterization of Actinidia virus 1, a new member of the family *Closteroviridae* encoding a thaumatin-like protein. *Arch. Virol.* 163, 229–234. doi: 10.1007/s00705-017-3610-z

Blouin, A. G., Chavan, R. R., Pearson, M. N., MacDiarmid, R. M., and Cohen, D. (2012). Detection and characterisation of two novel vitiviruses infecting *Actinidia*. *Arch. Virol.* 157, 713–722. doi: 10.1007/s00705-011-1219-1

Blouin, A. G., Pearson, M. N., Chavan, R. R., Woo, E. N. Y., Lebas, B. S. M., Veerakone, S., et al. (2013). Viruses of kiwifruit (*Actinidia* species). *J. Plant Pathol.* 95, 221–235. doi: 10.4454/JPP.V95I2.013

Boulila, M. (2010). Selective pressure, putative recombination events and evolutionary relationships among members of the family *Closteroviridae*: a proposal for a new classification. *Biochem. Syst. Ecol.* 38, 1185–1192. doi: 10.1016/j.bse.2010.12.002

Bratlie, M. S., and Drablos, F. (2005). Bioinformatic mapping of AlkB homology domains in viruses. *BMC Genomics* 6:1. doi: 10.1186/1471-2164-6-1

Chare, E. R., and Holmes, E. C. (2006). A phylogenetic survey of recombination frequency in plant RNA viruses. *Arch. Virol.* 151, 933–946. doi: 10.1007/s00705-005-0675-x

Chavan, R. R., Blouin, A. G., Cohen, D., and Pearson, M. N. (2013). Characterization of the complete genome of a novel citrivirus infecting *Actinidia chinensis*. *Arch. Virol.* 158, 1679–1686. doi: 10.1007/s00705-013-1654-2

Chavan, R. R., Cohen, D., Blouin, A. G., and Pearson, M. N. (2012). Characterization of the complete genome of ribgrass mosaic virus isolated from *Plantago major* L. from New Zealand and *Actinidia* spp.from China. *Arch. Virol.* 157, 1253–1260. doi: 10.1007/s00705-012-1292-0

Clover, G. R. G., Pearson, M. N., Elliott, D. R., Tang, Z., Smales, T. E., and Alexander, B. J. R. (2003). Characterization of a strain of apple stem grooving virus in *Actinidia chinesis* from China. *Plant Pathol.* 52, 371–378. doi: 10.1046/j.1365-3059.2003.00857.x

Cretazzo, E., and Velasco, L. (2017). High-throughput sequencing allowed the completion of the genome of grapevine red globe virus and revealed recurring co-infection with other tymoviruses in grapevine. *Plant Pathol.* 66, 1202–1213. doi: 10.1111/ppa.12669

Cuellar, W. J., Tairo, F., Kreuze, J. F., and Valkonen, J. P. T. (2008). Analysis of gene content in sweet potato chlorotic stunt virus RNA1 reveals the presence of the p22 RNA silencing suppressor in only a few isolates: implications for viral evolution and synergism. *J. Gen. Virol.* 89, 573–582. doi: 10.1099/vir.0

Dolja, V. V., Kreuze, J. F., and Valkonen, J. (2006). Comparative and functional genomics of closteroviruses. *Virus Res.* 117, 38–51. doi: 10.1016/j.virusres.2006.02.002

Donda, B. P., Jarugula, S., and Naidu, R. A. (2017). An analysis of the complete genome sequence and subgenomic RNAs reveals unique features of the Ampelovirus, grapevine leafroll-associated virus 1. *Phytopathology* 107, 1069–1079. doi: 10.1094/PHYTO-02-17-0061-R

Erwin, V. D. B., Omelchenko, M. V., Anders, B., Vibeke, L., Koonin, E. V., Dolja, V. V., et al. (2008). Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic Acids Res.* 17, 5451–5461. doi: 10.1093/nar/gkn519

Esteves, F., Teixeira Santos, M., EirasDias, J. E., and Fonseca, F. (2012). Occurrence of Grapevine leafroll-associated virus 5 in Portugal: genetic variability and population structure in field-grown grapevines. *Arch. Virol.* 157, 1747–1765. doi: 10.1007/s00705-012-1371-2

Fan, X., Hong, N., Dong, Y., Ma, Y., Zhang, Z., Ren, F., et al. (2015). Genetic diversity and recombination analysis of grapevine leafroll-associated virus 1 from China. *Arch. Virol.* 160, 1669–1678. doi: 10.1007/s00705-015-2437-8

Farooq, A. B. U., Ma, Y., Wang, Z., Zhuo, N., Xu, W., Wang, G., et al. (2012). Genetic diversity analyses reveal novel recombination events in grapevine leafroll-associated virus 3 in China. *Virus Res.* 171, 15–21. doi: 10.1016/j.virusres.2012.10.014

Folimonova, S. Y. (2020). Citrus tristeza virus: A large RNA virus with complex biology turned into a valuable tool for crop protection. *Plos Pathog.* 16:e1008416. doi: 10.1371/journal.ppat.1008416

García-Arenal, F., and Fraile, A. (2011). "Population dynamics and genetics of plant infection by viruses," in *Recent Advances in Plant Virology*, eds C. Caranta, M. A. Aranda, M. Tepfer, and J. J. Lopez-Moya (Norfolk: Caister Academic Press), 263–281.

Ghanem-Sabanadzovic, N. A., Sabanadzovic, S., Gugerli, P., and Rowhani, A. (2012). Genome organization, serology and phylogeny of grapevine leafroll-associated viruses 4 and 6: taxonomic implications. *Virus Res.* 163, 120–128. doi: 10.1016/j.virusres.2011.09.001

Gowda, S., Tatineni, S., Ayllon, M. A., Moreno, P., Flores, R., and Dawson, W. O. (2003). The conserved structures of the 5' nontranslated region of citrus tristeza virus are involved in replication and virion assembly. *Virology* 317, 50–64. doi: 10.1016/j.virol.2003.08.018

Hadidi, A., Flores, R., Candresse, T., and Barba, M. (2016). Next-generation sequencing and genome editing in plant virology. *Front. Microbiol.* 7:1325. doi: 10.3389/fmicb.2016.01325

Harper, S. J., Chooi, K. M., and Pearson, M. N. (2008). First report of citrus leaf blotch virus in New Zealand. *Plant Dis.* 92, 1470–1470. doi: 10.1094/PDIS-92-10-1470C

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and VandePol, S. (1982). Rapid evolution of RNA genomes. *Science* 215, 1577–1588. doi: 10.1126/science.7041255

Iglesias, N. G., Gago-Zachert, S. P., Robledo, G., Costa, N., Plata, M. I., Vera, O., et al. (2008). Population structure of citrus tristeza virus from field Argentinean isolates. *Virus Genes* 36, 199–207. doi: 10.1007/s11262-007-0169-x

Ito, T., Sato, A., and Suzaki, K. (2015). An assemblage of divergent variants of a novel putative closterovirus from American persimmo. *Virus Genes* 51, 105–111. doi: 10.1007/s11262-015-1202-0

James, D. (1999). A simple and reliable protocol for the detection of apple stem grooving virus by RT-PCR and in a multiplex PCR assay. *J. Virol. Methods* 83, 1–9. doi: 10.1016/S0166-0934(99)00078-6

James, D., and Phelan, J. (2016). Complete genome sequence of a strain of Actinidia virus X detected in *Ribes nigrum* cv. Baldwin showing unusual symptoms. *Arch. Virol.* 161, 507–511. doi: 10.1007/s00705-015-2678-6

Jarugula, S., Gowda, S., Dawson, W. O., and Naidu, R. A. (2010). 3'-Coterminal subgenomic RNAs and putative cis-acting elements of grapevine leafroll-associated virus 3 reveals 'unique' features of gene expression strategy in the genus *Ampelovirus*. *Virol. J.* 7:180. doi: 10.1186/1743-422X-7-180

Jarugula, S., Gowda, S., Dawson, W. O., and Naidu, R. A. (2018). Development of infectious cDNA clones of grapevine leafroll-associated virus 3 and analyses of the 5′ non-translated region for replication and virion formation. *Virology* 523, 89–99. doi: 10.1016/j.virol.2018.07.023

Kang, S., Atallah, O. O., Sun, Y., and Folimonova, S. Y. (2018). Functional diversification upon leader protease domain duplication in the citrus tristeza

virus genome: role of RNA sequences and the encoded proteins. *Virology* 514, 192–202. doi: 10.1016/j.virol.2017.11.014

Karasev, A. V. (2000). Genetic diversity and evolution of closteroviruses. *Annu. Rev. Phytopathol.* 38, 293–324. doi: 10.1146/annurev.phyto.38.1.293

Katsiani, A. T., Maliogka, V. I., Amoutzias, G. D., Efthimiou, K. E., and Katis, N. I. (2015). Insights into the genetic diversity and evolution of little cherry virus 1. *Plant Pathol.* 64, 817–824. doi: 10.1111/ppa.12309

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., et al. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1–7. doi: 10.1016/j.virol.2009.03.024

Kreuze, J. F., Savenkov, E. I, and Valkonen, J. P. T. (2002). Complete genome sequence and analyses of the subgenomic RNAs of sweet potato chlorotic stunt virus reveal several newfeatures for the genus *Crinivirus*. *J. Virol.* 76, 9260–9270. doi: 10.1128/JVI.76.18.9260-9270.2002

Liu, H., Wang, G., Yang, Z., Wang, Y., Zhang, Z., Li, L., et al. (2020). Identification and characterization of a pear chlorotic leaf spot-associated virus, a novel emaravirus associated with a severe disease of pear trees in China. *Plant Dis.* 104, 2786–2798. doi: 10.1094/PDIS-01-20-0040-RE

Liu, Y., Peremyslov, V. V., Vicente Medina, V., and Dolja, V. V. (2009). Tandem leader proteases of grapevine leafroll-associated virus 2: host-specific functions in the infection cycle. *Virology* 383, 291–299. doi: 10.1016/j.virol.2008.09.035

Liveratos, I. C., Eliasco, E., Muller, G., Olsthoorn, R. C. L., Salazar, L. F., Pleij, C. W. A., et al. (2004). Analysis of the RNA of potato yellow vein virus: evidence for a tripartite genome and conserved 3′-terminal structures among members of the genus *Crinivirus*. *J. Gen. Virol.* 85, 2065–2075. doi: 10.1099/vir.0.79910-0

Lopez, C., Ayllon, M. A., Navas-Castillo, J., Guerri, J., Moreno, P., and Flores, R. (1998). Molecular variability of the 5′- and 3′-terminal regions of citrus tristeza virus RNA. *Phytopathology* 88, 685–691. doi: 10.1094/PHYTO.1998.88.7.685

Maree, H. J., Freeborough, M. J., and Burger, J. T. (2008). Complete nucleotide sequence of a South African isolate of grapevine leafroll-associated virus 3 reveals a 5′ UTR of 737 nucleotides. *Arch. Virol.* 153, 755–757. doi: 10.1007/s00705-008-0040-y

Martelli, G. P., Agranovsky, A. A., Bar-Joseph, M., Boscia, D., Candresse, T., Coutts, R. H. A., et al. (2012). "Family *Cloteroviridae*," 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), 987–1001.

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

Mawassi, M., Mietkiewska, E., Gofman, R., Yang, G., and Bar-Joseph, M. (1996). Unusual sequence relationships between two isolates of citrus tristeza virus. *J. Gen. Virol.* 77, 2359–2364. doi: 10.1099/0022-1317-77-9-2359

Meng, B., Li, C., Goszczynski, D. E., and Gonsalves, D. (2005). Genome sequences and structures of two biologically distinct strains of grapevine leafroll-associated virus 2 and sequence analysis. *Virus Genes* 31, 31–41. doi: 10.1007/s11262-004-2197-0

Mongkolsiriwattana, C., Zhou, J., and Ng, J. C. K. (2016). A 3′ end structure in RNA2 of a crinivirus is essential for viral RNA synthesis and contributes to replication-associated translation activity. *Sci. Rep.* 6:34482. doi: 10.1038/srep34482

Pearson, M. N., Cohen, D., Chavan, R., and Blouin, A. G. (2011). *Actinidia* is a natural host to a wide range of plant viruses. *Acta Hort.* 913, 467–472. doi: 10.17660/ActaHortic.2011.913.62

Peng, C., Peremyslov, V. V., Mushegian, A. R., Dawson, W. O., and Dolja, V. V. (2001). Functional specialization and evolution of leader proteinases in the family *Closteroviridae*. *J. Virol.* 75, 12153–12160. doi: 10.1128/JVI.75.24.12153-12160.2001

Peng, Q., Lv, R., Ning, J., Yang, T., Lin, H., Xi, D., et al. (2018). First report of Actinidia virus 1 infecting *Actinidia chinensis* in China. *Plant Dis.* 103, 782–782. doi: 10.1094/PDIS-08-18-1334-PDN

Qu, L., Cui, H., Wu, G., Zhou, J., Su, J., Wang, G., et al. (2014). Genetic diversity and molecular evolution of Plum bark necrosis stem pitting-associated virus from China. *PLoS One* 9:e105443. doi: 10.1371/journal.pone.0105443

Roossinck, M. J. (2011). The big unknown: plant virus biodiversity. *Curr. Opin. Virol.* 1, 63–67. doi: 10.1016/j.coviro.2011.05.022

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

Roy, A., and Brlansky, R. (2009). Population dynamics of a Florida citrus tristeza virus isolate and aphid-transmitted subisolates: identification of three genotypic groups and recombinants after aphid transmission. *Phytopathology* 99, 1297–1306. doi: 10.1094/PHYTO-99-11-1297

Roy, A., Manjunath, K., and Brlansky, R. (2005). Assessment of sequence diversity in the 5′ terminal region of citrus tristeza virus from India. *Virus Res.* 113, 132–142. doi: 10.1016/j.virusres.2005.04.023

Rubio, L., Ayllo, M. A., Kong, P., Fernandez, A., Polek, M., Guerri, J., et al. (2001). Genetic variation of citrus tristeza virus isolates from California and Spain: evidence for mixed infections and recombination. *J. Virol.* 75, 8054–8062. doi: 10.1128/JVI.75.17.8054-8062.2001

Rubio, L., Guerri, J., and Moreno, P. (2013). Genetic variability and evolutionary dynamics of viruses of the family *Closteroviridae*. *Front. Microbiol.* 4:151. doi: 10.3389/fmicb.2013.00151

Ruiz, L., Simon, A., Garcıa, C., Velasco, L., and Janssen, D. (2018). First natural crossover recombination between two distinct species of the family *Closteroviridae* leads to the emergence of a new disease. *PLoS One* 13:e0198228. doi: 10.1371/journal.pone.0198228

Sztuba-Solińska, 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

Tatineni, S., Gowda, S., Ayllon, M. A., Albiach-Marti, M. R., and Dawson, W. O. (2002). Mutational analysis of the replication signals in the 3′-nontranslated region of citrus tristeza virus. *Virology* 300, 140–152. doi: 10.1006/viro.2002.1550

Thompson, J. R., Fuchs, M., and Perry, K. L. (2012). Genomic analysis of grapevine leafroll-associated virus 5 and related viruses. *Virus Res.* 163, 19–27. doi: 10.1016/j.virusres.2011.08.006

Turturo, C., Saldarelli, P., Dong, Y., Digiaro, M., Minafra, A., Savino, V., et al. (2005). Genetic variability and population structure of grapevine leafroll-associated virus 3 isolates. *J. Gen. Virol.* 86, 217–224. doi: 10.1099/vir.0.80395-0

Veerakone, S., Liefting, L. W., Tang, J., and Ward, L. I. (2018). The complete nucleotide sequence and genome organisation of a novel member of the family *Betaflexiviridae* from *Actinidia chinensis*. *Arch. Virol.* 163, 1367–1370. doi: 10.1007/s00705-017-3701-x

Villamor, D. E. V., Ho, T., Rwahnih, M. A., Martin, R. R., and Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109, 716–725. doi: 10.1094/PHYTO-07-18-0257-RVW

Wang, D., Liu, X., Li, T., Li, Q., Liu, Y., Gong, G., et al. (2018). First report of cucumber mosaic virus infection in kiwifruit (*Actinidia chinensis*) in China. *Plant Dis.* 102, 1180–1180. doi: 10.1094/PDIS-06-17-0824-PDN

Wang, Y., Wang, G., Yang, Z., Wang, L., Li, L., and Hong, N. (2016). First report of the tospovirus tomato necrotic spot associated virus infecting kiwifruit (*Actinidia* sp.). *Plant Dis.* 100, 2539–2539. doi: 10.1094/PDIS-05-16-0629-PDN

Wang, Y., Zhai, L., Wen, S., Yang, Z., Wang, G., and Hong, N. (2020). Molecular characterization of a novel emaravrius infecting *Actinidia* spp. in China. *Virus Res.* 275:197736. doi: 10.1016/j.virusres.2019.197736

Wang, Y., Zhuang, H., Yang, Z., Wen, L., Wang, G., and Hong, N. (2018). Molecular characterization of an apple stem grooving virus isolate from kiwifruit (*Actinidia chinensis*) in China. *Can. J. Plant Pathol.* 40, 76–83. doi: 10.1094/PDIS-06-17-0824-PDN

Wen, S., Zhu, C., Yang, Z., Wang, Y., Wang, G., Wang, L., et al. (2018). First report of Actinidia virus 1 infecting kiwifruit in China. *Plant Dis.* 103:780. doi: 10.1094/PDIS-07-18-1123-PDN

Wu, G., Tang, M., Wang, G., Jin, F., Yang, Z., Cheng, L., et al. (2015). Genetic diversity and evolution of two capsid protein genes of citrus tristeza virus isolates from China. *Arch. Virol.* 160, 787–794. doi: 10.1007/s00705-014-2281-2

Wu, Q., Ding, S., Zhang, Y., and Zhu, S. (2015). Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. *Ann. Rev. Phytopathol.* 53, 425–444. doi: 10.1146/annurev-phyto-080614-120030

Zhang, S., Yang, L., Ma, L., Tian, X., Li, R., Zhou, C., et al. (2020). Virome of *Camellia japonica*: discovery and molecular characterization of new viruses of different taxa in *Camellias*. *Front. Microbiol.* 11:945. doi: 10.3389/fmicb.2020.00945

Zhao, L., Cao, M., Huang, Q., Jing, M., Bao, W., Zhang, Y., et al. (2020). Occurrence and molecular characterization of Actinidia virus (AcVC), a novel vitivirus infecting kiwifruit (*Actinidia* spp.) in China. *Plant Pathol.* 69, 775–782. doi: 10.1111/ppa.13171

Zhao, L., Yang, W., Zhang, Y., Wu, Z., Wang, Q., and Wu, Y. (2019). Occurrence and molecular variability of kiwifruit viruses in *Actinidia deliciosa* " Xuxiang " in the Shaanxi province of China. *Plant Dis.* 103, 1309–1318. doi: 10.1094/PDIS-09-18-1570-RE

Zheng, Y., Navarro, B., Wang, G., Wang, Y., Yang, Z., Xu, W., et al. (2017). Actinidia chlorotic ringspot-associated virus: a novel emaravirus infecting kiwifruit plants. *Mol. Plant Pathol.* 18, 569–581. doi: 10.1111/mpp.12421

Zheng, Y., Wang, G., Zhou, J., Zhu, C., Wang, L., Xu, W., et al. (2014). First report of Actinidia virus A and Actinidia virus B on kiwifruit in China. *Plant Dis.* 98, 1590–1590. doi: 10.1094/PDIS-04-14-0420-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.

Copyright © 2020 Wen, Wang, Yang, Wang, Rao, Lu and Hong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 17 December 2020
doi: 10.3389/fmicb.2020.610872

Efficient, Rapid, and Sensitive Detection of Plant RNA Viruses With One-Pot RT-RPA–CRISPR/Cas12a Assay

*Rashid Aman, Ahmed Mahas, Tin Marsic, Norhan Hassan and Magdy M. Mahfouz**

Laboratory for Genome Engineering and Synthetic Biology, Division of Biological Sciences, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Most viruses that infect plants use RNA to carry their genomic information; timely and robust detection methods are crucial for efficient control of these diverse pathogens. The RNA viruses, potexvirus (*Potexvirus*, family *Alphaflexiviridae*), potyvirus (*Potyvirus*, family *Potyviridae*), and tobamovirus (*Tobamovirus*, family *Virgaviridae*) are among the most economically damaging pathogenic plant viruses, as they are highly infectious and distributed worldwide. Their infection of crop plants, alone or together with other viruses, causes severe yield losses. Isothermal nucleic acid amplification methods, such as loop-mediated isothermal amplification (LAMP), recombinase polymerase amplification (RPA), and others have been harnessed for the detection of DNA- and RNA-based viruses. However, they have a high rate of non-specific amplification and other drawbacks. The collateral activities of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated nuclease Cas systems such as Cas12 and Cas14 (which act on ssDNA) and Cas13 (which acts on ssRNA) have recently been exploited to develop highly sensitive, specific, and rapid detection platforms. Here, we report the development of a simple, rapid, and efficient RT- RPA method, coupled with a CRISPR/Cas12a-based one-step detection assay, to detect plant RNA viruses. This diagnostic method can be performed at a single temperature in less than 30 min and integrated with an inexpensive commercially available fluorescence visualizer to facilitate rapid, in-field diagnosis of plant RNA viruses. Our developed assay provides an efficient and robust detection platform to accelerate plant pathogen detection and fast-track containment strategies.

Keywords: RT-RPA, CRISPR-Cas12a, plant virus RNA, biosensors, diagnostics

OPEN ACCESS

Edited by:
Ahmed Hadidi,
United States Department of Agriculture (USDA), United States

Reviewed by:
José-Antonio Daròs,
Polytechnic University of Valencia, Spain
Jin Wang,
Chinese Academy of Sciences (CAS), China

***Correspondence:**
Magdy M. Mahfouz
magdy.mahfouz@kaust.edu.sa

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

Received: *27 September 2020*
Accepted: *30 November 2020*
Published: *17 December 2020*

Citation:
Aman R, Mahas A, Marsic T, Hassan N and Mahfouz MM (2020) Efficient, Rapid, and Sensitive Detection of Plant RNA Viruses With One-Pot RT-RPA–CRISPR/Cas12a Assay. Front. Microbiol. 11:610872. doi: 10.3389/fmicb.2020.610872

INTRODUCTION

Despite the significant increase of ssDNA viruses in recent years, RNA viruses are still the most prevalent disease-causing agents in plants and pose a major threat to agriculture and food security. Most plant RNA viruses have a plus single-stranded RNA (+ ssRNA) genome that encodes the proteins required for replication, translation, encapsidation, cell to cell movement, and others (Joshi and Haenni, 1984). Their genomes take on the role of messenger RNA (mRNA) in host cells and are

translated into functional proteins that are involved, directly or indirectly, in virus replication. Plant RNA viruses with a negative double-stranded RNA (-dsRNA) or negative single-stranded RNA (-ssRNA) genome are first transcribed and the resulting RNAs are then translated into functional viral proteins (Joshi and Haenni, 1984). The + ssRNA-based viruses potato virus X (PVX), potato virus Y (PVY), and tobacco mosaic virus (TMV) are among the most economically important viruses based on their worldwide distribution and effect on crop yield and quality (Scholthof et al., 2011; Rybicki, 2015). PVX predominantly spreads via mechanical transmission and causes 10–15% yield loss in potato (*Solanum tuberosum*) (Kreuze et al., 2020). Its symptoms vary from weak to intense, resulting in smaller leaves and stunted plants. PVY is transmitted by over 25 different aphid species and can also be mechanically transmitted, and poses a significant threat to tuber yield and quality as well as seed production of potato worldwide (Lawson et al., 1990; Kreuze et al., 2020). TMV, the iconic first recognized virus, also spreads via mechanical transmission and is of major economic importance, as TMV infections cause mottled browning of tobacco (*Nicotiana tabacum*) leaves (Hadidi et al., 2020). TMV also infects other Solanaceous crops plants, particularly tomato (*Solanum lycopersicum*) (Rifkind and Freeman, 2005; Kumar et al., 2011).

The development of a robust and sensitive diagnostic method is of utmost importance to control plant viruses. Several methods have been developed, including, but not limited to, enzyme-linked immunosorbent assay (ELISA), molecular hybridization, polymerase chain reaction (PCR), reverse transcription followed by PCR (RT-PCR), loop (or RT-loop) mediated isothermal amplifications (LAMP and RT-LAMP), and microarrays (Baranwal et al., 2020; Hadidi et al., 2020; Rubio et al., 2020). ELISA depends on binding viral-specific antibodies to the virus coat protein (Clark and Adams, 1977). Molecular hybridization techniques like Southern and northern blotting use sequence-specific probes to identify DNA or RNA molecules, respectively (Hadidi et al., 2020), that can be later visualized by different methods, including simple chemiluminescent detection. PCR, RT-PCR, RT-LAMP, and LAMP, which do not require cycling of temperature, in contrast to PCR, have further streamlined plant pathogens detection (Faggioli et al., 2017). Real-time quantitative PCR (qPCR) and (RT-qPCR) rely on the binding of a virus sequence-specific fluorescent probe to the PCR-amplified region and are conventional techniques used for the detection of DNA and RNA viruses (Lin et al., 2013; Faggioli et al., 2017; Jarošová et al., 2018). Although widely used, qPCR and RT-qPCR detection methods require well-equipped laboratories and trained personnel, impeding the use of these technologies at point-of-care or resource-limited areas.

By contrast, isothermal amplification techniques such as helicase-dependent amplification (HDA), nucleic acid sequence-based amplification (NASBA), LAMP, and recombinase polymerase amplification (RPA) allow nucleic acid amplification at a single temperature, thus supporting their use in the field for on-site diagnosis in a low-resource environment (Compton, 1991; Notomi et al., 2000; Vincent et al., 2004; Piepenburg et al., 2006; Rubio et al., 2020). LAMP and RPA DNA amplification approaches have been further customized as RT-LAMP and RT-RPA when combined with a suitable reverse transcriptase for pathogens harboring an RNA genome. RT-LAMP and RT-RPA have been utilized for the detection of many plant viruses, including PVX and PVY (Fukuta et al., 2003; Nie, 2005; Euler et al., 2013; Uehara-Ichiki et al., 2013; Budziszewska et al., 2016; Faggioli et al., 2017; Silva et al., 2018). Although these isothermal amplification methods are highly sensitive, they also demand additional care to minimize non-specific amplification that often results in false-positive results.

Clustered regularly short palindromic repeat-associated systems (CRISPR-Cas) are derived from a natural immune-like system that provides molecular immunity to microorganisms (bacterial and archaeal species) against foreign invading species. CRISPR-Cas has been extensively exploited in eukaryotic species for genome engineering, molecular immunity, and transcriptome regulation (Ran et al., 2013; Hsu et al., 2014; Ali et al., 2015; Aman et al., 2018; Mahas et al., 2019). In several CRISPR systems target identification and cleavage results in the activation of their non-specific endonuclease activity (Aman et al., 2020), which has been harnessed for nucleic acid detection. For instance, the well-known nucleic acid detection platforms SHERLOCK (Specific High-sensitivity Enzymatic Reporter UnLOCKing) and DETECTR (DNA Endonuclease-Targeted CRISPR Trans Reporter) employ the collateral activities of Cas13 and Cas12, respectively, when using ssRNA or ssDNA reporters (Chen et al., 2018; Gootenberg et al., 2018; Li et al., 2018a,b; Kellner et al., 2019). End-point or quantitative measurements can be obtained by using a variety of fluorophore- or biotin-labeled reporters.

LAMP and RPA methods have been successfully integrated with CRISPR/Cas technology for rapid and portable detection of viral nucleic acids (Kellner et al., 2019; Broughton et al., 2020). We recently developed the ***in vitro*** **S**pecific **C**RISPR-based **A**ssay for **N**ucleic acids detection (iSCAN) for sensitive and rapid detection of SARS-CoV-2, the causal agent of COVID-19, which combines the target sequence amplification via RT-LAMP with the specific detection and subsequent collateral activity of CRISPR/Cas12 (Ali et al., 2020). To mitigate the non-specific amplification and cross-contamination issues inherent in these amplification methods, one-pot assays have been recently developed for virus detection (Joung et al., 2020). In this study, we have applied iSCAN, after replacing RT-LAMP with RT-RPA, to develop a one-pot detection assay named as iSCAN-one-pot (iSCAN-OP), for the detection of plant RNA viruses. We report that our developed iSCAN-OP is simple, specific, rapid, and sensitive for the detection of plant RNA viruses. The RT-RPA pre-amplification step converts the RNA genome of the virus into dsDNA that serves as a substrate for the Cas12a cis activity. Cas12a targeting of the dsDNA triggers its collateral activity, which in turn cleaves the ssDNA reporter molecules and releasing the signal (Chen et al., 2018; Ding et al., 2020). We further combined the iSCAN-OP detection assay with a commercially available P51 fluorescence viewer device to facilitate quick, affordable, in-field diagnosis of plant RNA viruses. Our iSCAN-OP detection assay can be easily adapted for large-scale virus screenings in the field and will constitute

a critical tool in agriculture biotechnology for rapid diagnostics of plant diseases.

MATERIALS AND METHODS

Nucleic Acid Preparation

For *in vitro*-transcribed viral RNA, the coding sequence for the coat protein (CP) of each virus was amplified via PCR with the appropriate virus-specific oligonucleotides (**Supplementary Table S1**). PCR products were then purified using a QIAquick PCR purification kit as per the manufacturer's protocol. A total of 1–1.5 μg of each PCR product was *in vitro*-transcribed at 37°C overnight (8 h) with the TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific K0441). *In vitro*-transcribed viral RNA was then purified using the DIRECTZOL KIT from ZymoBiomics (Direct-Zol RNA Miniprep kit; catalog #R2070) following the manufacturer's protocol. For real viral samples, 3-week-old *Nicotiana benthamiana* plants were infected with the virus, as previously described (Aman et al., 2018). Total plant RNA was extracted from systemic leaves 7–10 days after infiltration (dai) using the DIRECTZOL KIT from ZymoBiomics (Direct-Zol RNA Miniprep kit; catalog #R2070). The concentrations of *in vitro* transcribed RNA and total plant RNA templates used in the RT-RPA reactions were adjusted to 100 ng/μL with sterile water.

All CRISPR RNAs (crRNAs) were produced by *in vitro* transcription of commercially synthesized ssDNA oligos with an appended T7 promoter sequence (Sigma), as previously reported (Abudayyeh et al., 2019; **Supplementary Table S2**). The synthetized ssDNAs were annealed to a forward-directed T7 promoter primer (Sigma) to initiate *in vitro* transcription of crRNAs at 37°C overnight (8 h). All crRNAs were purified using the DIRECTZOL KIT from ZymoBiomics.

Cas12 Protein Purification and Assessment of Activity

LbCas12a protein from *Lachnospiraceae bacterium* (LbCas12a) was purified as previously described by Chen et al. (2018) and Ali et al. (2020) protocol. The target-specific endonuclease activity of Cas12a was tested and validated against a PCR product amplified from the PVX, PVY, and TMV CP coding sequences, as previously described (Ali et al., 2020). The collateral *trans* cleavage activity of LbCas12a was measured by providing the Hex-labeled ssDNA reporter (/5HEX/TTTTTTT/3IABkFQ/) in the restriction reaction. The fluorescence signal as a result of trans-cleavage activity of Cas12a was measured with the P51 Molecular Fluorescence Viewe[1].

Design and Screening of RPA Primers

For each virus template (TMV NCBI ID: MK087763.1, PVX NCBI ID: M95516.1 and PVY NCBI ID: HM367076.1), 3–4 sets of RPA primers (**Supplementary Table S3**) were designed as per the manufacturer's instructions to amplify a region of 100–200 bp. The length and melting temperatures of the primers varied from 30 to 35 bases and 54–67°C, respectively. All primers used in RPA assays were from Sigma. The efficiency of different sets of RPA primers was tested with *in vitro*-transcribed viral RNA, and the signal was examined using a Tecan plate reader (M200) or P51 Molecular Fluorescence Viewer.

iSCAN-Two-Pot (iSCAN-TP) Detection Assay

Total RNA extracted from infected plants was subjected to iSCAN-TP or iSCAN one-pot (iSCAN-OP) before being handed over to CRISPR/Cas-based detection in a two- or one-pot reaction, respectively. The fluorescent signal resulting from Cas12a-based collateral cleavage of the single-stranded DNA-fluorescent quencher (ssDNA-FQ) reporter was visualized via a fluorescence viewer (**Figure 1A**).

The iSCAN-TP detection assays were performed in two steps. The RT-RPA assay was carried out according to the manufacturer's protocol (TwistAmp^{TM} Basic Kit "Improved Formulation," cat. no. TABAS03KIT) with slight modifications. Briefly, first resuspending the RPA pellet with 29.5 μL resuspension buffer, then 11.8 μL of the resuspended RPA mixture was mixed with 0.5 μL Superscript III Reverse Transcriptase, 0.5 μL forward primer (10 μM), 0.5 μL reverse primer (10 μM), 1 μL *in vitro* transcribed or total plant RNA (100 ng) and 0.2 μL nuclease-free water. Reactions tubes were mixed thoroughly by pipetting, before adding 2 μL of 280 nM magnesium acetate. The tubes were incubated at 42°C to start the RT-RPA reaction. Next, for Cas12 based detection assays, 250 nM of LbCas12a protein was first pre-incubated with 250 nM of LbCas12a crRNAs in 1x Cas12 reaction Buffer (20 mM HEPES (pH 7.5), 150 mM KCl, 10 mM MgCl2, 1% glycerol and 0.5 mM DTT) for 30 min at 37°C to assemble Cas12-crRNA ribonucleoprotein (RNP) complexes as previously described (Chen et al., 2018). The reaction was diluted 4 times with 1x binding buffer (20 mM Tris-HCl, pH 7.5, 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 5% glycerol, 50 μg mL-1 heparin) (Chen et al., 2018). Two microliter aliquot of the RT-RPA products was added to a pre-assembled and diluted Cas12a/crRNA ribonucleoprotein (RNP) complex with the addition of the Hex-labeled ssDNA reporter for fluorescence-based detection. The reaction was then incubated for another 15–30 min at 37°C and the fluorescent signal was monitored using either a Tecan plate reader (M200) or a P51 Molecular Fluorescence Viewer.

iSCAN-OP Detection Assay

The difference between iSCAN one-pot and two-pot assays is that, in two-pot, the target sequence is amplified first by isothermal amplification methods such as RPA in a separate tube. The amplified product is then added into another separate tube that has the detection reagents such as Cas12a, crRNA and the reporter. In case of one-pot assays, both steps (target sequence amplification and Cas12a-based detection) are performed in the same tube at the same temperature (Kellner et al., 2019).

The iSCAN-OP detection assays were performed by first resuspending the RPA pellet with 29.5 μL resuspension buffer,

[1]https://www.minipcr.com/product/p51-molecular-glow-lab/

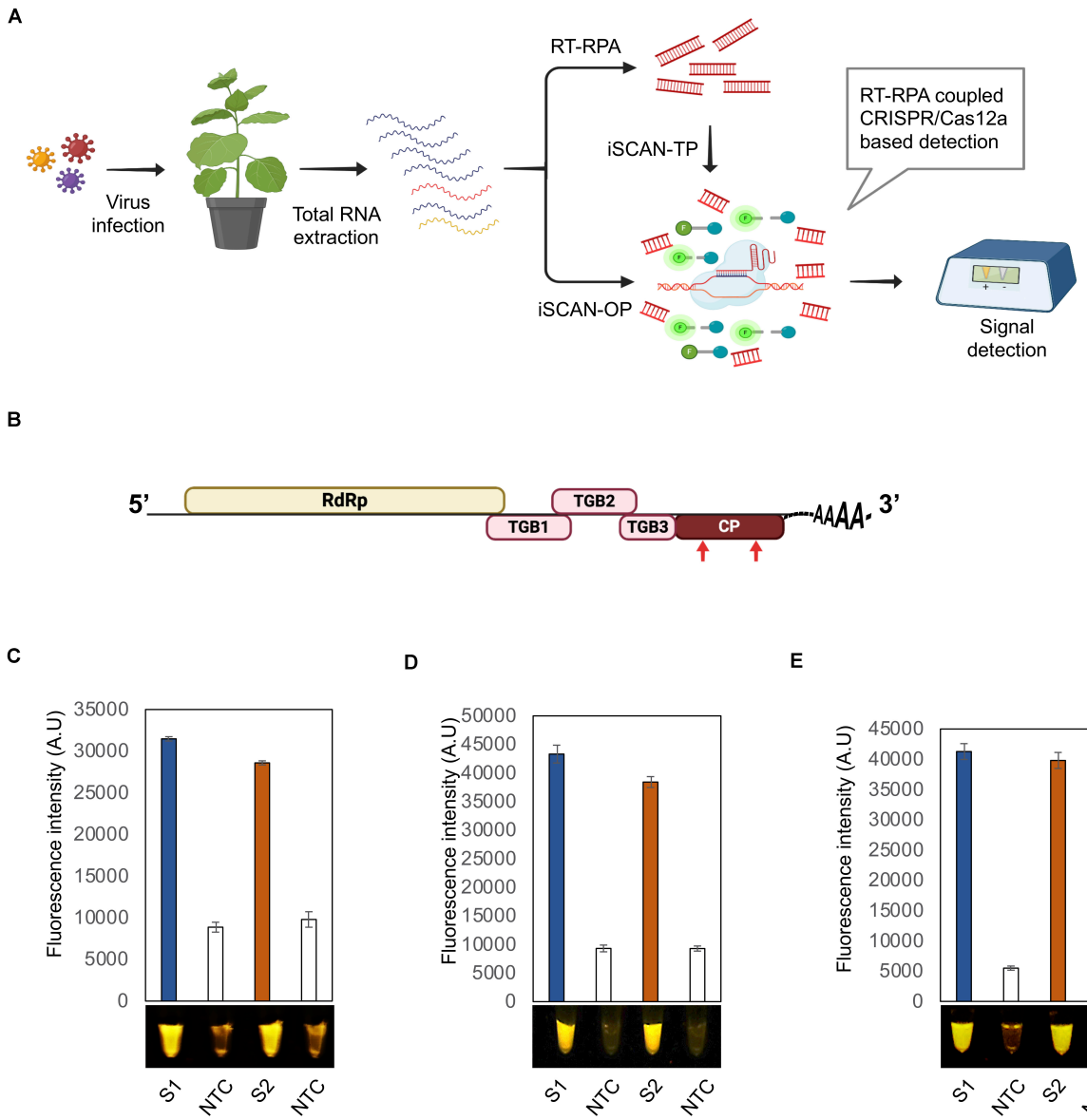


FIGURE 1 | Optimization of the iSCAN-OP detection assay for PVX and PVY detection**. (A)** Workflow of RT-RPA coupled CRISPR/Cas12a based detection system. **(B)** A schematic view of the potato virus X (PVX) genome. Red arrows indicate the regions targeted by the RT-RPA coupled CRISPR/Cas12a based detection system. **(C)** End-point fluorescence visualization of the iSCAN-TP detection assay (*in vitro* transcribed RNA) with P51 Molecular Fluorescence Viewer. S1: Primer set 1, S2: Primer set 2, NTC: no template control. Values are shown in the graph as means ± SD (n = 3). **(D)** End-point fluorescence visualization of the iSCAN-OP detection assay (*in vitro* transcribed RNA) with P51 Molecular Fluorescence Viewer. S1: Primer set 1, S2: Primer set 2, NTC: no template control. Values are shown in the graph as means ± SD (n = 3). **(E)** End-point fluorescence visualization of the iSCAN-OP detection assay (*in vitro* transcribed RNA) with P51 Molecular Fluorescence Viewer. S1: Primer set 1, S2: Primer set 2, NTC: no template control. Values are shown in the graph as means ± SD (n = 3).

followed by the addition of 1 μL Superscript III Reverse Transcriptase (Invitrogen), 1.5 μL forward primer (10 μM), 1.5 μL reverse primer (10 μM), 2 μL Cas12a (5 μM), 2 μL crRNA (5 μM), 2.5 μL nuclease-free water, and 4 μL Hex reporter (10 μM). Reaction tubes were mixed by pipetting, and the mixture was divided equally 22 μL into two tubes. 1 μL of synthetic or total plant RNA (100 ng) and 2 μL of 280 mM magnesium acetate was then added to the reaction tube and incubated at 42°C for 20 min. The end-point fluorescence signal was monitored using either a Tecan plate reader (M200) or a P51 Molecular Fluorescence Viewer.

Real-Time Detection of *in vitro* Transcribed TMV RNA

To determine the sensitivity of RT-RPA-CRISPR/Cas12a detection method, we diluted *in vitro* transcribed TMV RNA down to femtomolar concentrations and set up one-pot RT-RPA

as described above, with the addition of 250 nM FAM reporter (5′-/56-FAM/TTATT/3IABKFQ/-3′, IDT) instead of HEX reporter. To detect the FAM signal in real-time, the reaction samples were run in an Applied Biosystems StepOne Real-time PCR system for 1 h at 42°C with fluorescent measurements every 2 min. As a no-template control, we added water instead of TMV synthetic RNA.

RESULTS

Optimization of the iSCAN-OP Detection Assay for PVX and PVY Detection

To test the activity of our system, we selected two economically significant plant RNA viruses: PVX and PVY. We designed multiple RPA primer sets against the coding sequence of the viral CP (**Figure 1B**, **Supplementary Figure 2A**, and **Supplementary Table S1**). We also designed two independent crRNAs, and assessed their CRISPR/Cas12a-based *cis*-cleavage activity against the PCR products amplified from the CP coding sequences. Both crRNAs very efficiently cleaved their intended targets (**Supplementary Figure 1A**). To confirm the *trans* collateral cleavage activity of CRISPR/Cas12a, we added a single-stranded HEX-labeled DNA reporter to the reaction. In contrast to control samples containing no crRNA, we observed a strong fluorescent signal in the presence of CRISPR/Cas12a. In addition, Cas12a cleaved the ssDNA reporter only when pre-loaded with crRNA (**Supplementary Figure 1B**).

To test the efficiency of the designed primers, we subjected the PCR products amplified from the CP to *in vitro* transcription, followed by purification and applied to iSCAN-TP detection assay. All primer sets were effective in the iSCAN-TP reaction format, although we noticed that the first two sets were more active (**Supplementary Figure 1C**). We thus proceeded with the first two primer sets in later PVX RT-RPA experiments (**Figure 1C**).

To overcome the contamination of the work space problem which may lead to false negative assays, we developed an iSCAN-OP detection assay for plant RNA viruses (RT-RPA-CRISPR/Cas12a). Since the optimal temperature for reverse transcriptase enzymes is typically 42°C or more, we tested whether the Cas12a *cis* and *trans* endonuclease activities might also be active at temperatures other than 37°C. Cas12a exhibited cis and trans activities at all temperatures from 37 to 42°C, that is in line with the previously reported studies (Huang et al., 2020), paving the way for its use for iSCAN-OP detection assay, which we tested next on *in vitro* transcribed PVX-CP. Our detection assay efficiently detected the PVX-CP RNA in a reaction mixture compared to the no template control (NTC) (**Figure 1D**). To validate our detection system with freshly prepared viral samples, we isolated total RNA from *Nicotiana benthamiana* plants infected with PVX and repeated the iSCAN-OP detection assay. We determined that our detection system also effectively detected viral RNA in total RNA extracted from infected plants (**Figure 1E**).

Next, we assessed our RT-RPA-CRISPR/Cas12a detection assay against another plant RNA virus, PVY. To this end, we designed two crRNAs targeting the CP coding region of the PVY genome (**Supplementary Table S3**) and evaluated their activity against a PCR product covering the PVY-CP sequence. The CRISPR/Cas12a-based *cis* cleavage of the PVY-CP PCR product indicated that both crRNAs efficiently directed Cas12a for target cleavage (**Supplementary Figure 2B**). The addition of the HEX reporter also succeeded in identifying Cas12a target-based collateral activity in the reaction mixture (**Supplementary Figure 2C**). We measured a strong signal in the reaction mixture compared to NTC, indicating that both crRNAs activated Cas12a *cis* and *trans* cleavage activities. Subsequently, we designed four sets of RPA primers to assess our RT-RPA-CRISPR/Cas12a system against *in vitro*-transcribed PVY-CP RNA. We detected a strong signal in all reactions relative to NTC, with the exception of primer set 4 (S4), where we observed non-specific amplification in the control reaction (**Supplementary Figure 2D**). This result underscores the need to screen multiple primers sets in isothermal reactions.

Detection of TMV With the iSCAN-OP Detection Assay

A robust and useful detection assay should be effective in detecting various viruses in addition to PVX and PVY. We therefore selected another plant pathogenic virus with a + ssRNA genome: tobacco mosaic virus (TMV). As described above for PVX and PVY, we designed two crRNAs against the CP coding region of TMV and tested their Cas12a-based *cis* and *trans* cleavage potential (**Figure 2A**). As shown in **Supplementary Figure 3A**, a TMV-CP PCR product was efficiently cleaved in a Cas12a/crRNA-dependent manner compared to the control reaction. Both crRNAs also directed Cas12a target-based collateral activity, as evidence by the associated fluorescence signal relative to the control (**Supplementary Data 3B**).

To generalize the applicability of our detection system, we designed three different RPA primer sets to amplify a region of the TMV-CP and tested their efficacy in an iSCAN-OP detection assay with *in vitro*-transcribed TMV-CP RNA. All primers sets were equally efficient and able to detect TMV-CP-RNA in a iSCAN-OP reaction (**Figure 2B**). Next, to test the efficiency of these primers on freshly collected samples, we isolated total RNA from *N. benthamiana* plants infected with TMV and subjected them to our detection assay. Again, all primer sets efficiently detected the TMV viral genome when the fluorescent signal was compared to the control reaction (**Figure 2C**). Collectively, our results suggest that this detection assay can be used and generalized for any plant RNA virus. Such a robust and fast platform will undoubtedly advance the timely detection of any pathogenic plant virus, thus allowing early diagnosis and the implementation of preventive strategies.

Multiplexing and Sensitivity of the iSCAN-OP Detection Assay

To determine how our iSCAN-OP detection assay performs when plants are infected by more than one virus, we infected 3-week-old *N. benthamiana* plants with PVX and

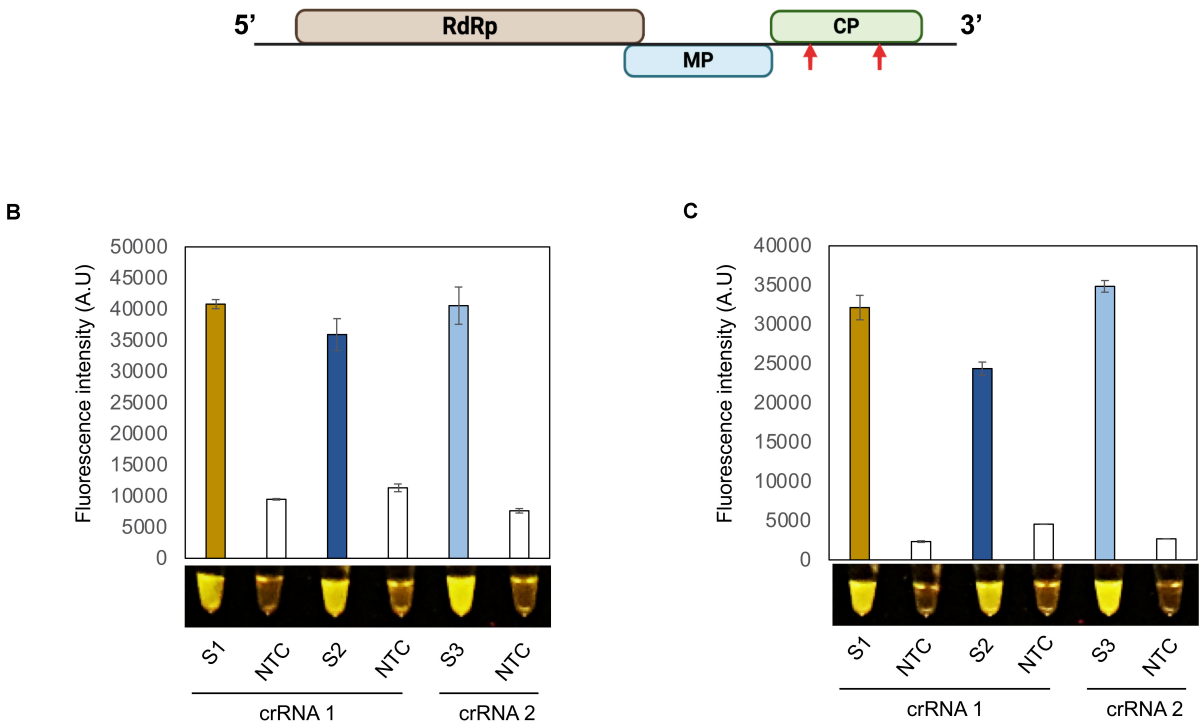


FIGURE 2 | Detection of TMV with the iSCAN-OP detection assay.**(A)** A schematic view of tobacco mosaic virus (TMV) genome design. Red arrows indicate the regions targeted by RT-RPA coupled CRISPR/Cas12a based detection system. **(B)** End-point fluorescence visualization of the iSCAN-OP detection assay (*in vitro* transcribed RNA) with P51 Molecular Fluorescence Viewer. S1: Primer set 1, S2: Primer set 2, S3: primer set 3, NTC: no template control. Values are shown in the graph as means ± SD (n = 3). **(C)** End-point fluorescence visualization of the iSCAN-OP detection assay (with plant total RNA) with P51 Molecular Fluorescence Viewer. S1: Primer set 1, S2: Primer set 2, S3: Primer set 3 NTC: no template control. Values are shown in the graph as means ± SD (n = 3).

TMV, separately or together. We extracted total RNA from plants 7 dai and performed iSCAN-OP detection assays at 42°C for 20 min. We only detected PVX in mixed infection samples when using PVX-specific primers and the appropriate crRNA. We obtained no signal amplification from plants that had been infected only with TMV when using the PVX-specific primers and crRNA, indicating that the observed signal originated from PVX viral RNA exclusively (**Figure 3A**). Similarly, TMV-specific RPA primers and crRNA only detected TMV viral RNA but not PVX, demonstrating that our detection assay is specific, robust and can be utilized for multiplexed detection of different viruses in situations of mixed infections.

Last, we investigated the detection sensitivity of our RT-RPA-CRISPR/Cas12a detection assay in real-time. To this end, we diluted our *in vitro*-transcribed TMV RNA down to femtomolar concentrations and subjected the serial dilutions to our iSCAN-OP detection assay with the S3 set of RPA primers together with a ssDNA FAM fluorescent reporter to measure fluorescence in real-time. Our results indicate that our CRISPR-based detection assay can detect viral RNA down to the picomolar range, however, longer reaction time can improve the detection limit of the assay down to femtomolar, which is in line with previously reported data (**Figure 3B**; Cha et al., 2020). However, we noticed that a longer incubation period was required for the effective detection of viral RNA molecules at low concentrations.

DISCUSSION

Despite various control strategies to prevent viral spread in crop fields, RNA viruses remain a threat to plant health and crop productivity worldwide. Specific and robust detection of plant viruses is crucial for early detection and containment. As the majority of viruses that infect plants harbor RNA genomes, we established a robust and efficient iSCAN-OP detection method combining the RT-RPA and CRISPR/Cas12a-based detection platforms for plant RNA viruses. The development of robust, simple, and affordable detection platforms thus has great potential to lower crop losses. To address this, we combined two methods into such a detection assay: first, RPA amplifies the target sequence at one temperature (isothermal reaction); second, CRISPR/Cas12a collateral nuclease activity is activated in the presence of its target, which was just amplified by RPA. RPA is more convenient and economical than LAMP, as it entails a single pair of primers in comparison to 4–6 primers needed for LAMP. Another advantage of RPA over LAMP and

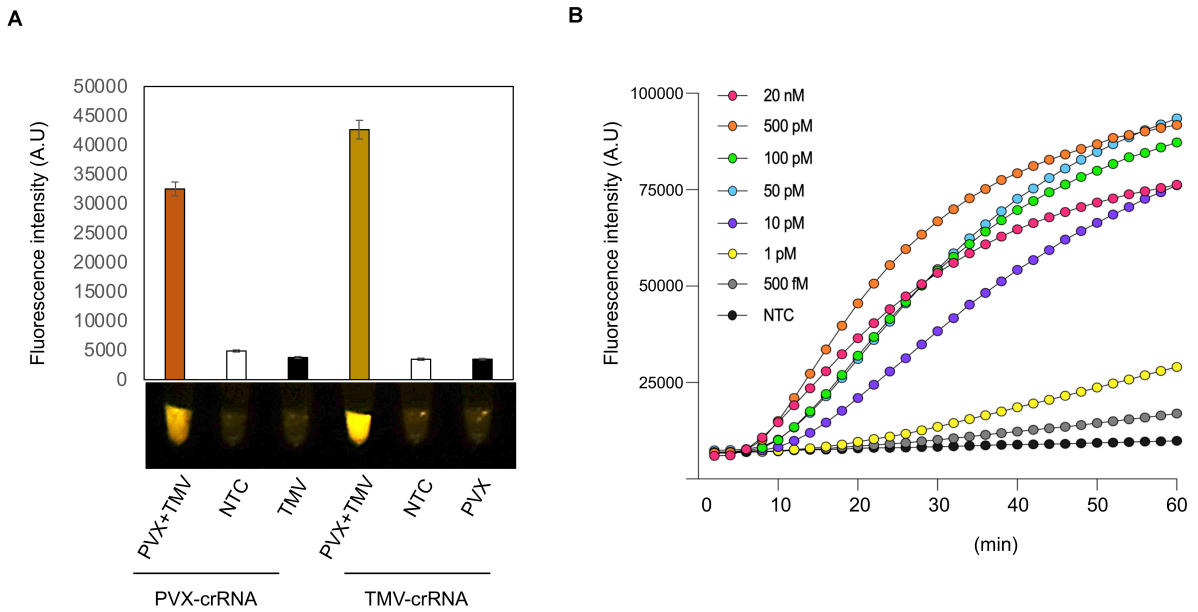


FIGURE 3 | Multiplexing and sensitivity of the iSCAN-OP detection assay. **(A)** End-point fluorescence visualization of the iSCAN-OP detection assay (with plant total RNA) with P51 Molecular Fluorescence Viewer. Values are shown in the graph as means ± SD (n = 3). **(B)** Real-time detection of Cas12a trans-cleavage activity using a ssDNA-FQ FAM labeled reporter. Values are shown in the graph as means (n = 3). NTC: no template control.

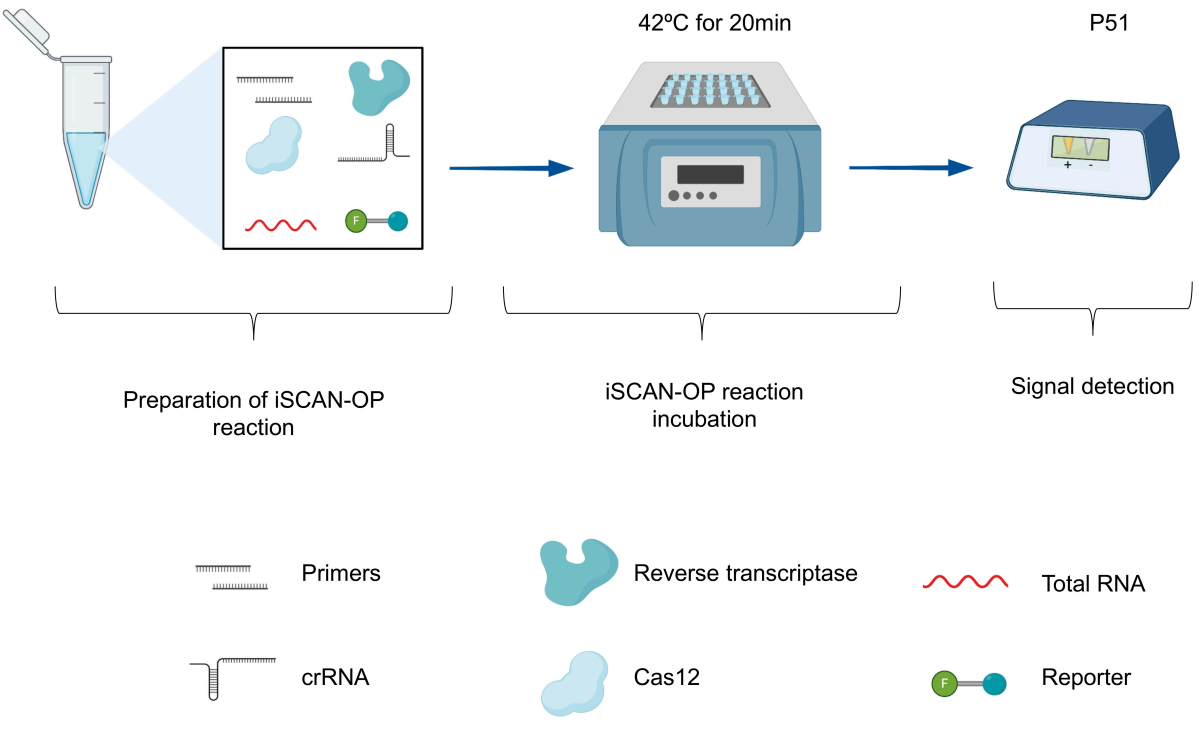


FIGURE 4 | Overview of the iSCAN-OP detection assay. Schematic illustration of the iSCAN-OP detection assay. The diagram describes the addition of all necessary reagents: virus-specific RPA primers, reverse transcriptase, Cas12a protein, crRNA and HEX reporter to the tubes containing the RPA substrate, resuspended from the lyophilized pellet in resuspension buffer. After mixing, total RNA extracted from plant tissue and magnesium acetate is added to the reaction mixture. The reaction mixture is incubated at an isothermal temperature of 42°C for 20 min and the fluorescence signal is measured with the P51 Molecular Fluorescence Viewer.

other isothermal amplification methods is its short reaction time (Zaghloul and El-Shahat, 2014).

The collateral activities of several Cas proteins (Cas13, Cas12, and Cas14) have been co-opted to add further specificity in CRISPR/Cas-based diagnostic platforms (Chaijarasphong et al., 2019; Aman et al., 2020; Joung et al., 2020; Kanitchinda et al., 2020). LAMP and RPA are highly robust nucleic acid amplification methods that amplify the target sequence in a much shorter time than traditional PCR. However, the robustness of these methods also constitutes a major limitation of isothermal amplification methods due to amplification of minute amounts of template (oftentimes resulting from cross-contamination), leading to false positives (Nagai et al., 2016; Ali et al., 2020). To overcome these issues, screening more than one pair of primers is recommended to select the best pair. It is also recommended to designate dedicated pre and post-amplification spaces to minimize the chances of cross-contamination (Panno et al., 2020). Isothermal amplification methods have recently been successfully coupled with CRISPR/Cas technology in single-pot reactions to reduce the potential for cross-contamination. The amplification temperature of RPA permits its pairing with CRISPR/Cas12a nuclease activity in a single-pot reaction.

CRISPR-based diagnostics have been used to diagnose different human pathogens. We have recently developed iSCAN assay for specific, sensitive and rapid detection of SARS-CoV-2 (Ali et al., 2020). Here, we report a highly robust and specific iSCAN-OP detection assay for plant RNA viruses; that can detect viral RNA in total RNA extracted from plants in 20 min. Different from the iSCAN assay for SARS-CoV-2, we chose to use RT-RPA instead of RT-LAMP for the development of one-pot assay in iSCAN-OP because RT-RPA and Cas12a detection can be performed in the same temperature.

During optimization of the iSCAN-OP detection assay, we tested the activity of CRISPR/Cas12a at different temperatures and determined that *cis* and *trans*-cleavage mediated by the Cas12a protein was similarly efficient at all tested temperatures, with a slightly higher activity at 42°C. The unequal efficiency of different RPA primers also indicates that all primer sets are not equivalent, pointing to the need to screen several primer sets. Furthermore, we confirmed the efficacy of our iSCAN-OP detection system against several plant RNA viruses, corroborating the efficiency of such systems, not only in plants but also in other species, since their detection principles are independent from the organism. Our assay may be applicable for viroid detection as it has been recently reported that CRISPR-Cas 13a, and 13b have a good potential for application as a rapid and accurate diagnostic assay for known viroids (Hadidi, 2019).

Our data also suggest that the additional specificity conferred by the CRISPR/Cas assay allows the detection of a specific virus during mixed infections when virus-specific RPA primers are used. We observed strong signals with detecting TMV from plants infected with both TMV and PVX, which we attribute to the high replicative nature of TMV in *N. benthamiana* (relative to PVX). We only detected a single virus in plants infected with TMV or PVX, indicating that the one-pot RT-RPA-CRISPR/Cas12a detection assay is highly specific, and can be clearly adopted for the detection of RNA-based viruses. In addition, the iSCAN-OP detection assay is highly sensitive as we detected TMV synthetic RNA molecules down to femtomolar concentrations. CRISPR/Cas12a can detect RT-RPA-amplified products in 20 min, although a longer incubation time of up to 1 h may be required in the case of low viral titer. Finally, to support in-field diagnosis, we coupled our detection assay with a commercially available and affordable fluorescent reader device (P51 Molecular Fluorescence Viewer). The reaction, incubated at the isothermal temperature of 42°C for 20 min, is transferred to a P51 fluorescent viewer for signal capture (**Figure 4**). The high specificity of our developed detection assay makes it amenable to be used in mixed infection of two or more viruses.

In conclusion, we have shown that a CRISPR/Cas12a-based detection assay coupled with nucleic acid amplification can detect plant RNA viruses with high specificity and sensitivity. The single-tube detection format contributes to minimizing potential cross-contamination and thus reduces false positives. Moreover, the single temperature requirement of the iSCAN-OP detection assay, together with the freedom from expensive equipment or highly trained personnel, makes this method amenable to be implemented at point-of-care, if coupled with an easy total RNA extraction protocol from plant tissues (Daher et al., 2016). As RPA has been reported to be tolerant to reaction inhibitors compared to other amplification methods such as qPCR, it will be more suitable when coupled with a quick total RNA extraction from plants, as reported for the detection of PVX from *N. benthamiana* (Shahin et al., 2018; Silva et al., 2018). Overall, our developed assay constitutes an efficient and robust detection method that has the potential to be used for RNA plant virus and viroid diagnosis.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MM conceived the research, analyzed the data, and managed the project. RA, AM, TM, and NH designed the research. RA and AM performed the research. RA, AM, and MM wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the King Abdullah University of Science and Technology (KAUST) Competitive Research Grants (CRG8) under Award No. CRG8-URF/1/4026-01-01.

ACKNOWLEDGMENTS

We would like to thank members of the Genome Engineering and Synthetic Biology Laboratory for insightful discussions and technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.610872/full#supplementary-material

REFERENCES

Abudayyeh, O. O., Gootenberg, J. S., Kellner, M. J., and Zhang, F. (2019). Nucleic acid detection of plant genes using CRISPR-Cas13. *CRISPR J.* 2, 165–171.

Ali, Z., Abulfaraj, A., Idris, A., Ali, S., Tashkandi, M., and Mahfouz, M. M. (2015). CRISPR/Cas9-mediated viral interference in plants. *Genome Biol.* 16:238.

Ali, Z., Aman, R., Mahas, A., Rao, G. S., Tehseen, M., Marsic, T., et al. (2020). iSCAN: an RT-LAMP-coupled CRISPR-Cas12 module for rapid, sensitive detection of SARS-CoV-2. *Virus Res.* 288:198129.

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., et al. (2018). RNA virus interference via CRISPR/Cas13a system in plants. *Genome Biol.* 19:1.

Aman, R., Mahas, A., and Mahfouz, M. (2020). Nucleic acid detection using CRISPR/Cas biosensing technologies. *ACS Synth. Biol.* 9, 1226–1233.

Baranwal, V. K., Kapoor, R., Kumar, S., and Srivastava, N. (2020). "Chapter 2 - Recent advances of virus diagnostics in horticultural crops," in *Applied Plant Virology*, ed. L. P. Awasthi (Cambridge, MA: Academic Press), 27–37.

Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., et al. (2020). CRISPR-Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* 38, 870–874.

Budziszewska, M., Wieczorek, P., and Obrepalska-Steplowska, A. (2016). One-step reverse transcription loop-mediated isothermal amplification (RT-LAMP) for detection of tomato Torrado virus. *Arch. Virol.* 161, 1359–1364.

Cha, D., Kim, D., Choi, W., Park, S., and Han, H. (2020). Point-of-care diagnostic (POCD) method for detecting *Bursaphelenchus xylophilus* in pinewood using recombinase polymerase amplification (RPA) with the portable optical isothermal device (POID). *PLoS One* 15:e0227476. doi: 10.1371/journal.pone.0227476

Chaijarasphong, T., Thammachai, T., Itsathitphaisarn, O., Sritunyalucksana, K., and Suebsing, R. (2019). Potential application of CRISPR-Cas12a fluorescence assay coupled with rapid nucleic acid amplification for detection of white spot syndrome virus in shrimp. *Aquaculture* 512:734340.

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., et al. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 360, 436–439.

Clark, M. F., and Adams, A. N. (1977). Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J. Gen. Virol.* 34, 475–483.

Compton, J. (1991). Nucleic acid sequence-based amplification. *Nature* 350, 91–92.

Daher, R. K., Stewart, G., Boissinot, M., and Bergeron, M. G. (2016). Recombinase polymerase amplification for diagnostic applications. *Clin. Chem.* 62, 947–958.

Ding, X., Yin, K., Li, Z., Lalla, R. V., Ballesteros, E., Sfeir, M. M., et al. (2020). Ultrasensitive and visual detection of SARS-CoV-2 using all-in-one dual CRISPR-Cas12a assay. *Nat. Commun.* 11:4711.

Euler, M., Wang, Y., Heidenreich, D., Patel, P., Strohmeier, O., Hakenberg, S., et al. (2013). Development of a panel of recombinase polymerase amplification assays for detection of biothreat agents. *J. Clin. Microbiol.* 51, 1110–1117.

Faggioli, F., Luigi, M., and Boubourakas, I. N. (2017). "Chapter 36 - Viroid amplification methods: RT-PCR, real-time RT-PCR, and RT-LAMP," in *Viroids and Satellites*, eds A. Hadidi, R. Flores, J. W. Randles, and P. Palukaitis (Boston, MA: Academic Press), 381–391.

Fukuta, S., Iida, T., Mizukami, Y., Ishida, A., Ueda, J., Kanbe, M., et al. (2003). Detection of Japanese yam mosaic virus by RT-LAMP. *Arch. Virol.* 148, 1713–1720.

Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., and Zhang, F. (2018). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* 360, 439–444.

Hadidi, A. (2019). Next-generation sequencing and CRISPR/Cas13 editing in viroid research and molecular diagnostics. *Viruses* 11:120.

Hadidi, A., Kyriakopoulou, P. E., and Barba, M. (2020). "Chapter 1 - Major advances in the history of plant virology," in *Applied Plant Virology*, ed. L. P. Awasthi (Cambridge, MA: Academic Press), 3–24.

Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278.

Huang, W., Yu, L., Wen, D., Wei, D., Sun, Y., Zhao, H., et al. (2020). A CRISPR-Cas12a-based specific enhancer for more sensitive detection of SARS-CoV-2 infection. *EBioMedicine* 61:103036.

Jarošová, J., Ripl, J., Fousek, J., and Kundu, J. K. (2018). TaqMan multiplex real-time qPCR assays for the detection and quantification of *Barley yellow dwarf virus*, *Wheat dwarf virus* and *Wheat streak mosaic virus* and the study of their interactions. *Crop Pasture Sci.* 69, 755–764.

Joshi, S., and Haenni, A.-L. (1984). Plant RNA viruses: strategies of expression and regulation of Viral Genes. *FEBS Lett.* 177, 163–174.

Joung, J., Ladha, A., Saito, M., Kim, N. G., Woolley, A. E., Segel, M., et al. (2020). Detection of SARS-CoV-2 with SHERLOCK one-pot testing. *N. Engl. J. Med.* 383, 1492–1494.

Kanitchinda, S., Srisala, J., Suebsing, R., Prachumwat, A., and Chaijarasphong, T. (2020). CRISPR-Cas fluorescent cleavage assay coupled with recombinase polymerase amplification for sensitive and specific detection of *Enterocytozoon hepatopenaei*. *Biotechnol. Rep.* 27:e00485.

Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O., and Zhang, F. (2019). SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat. Protoc.* 14, 2986–3012.

Kreuze, J. F., Souza-Dias, J. A. C., Jeevalatha, A., Figueira, A. R., Valkonen, J. P. T., and Jones, R. A. C. (2020). "Viral diseases in potato," in *The Potato Crop: Its Agricultural, Nutritional and Social Contribution to Humankind*, eds H. Campos and O. Ortiz (Cham: Springer International Publishing), 389–430.

Kumar, S., Udaya Shankar, A. C., Nayaka, S. C., Lund, O. S., and Prakash, H. S. (2011). Detection of *Tobacco mosaic virus* and *Tomato mosaic virus* in pepper and tomato by multiplex RT–PCR. *Lett. Appl. Microbiol.* 53, 359–363.

Lawson, C., Kaniewski, W., Haley, L., Rozman, R., Newell, C., Sanders, P., et al. (1990). Engineering resistance to mixed virus infection in a commercial potato cultivar: resistance to potato virus X and *Potato virus Y* in transgenic russet Burbank. *Biotechnology* 8, 127–134.

Li, S. Y., Cheng, Q. X., Liu, J. K., Nie, X. Q., Zhao, G. P., and Wang, J. (2018a). CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Res.* 28, 491–493.

Li, S. Y., Cheng, Q. X., Wang, J. M., Li, X. Y., Zhang, Z. L., Gao, S., et al. (2018b). CRISPR-Cas12a-assisted nucleic acid detection. *Cell Discov.* 4:20.

Lin, L., Li, R., Bateman, M., Mock, R., and Kinard, G. (2013). Development of a multiplex TaqMan real-time RT-PCR assay for simultaneous detection of Asian prunus viruses, plum bark necrosis stem pitting associated virus, and peach latent mosaic viroid. *Eur. J. Plant Pathol.* 137, 797–804.

Mahas, A., Aman, R., and Mahfouz, M. (2019). CRISPR-Cas13d mediates robust RNA virus interference in plants. *Genome Biol.* 20:263.

Nagai, K., Horita, N., Yamamoto, M., Tsukahara, T., Nagakura, H., Tashiro, K., et al. (2016). Diagnostic test accuracy of loop-mediated isothermal amplification assay for *Mycobacterium tuberculosis*: systematic review and meta-analysis. *Sci. Rep.* 6:39090.

Nie, X. (2005). Reverse transcription loop-mediated isothermal amplification of DNA for Detection of *Potato virus Y*. *Plant Dis.* 89, 605–610.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., et al. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 28:E63.

Panno, S., Matic, S., Tiberini, A., Caruso, A. G., Bella, P., Torta, L., et al. (2020). Loop mediated isothermal amplification: principles and applications in plant virology. *Plants* 9:461.

Piepenburg, O., Williams, C. H., Stemple, D. L., and Armes, N. A. (2006). DNA detection using recombination proteins. *PLoS Biol.* 4:e204. doi: 10.1371/journal.pbio.0040204

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308.

Rifkind, D., and Freeman, G. L. (2005). "Introduction to part D: viruses," in *The Nobel Prize Winning Discoveries in Infectious Diseases*, eds D. Rifkind and G. L. Freeman (London: Academic Press), 77–80.

Rubio, L., Galipienso, L., and Ferriol, I. (2020). Detection of plant viruses and disease management: relevance of genetic diversity and evolution. *Front. Plant Sci.* 11:1092. doi: 10.3389/fpls.2020.01092

Rybicki, E. P. (2015). A top ten list for economically important plant viruses. *Arch. Virol.* 160, 17–20.

Scholthof, K.-B. G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., et al. (2011). Top 10 plant viruses in molecular plant pathology. *Mol. Plant Pathol.* 12, 938–954.

Shahin, K., Gustavo Ramirez-Paredes, J., Harold, G., Lopez-Jimena, B., Adams, A., and Weidmann, M. (2018). Development of a recombinase polymerase amplification assay for rapid detection of *Francisella noatunensis* subsp. *orientalis*. *PLoS One* 13:e0192979. doi: 10.1371/journal.pone.0192979

Silva, G., Oyekanmi, J., Nkere, C. K., Bomer, M., Kumar, P. L., and Seal, S. E. (2018). Rapid detection of potyviruses from crude plant extracts. *Anal. Biochem.* 546, 17–22.

Uehara-Ichiki, T., Shiba, T., Matsukura, K., Ueno, T., Hirae, M., and Sasaya, T. (2013). Detection and diagnosis of rice-infecting viruses. *Front. Microbiol.* 4:289. doi: 10.3389/fmicb.2013.00289

Vincent, M., Xu, Y., and Kong, H. (2004). Helicase-dependent isothermal DNA amplification. *EMBO Rep.* 5, 795–800.

Zaghloul, H., and El-Shahat, M. (2014). Recombinase polymerase amplification as a promising tool in hepatitis C virus diagnosis. *World J. Hepatol.* 6, 916–922.

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.

Copyright © 2020 Aman, Mahas, Marsic, Hassan and Mahfouz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 December 2020
doi: 10.3389/fmicb.2020.612936

A Mixed Infection of Helenium Virus S With Two Distinct Isolates of Butterbur Mosaic Virus, One of Which Has a Major Deletion in an Essential Gene

John Hammond[1]*, *Michael Reinsel*[1], *Samuel Grinstead*[2], *Ben Lockhart*[3], *Ramon Jordan*[1] and *Dimitre Mollov*[2]

[1] Floral and Nursery Plants Research Unit, United States National Arboretum, United States Department of Agriculture-Agricultural Research Service, Beltsville, MD, United States, [2] National Germplasm Resources Laboratory, Beltsville Agricultural Research Center, United States Department of Agriculture-Agricultural Research Service, Beltsville, MD, United States, [3] Department of Plant Pathology, University of Minnesota, St. Paul, MN, United States

OPEN ACCESS

Edited by:
Ahmed Hadidi,
United States Department of Agriculture (USDA), United States

Reviewed by:
Nicolas Bejerman,
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina
Pedro Luis Ramos-González,
Instituto Biológico, Brazil

***Correspondence:**
John Hammond
john.hammond@usda.gov

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

Received: 01 October 2020
Accepted: 27 November 2020
Published: 21 December 2020

Citation:
Hammond J, Reinsel M, Grinstead S, Lockhart B, Jordan R and Mollov D (2020) A Mixed Infection of Helenium Virus S With Two Distinct Isolates of Butterbur Mosaic Virus, One of Which Has a Major Deletion in an Essential Gene. Front. Microbiol. 11:612936. doi: 10.3389/fmicb.2020.612936

Multiple carlaviruses infect various ornamental plants, often having limited host ranges and causing minor symptoms, yet often reducing yield or quality. In this study we have identified a mixed infection of butterbur mosaic virus (ButMV) and helenium virus S (HelVS) from a plant of veronica (*Veronica* sp.) showing foliar mosaic and distortion. Carlavirus-like particles were observed by transmission electron microscopy (TEM), and RNA from partially purified virions was amplified by random RT-PCR, yielding clones of 439–1,385 bp. Two partially overlapping clones including coat protein (CP) sequence, and two of four partial replicase clones, were closely related to ButMV-J (AB517596), previously reported only from butterbur (*Petasites japonicus*) in Japan. Two other partial replicase clones showed lower identity to multiple carlaviruses. Generic primers which amplify the 3′-terminal region of multiple carlaviruses yielded clones of three distinct sequences: (1) with 98% nt identity to HelVS; (2) ButMV-A, showing 82% nt identity to ButMV-J; and (3) ButMV-B, with 78% nt identity to each of ButMV-J and ButMV-A. Further amplification of upstream fragments revealed that ButMV-B had an internal deletion in TGB1, confirmed using isolate-specific primers. Near-complete genomes of both ButMV-A and ButMV-B were obtained from next-generation sequencing (NGS), confirming the deletion within ButMV-B, which is presumably maintained through complementation by ButMV-A. HelVS was previously reported only from *Helenium* hybrids and *Impatiens holstii*. A near-complete HelVS genome was obtained for the first time by NGS from the same sample. Additional *Veronica* hybrids infected with HelVS were identified by TEM and RT-PCR, including cv. ‘Sunny Border Blue’ which was also subjected to NGS. This resulted in assembly of an 8,615 nt near-complete HelVS genome, with high identity to that from the mixed infection. The predicted CP sequence has 96% amino acid (aa) identity to HelVS from helenium (Q00556). Other

ORFs show a maximum of 54% (TGB3) to 68% (NABP) aa identity to the equivalent ORFs of other carlaviruses. These results demonstrate for the first time maintenance by complementation of a carlavirus isolate with a major deletion in an essential gene, and confirm that HelVS is a distinct species in the genus *Carlavirus*.

Keywords: carlavirus, defective isolate, veronica, complementation, next-generation sequencing

INTRODUCTION

Many ornamental plants are propagated vegetatively by cuttings, or through micropropagation, in order to preserve unique characters obtained and selected through breeding, but which are not stably maintained through seed. Unless breeding populations are rigorously screened for viruses, with only virus-free seedling plants retained for propagation, this leads to the maintenance of viruses present in the original selection. Similarly, if viruses are introduced by vectors such as aphids, or by mechanical transmission due to handling of plants during horticultural operations such as transplanting, pruning, and division, this often results in mixed infections. Aphid-transmitted carlaviruses infect various ornamental plants, and often have limited host ranges; in addition, carlaviruses often produce relatively mild symptoms, which may not be identified by many growers as being due to a viral infection (e.g., Wetter and Milne, 1981). However, even viruses producing minimal symptoms in a particular host may cause more significant symptoms in a different host, or in mixed infections, and even in the absence of obvious symptoms, may reduce yield and/or quality, and ease of propagation (e.g., Brierley and Smith, 1944; Allen, 1975; Valverde et al., 2012). One classical example of a carlavirus affecting ornamentals is the case of lily symptomless virus (LSV) in Easter lily (*Lilium longiflorum*) and hybrid lilies; in single infections of most cultivars of *L. longiflorum*, LSV may cause minimal visible symptoms. However, in comparison of the growth of virus-free Asiatic hybrid lilies obtained from an LSV-infected stock, the virus-free "Enchantment" lilies were found to grow 25–50% taller than LSV-infected stock of the same cultivar, with more uniform height of the crop (Allen, 1975). The combination of LSV with cucumber mosaic virus (CMV) in a mixed infection of at least some Easter lily cultivars was found to cause necrotic stripe in the leaves (Brierley and Smith, 1944).

No carlaviruses, and few other viruses, have previously been reported from any species of *Veronica*. Impatiens necrotic spot virus was reported (as TSWV-I) in ornamental *Veronica* spp. by Hausbeck et al. (1992), while the weedy species *V. persica* has been reported to be naturally infected by CMV (Fletcher, 1989). *V. persica* was experimentally infected with Arabis mosaic virus and strawberry latent ringspot virus using the nematode vector *Xiphinema diversicaudatum* (Thomas, 1970). The weeds *V. arvensis, V. agrestis,* and *V. hederifolia* are reported to be susceptible to CMV, and *V. longifolia* to alfalfa mosaic virus (Zitter, 2001), while plum pox virus has been reported on *V. hederifolia* in Bulgaria (Milusheva and Rankova, 2002).

We were therefore interested to identify the virus(es) associated with obvious mosaic in a plant of veronica submitted to the University of Minnesota Plant Disease Clinic. In this research we identified and characterized (a) a mixed infection of helenium virus S (HelVS) and two distinct isolates of butterbur mosaic virus (ButMV), one of which lacks a major portion of the "essential" TGB1 gene; (b) identified single infection of HelVS in other cultivars of veronica, established the near-complete genome of this isolate by next-generation sequencing, and confirmed it by Sanger sequencing.

Carlavirus species (family *Betaflexiviridae*, genus *Carlavirus*) are currently discriminated by the criteria of having less than about 72% nt identity or 80% aa identity, between their coat protein (CP) or polymerase genes. The complete sequence of ButMV has been available (Hashimoto et al., 2009), and clearly meets the current criteria. Both *Helenium virus S* and *Butterbur mosaic virus* are ICTV-recognized species of the genus *Carlavirus*, and while HelVS has been accepted as a distinct carlavirus species since 1982 (Matthews, 1982), to date there have been only fragments of the HelVS genome available (Foster et al., 1990; Turner et al., 1993). Indeed, to date the "RefSeq" for HelVS (NC_038323) is only a 1,389 nt 3′-proximal fragment including the CP and nucleic acid binding protein (NABP) of Foster et al. (1990).

Carlaviruses have slightly flexuous virions of c.610–700 nm long, with a single-stranded poly-adenylated genome of c.8.3–8.8 kb; the genome encodes six ORFs, with ORF1 being the viral replicase, ORFs2-4 the triple gene block, ORF5 the CP, and ORF6 an NABP (ICTV 9th Report, 2011). Previously HelVS had been reported only from Helenium hybrids (e.g., Kuschki et al., 1978) and *Impatiens* species (Koenig et al., 1983), while ButMV has been reported only from *Petasites japonicus* in Japan (Hashimoto et al., 2009) and Korea.

The results generated here represent the first near-complete genome sequence of HelVS, and demonstrate veronica as a new host for both HelVS and ButMV. Although mixed infections of different species of carlavirus have been reported previously (e.g., Van Dijk, 1993; Eastwell and Druffel, 2012; Richert-Pöggeler et al., 2015; Ho et al., 2016), and mixed isolates of other viruses in the same plant are not uncommon (e.g., Lim et al., 2010), we are not aware of any previous reports of a defective carlavirus isolate lacking an essential gene being maintained in a mixed infection with a functional isolate.

MATERIALS AND METHODS

Plant Material

A plant of a *Veronica* hybrid showing mosaic symptoms (**Figure 1**) and submitted to the Plant Disease Clinic at the University of Minnesota, was the initial subject of investigation;

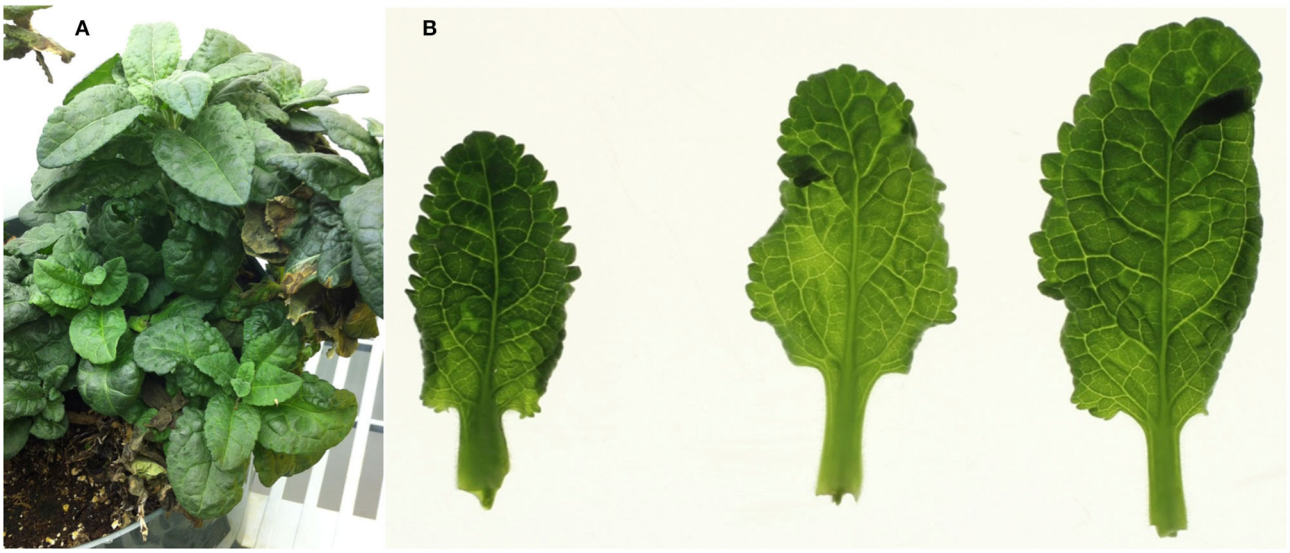


FIGURE 1 | Symptoms in the whole plant, and detached leaves, of the original (Minnesota; MN) plant of *Veronica* sp., with a mixed infection of HelVS-V with ButMV-A and ButMV-B. **(A)** Whole plant showing mild foliar mosaic and puckering; and **(B)** detached leaves viewed on a lightbox, more clearly showing mottle/mosaic and interveinal chlorosis, with some puckering.

the plant was subsequently transferred to the USDA-ARS Beltsville Agricultural Research Center, Beltsville, Maryland, where it was maintained in a growth chamber until the plant eventually died. This plant is hereafter referred to as the "MN" (Minnesota) plant. After the identification of a mixed infection in this original plant, three additional plants of *Veronica*, cultivars 'Sunny Border Blue,' 'Red Fox,' and 'Purpleicious,' showing mild mottle or mosaic symptoms were obtained from a local (Maryland, MD) nursery. These plants were maintained in a greenhouse under nominal day/night conditions of ∼24/21°C, under natural daylight with supplemental light as necessary under cloudy conditions, and to extend day length to 14 h as required.

Transmission Electron Microscopy (TEM)

Symptomatic leaves were initially examined at the University of Minnesota by electron microscopy. Extracts of liquid nitrogen-powdered leaves were partially purified and concentrated using ultracentrifugation through a 30% sucrose pad as described (Mollov et al., 2013). Grids were stained with 2% sodium phosphotungstic acid (PTA) pH 7.0 and visualized on a Philips electron microscope at the University of Minnesota Imaging Center. Leaf extracts of cvs. 'Sunny Border Blue,' 'Red Fox,' and 'Purpleicious' were prepared at USDA-ARS in Beltsville, MD, by grinding leaf pieces with a small volume of PTA and transferring a drop of the liquid to a formvar-coated copper grid for examination in a JEOL 100CX II transmission electron microscope (JEOL USA, Inc., Peabody, MA). Leaf tissue was also embedded and sectioned to examine cytopathological effects. Grids for either negatively-stained or thin sections of embedded tissue were typically examined at microscope magnifications of 33,000×, 50,000×, and 66,000×, and images captured with an AMT HR digital camera system (Advanced Microscopy Techniques Corp., Woburn, MA).

Initial Partial Purification of Virions, cDNA Production and Cloning

At the University of Minnesota, the partially purified virion inclusive pellet was resuspended in DNase buffer and treated with DNase. Virion nucleic acids were extracted with SDS-phenol and chloroform as described (Mollov et al., 2013). A genomic library from the virion RNA was generated by random PCR technique following the protocol described by Froussard (1992). Six distinct cloned random PCR amplicons were cloned and sequenced and subjected to online BLASTn analysis at the NCBI website.

Additional Cloning and Sequencing

Total RNA was extracted from additional leaves from the MN plant using the RNeasy Plant Mini Kit (Qiagen, Germantown, MD, USA), and cDNA was prepared using tagged oligo(dT) primer NSNC-odT (Hammond et al., 2006); PCR was performed using degenerate primer PxDeg and BNSNC, which amplify the 3′-terminal region of both potexviruses and carlaviruses (Hammond et al., 2006; Hammond and Reinsel, 2011). Additional primers were designed from the TGB2 region of the Japanese isolate of ButMV (AB517596), from the sequences of the cloned random PCR amplicons, and from the initially obtained three types of 3′-terminal region clones. Total nucleic acids (primarily RNA) was purified from leaves of *Veronica* cvs. 'Sunny Border Blue,' 'Red Fox,' and 'Purpleicious' using either the RNeasy Plant Mini Kit (for all isolates), and additionally the "CKC" method (Henderson and Hammond, 2013) for 'Sunny Border Blue' (see below), and initial cloning performed using

amplicons generated with PxDeg/BNSNC. Sequences of all of the primers noted in this work are listed in **Supplementary Table 1**.

PCR products were cloned using the TOPO cloning kit (Invitrogen) and sequenced. Primers designed from the sequences obtained from the initial random PCR products and the 3′-terminal PCR products were used to extend the sequences and determine which upstream products were derived from the same templates as the 3′-terminal clones.

In order to verify apparent differences in size of related PCR-derived sequences derived from extension of the initial 3′-proximal PCR products, one pair of degenerate primers capable of amplifying both variants (ButRepFdeg/ButTGBRdeg) was used to amplify cDNA prepared from fresh RNA extracts from the mixed infection; two additional reverse primers, each specific for one sequence variant (ButA-TGB; and ButB-TGB) were each paired separately with ButRepFdeg for amplification, and the products analyzed by agarose gel electrophoresis.

Library Preparation and Next-Generation Sequencing (NGS)

A total RNA extract was obtained from the original MN plant using the RNeasy kit (Qiagen). The purified RNA was submitted to Macrogen (South Korea) for cDNA library preparation and sequencing on an Illumina HiSeq2500 as 100 bp single end reads.

Total nucleic acids from the MD plant *Veronica* cv. 'Sunny Border Blue,' prepared using the CKC method, were submitted for ribosomal RNA depletion (RiboZero kit, Illumina) and cDNA library generation by Genewiz (South Plainfield, NJ, USA). The resulting cDNA library was sequenced in-house at USDA-ARS using 250 bp paired-end protocols on an Illumina MiSeq instrument, in parallel with other libraries, each prepared with different indices.

Sequence Analysis

The sequences of the six initial random PCR clones were analyzed by BLASTn and BLASTp against viral sequences in GenBank. Additional sequences were similarly analyzed by BLAST, and further analyzed by local alignments, CLUSTAL alignments, and translation using tools at NCBI, and Geneious Pro or Geneious Prime (Biomatters, Inc., San Diego, CA).

Next-Generation Sequencing

The bioinformatic pipelines utilized here were based on steps in either the CLC Workbench or Geneious packages, and are broadly comparable to other protocols that have been used by other groups. Although it has been noted that the yields of viral sequences is dependent on viral genome organization and the proportion of viral reads in the data (Pecman et al., 2017), Villamor et al. (2019) note that all of the many pipelines available share a common backbone. Massart et al. (2019) compared the ability of 21 different plant virology laboratories to detect viruses in the same 10 small RNA (21–24 nt) datasets from three different plant species, and found a high level of reproducibility between laboratories applying different pipelines to the same datasets. We therefore used the software packages available in our respective laboratories.

Next-Generation Sequencing Analysis of the MN Sample

The 100 nt SE raw data reads were trimmed and *de novo* assembled using CLC Workbench (CLC Bio; Qiagen, USA) (8.5.1) for the ButMV analyses and version 20.0.3 for the HelVS analyses. The assembled contigs were queried against a protein database containing RefSeq virus proteins, with UniProt Arabidopsis proteins included to reduce false positive hits. Viral candidate contigs were subjected to BLASTx against the full NCBI nr database (locally downloaded to enhance speed) to refine results.

For the HelVS, data QC parameters were base caller prediction reliability $p = 0.05$ with maximum two ambiguities; assembler parameters were auto bubble and word size with minimum contig length of 175 nt. Initial BLASTx (RefSeq virus plus Arabidopsis database) parameters were BLOSUM62, gap existence cost = 11 and extension cost = 1. The second round BLASTx to the full nr database was made with the same parameters. After BLAST analyses, raw data was mapped to contigs at the criteria of 80% ID and 50% query coverage. To obtain an optimal contig length, the data was sampled at 5% to reduce coverage from up to 18,000x to no more than 900x. Further subsampling of three, two, and one percent were evaluated to obtain optimal contig length. Final assembly was at 2 and 3% subsampling. The full NGS dataset was mapped back (90% ID and 75% coverage) against this putative reassembly to correct any contig disagreements and look for any gaps or evidence of different isolates. After examining the ORFs (start codon AUG only, both strands, standard genetic code), the sequence was reverse complemented.

For the ButMV assemblies, data quality was trimmed ($p = 0.05$, maximum 2 nt ambiguity) and assembled using default auto bubble and word size, with minimum contig length 125 nt. The same BLAST algorithms were used as with the HelVS assemblies, but no subsampling was necessary.

Next-Generation Sequencing Analysis of the MD Sample ('Sunny Border Blue')

The paired-end raw data reads (7,592,256 reads of 75–250 nt) from the MiSeq were binned by the specific index for the particular library, trimmed to remove adapter linkers, and *de novo* assembled into contigs using Geneious Pro R9. The Geneious Mapper parameters were set at Med-Low sensitivity and Fast assembly, 5 iterations, with 80% minimum overlap identity, allowing maximum 4 nt ambiguity, with minimum contig length 100 nt. The resulting 140,768 total contigs were analyzed by BLASTx (Blosum62, gap cost = 11, open extend = 1) against a local database of viral sequences from GenBank. Nineteen contigs (ranging in size from 1.1 kb to 5.9 kb) having high identity to carlaviruses were further edited and reassembled using the same Geneious Mapper parameters.

Validation of the NGS Sequence From MD Plant of 'Sunny Border Blue'

Primers (see **Supplementary Table 1**) designed from the assembled carlavirus-like sequences determined by NGS

(HelVS-Ver, NGS) were used to generate clones of seven overlapping PCR products representing almost the full genome determined by NGS (HelVS-Ver, PCR validation). The HelVS-Ver NGS genome was also compared to the assembled c.4.1 kb PCR-derived portion of the original MN *Veronica* isolate with the mixed infection (HelVS-V); to HelVS-VR21, the NGS sequence also derived from the MN *Veronica* isolate HelVS-V; and to the partial HelVS sequences available in GenBank.

Phylogenetic Analysis

The genomic nt coding region (including the full sequences of all recognized functional ORFs) of four virus isolates obtained in this study [HelVS-VR21 (=HelVS-V), ButMV-A, and ButMV-B isolates from the MN Veronica; and HelVS-Ver from Veronica 'Sunny Border Blue'] were used in phylogenetic analysis with 33 carlavirus genomes obtained from NCBI. Additionally, the deduced aa sequences of the RdRp and CP ORFs for the four virus isolates were used in phylogenetic analyses with the respective carlavirus RdRp and CP aa sequences of the same 33 carlaviruses. In all cases sequences were aligned using Clustal W.

Maximum likelihood phylogenies were built using Mega X, with substitution model Tamura-Nei (Tamura and Nei, 1993), with 1,000 bootstrap replications (Kumar et al., 2018), and a condensed tree was built by collapsing branches with <50% bootstrap support.

Genome reference sequences of the carlaviruses most closely linked to HelVS in the phylogenetic trees, and identified by BLASTn, were separately compared to HelVS-VR21 using the pairwise Geneious alignment function to determine % nt identities.

Pairwise Sequence Comparison (PASC) Analysis

The HelVS-Ver sequence was subjected to analysis using the Pairwise Sequence Comparison (PASC) tool at the NCBI website (Bao et al., 2014) to provide an addition means of verifying that HelVS is indeed a distinct species.

RESULTS

Visualization of Virus Particles

Examination of negatively-stained partial virus purifications at the University of Minnesota revealed slightly flexuous particles typical of carlaviruses; similar carlavirus-like virions were also detected from leaves of 'Sunny Border Blue,' 'Red Fox,' and 'Purpleicious' (data not shown). Examination of thin sections of embedded leaf material of 'Sunny Border Blue' revealed apparent slightly flexuous particles in the cytoplasm (**Figure 2**), similar to those observed by Koenig et al. (1983), but no banded inclusions as have been reported for various carlavirus infections by others (e.g., Wetter and Milne, 1981).

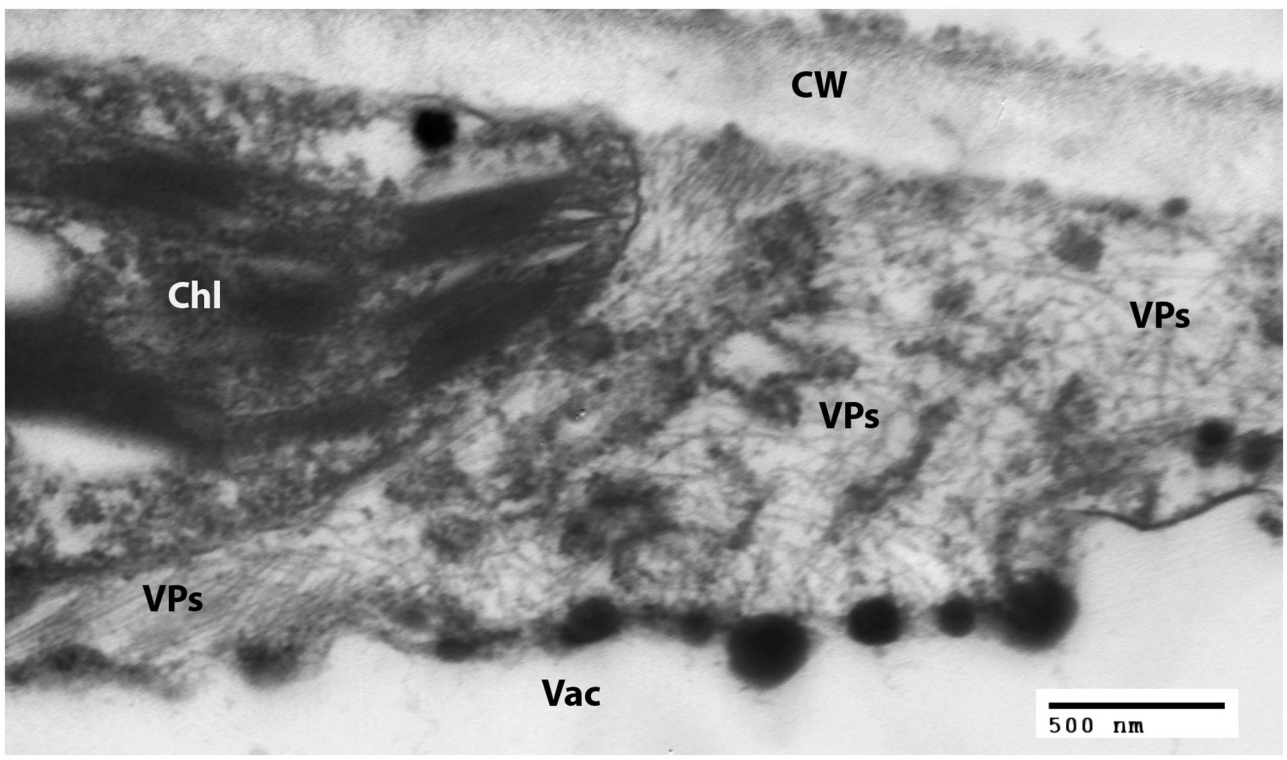


FIGURE 2 | Transmission electron micrograph of a thin section of a leaf of *Veronica spicata* cv. Sunny Border Blue, showing loose aggregations of viral particles in the cytoplasm. Chl, Chloroplast; CW, Cell wall; Vac, Vacuole; VPs, Viral particles.

Identification of the Origins of the Random PCR Products

The initial random PCR amplification yielded six distinct clones identified as of viral origin (**Table 1**); four of these clones were found to be derived from carlavirus RdRp sequences, with two partially overlapping sequences being closely related to the full genome of ButMV from Japan (ButMV-J; NC_013527) (RdRp-a, c.76%; and RdRp-c, c.80% nt identity), but distinct from each other (c.75% identity). Another partial sequence (RdRp-b) had only ∼60% coverage at c.71% nt identity to ButMV-J, but 97% coverage at 73% identity to Gaillardia latent virus (GLV; KJ415259) and 94.4% identity over a 177 nt overlap to the Impatiens isolate of HelVS (HelVS-I; FJ555524); the fourth partial RdRp clone (RdRP-d) had a 100% nt identity of its 3′-end in an 86 nt overlap with the 5′-end of RdRp-b, no significant identity to ButMV-J, and higher identities with incomplete coverage to GLV (KJ415259; 65% coverage, 73% identity), Phlox virus S (PhlVS, EF492068; 49% coverage, 77% identity), Phlox virus B (PhlVB, EU162589; 51% coverage, 72% identity), and Chrysanthemum virus B (CVB, AM118099; 60% coverage, 71% identity). Two additional clones were found to have an overlap of ∼450 nt in the coat protein (CP) sequence, but to share only c.78% nt identity within this overlap; the larger fragment (CP-a) included most of TGB2, all of TGB3, and most of the CP ORF, and had 79% nt identity to ButMV-J, while fragment CP-b was totally within the CP ORF, and had 82% nt identity to ButMV-J.

Thus, the initial random PCR clones provided evidence for the presence of two distinct isolates of ButMV, as well as suggesting the presence of an additional species of carlavirus, initially not clearly identified.

PCR-Based Cloning and Sequencing From the 3′ Region of the Viral Genomes

Cloning of the 3′-terminal region with either the degenerate primer PxDeg or a ButMV-specific TGB2 primer yielded three distinct sets of clones of c.1.7–1.85 kb; the sequence of these clones revealed two distinct sequences closely related to ButMV-J (designated as ButMV-A and ButMV-B), and a third sequence identified as HelVS by BLASTn analysis. We designate the HelVS sequence HelVS-V to distinguish it from the original Helenium isolate (D10454, S71594) and HelVS-I (from Impatiens; FJ555524) Further PCR amplifications were undertaken with specific forward primers designed from each of the RdRp clones, and specific reverse primers from each of the 3′-terminal clones (data not shown), or the BNSNC primer, in order to bridge the gap between the RdRp and 3′-terminal clones (**Supplementary Figure 1**) and generate continuous sequences of ButMV-A and ButMV-B.

This strategy resulted in generation of sets of clones enabling assembly of consensus sequences spanning 3,765 nt of ButMV-A, 3,165 nt for ButMV-B, and 3,052 nt for HelVS-V (not including the 3′ poly-A tract), with a minimum of three clones per fragment. The HelVS-V 3,052 nt sequence was then extended, by use of primer VerRdRpD-F2 designed from random PCR clone RdRp-d, paired with HVS-rep4 derived from the 3,052 nt HelVS-V sequence; after direct sequencing of the resulting PCR product using primer VerRdRpD-F1 (partially overlapping and extending downstream of VerRdRpD-F2), and sequencing the complementary strand using the PCR primer HVS-rep4. By this means the HelVS-V sequence was extended to 4,138 nt.

The HelVS-V 3′-terminal sequence was linked to RdRp-b (98.7% identity in ∼153 nt overlap). The HelVS-V sequence has 98% nt identity to both available fragments of the Helenium

TABLE 1 | Initial BLASTn and final identifications of origins of random PCR products.

Random PCR product, Acc. No.	Size (nt)	Initial identification	Query coverage, isolate	Query % identity, Acc. No.	Final NGS sequence identification	% identity to NGS sequence
RdRp-a MW226001	1,026	ButMV	99%[a] ButMV-J	76.27 NC_013527	ButMV-B	99.81
RdRp-b MW226002	950	GLV	97% (DSM) [b]	73.43[c] KJ415259	HelVS-V	99.58
RdRp-c MW226003	585	ButMV	99% ButMV-J	79.79 NC_013527	ButMV-A	99.49
RdRp-d MW226004	439	GLV[d]	65% (DSM)	76.21 KJ415259	HelVS-V	99.54
CP-a MW225999	1,385	ButMV	99% ButMV-J	79.23 NC_013527	ButMV-A	99.71
CP-b MW226000	775	ButMV	99% ButMV-J	81.63 NC_013527	ButMV-B	99.87

[a] *Pairwise alignment of RdRp-a and RdRp-c revealed 74.40% identity over the full length of RdRp-c.*

[b] *Deutsche Sammlung von Mikroorganismen (DSM) Gaillardia latent virus isolate 5/18-05-2010.*

[c] *Separate pairwise alignment with the ButMV-J sequence AB517596 revealed only 59% coverage at 71.01% identity. Pairwise BLASTn of RdRp-b against the partial RdRp sequence of the Impatiens isolate of HelVS revealed an overlap of the 3′-terminal 177 nt of RdRp-b with the 5′ FJ555524 at 92.39% identity.*

[d] *A pairwise BLASTn revealed no significant similarity to ButMV-J, but comparing RdRp-b and RdRP-d revealed a direct overlap (100% identity) of the 3′-terminal 86 nt of RdRp-d with the 5′-terminal region of RdRp-b, supporting their origin from the same source.*

TABLE 2 | Nucleotide identities of HeLVS sequences obtained from two Veronica sources and different methods, and compared to available GenBank accessions.

Sequence, type, length, GenBank accession no.	HelVS-VR21[a] (MN isolate), NGS, 8,668 nt, MW207172	HelVS-Ver (MD isolate, 'Sunny Border Blue'), NGS, 8,615 nt, MW246808
HelVS-VR21, NGS, 8,668 nt, MW207172		98.58%
HelVS-V[a], PCR clone assembly, 4,138 nt, MW246812	99.76%	98.66%
HelVS-Ver, NGS, 8,615 nt, MW246808	98.58%	
HelVS-Ver, PCR validation, 8,460 nt, MW246809	98.59%	99.95%
HelVS-I, partial RdRp and TGB1, 935 nt, FJ555524[b]	94.98%	94.93%
HelVS, CP, and NABP, 1,389 nt, D10454[c]	98.13%	97.95%
HelVS, partial TGB2, TGB3, 381 nt, S71594[c]	97.64%	97.90%

[a] *HelVS-VR21 and HelVS-V are, respectively, the NGS-determined, and PCR-assembled, sequences of the MN isolate termed HelVS-V.*
[b] *HelVS-I, isolate from Impatiens, from the Netherlands.*
[c] *HelVS, original Helenium isolate, from the United Kingdom.*

isolate (D10545, CP and NABP; S71594, partial TGB2 and TGB3), and 96% nt identity to HelVS-I (FJ555524) (**Table 2**).

ButMV-B Has a Significant Deletion With Respect to ButMV-A and ButMV-J

The ButMV-A 3′-terminal sequence was linked to RdRp-c (98.7% nt identity in ∼75 nt overlap), while the ButMV-B 3′-terminal sequence was linked to RdRp-a (99.4% nt identity in ∼475 nt overlap). However, the ButMV-B sequence showed a significant internal deletion compared to ButMV-A and ButMV-J; the extended ButMV-A showed 81% nt identity to ButMV-J, while the ButMV-B sequence showed 72 and 77% nt identity to ButMV-J for the portions separated by the deleted sequence. ButMV-A and ButMV-B showed 71 and 78% nt identity to each other for the portions separated by the deletion in ButMV-B (**Table 3**).

The presence of a significant deletion in the genome of ButMV-B was examined further using PCR with a degenerate forward primer designed to anneal to both ButMV-A and ButMV-B in the 3′ portion of the replicase gene, paired with: (a) ButMV isolate-specific reverse primers to span the presumed deletion; and (b) degenerate primers designed to amplify both ButMV-A and ButMV-B across the region of the deletion. Although dual PCR products corresponding to ButMV-A (1,212 bp) and ButMV-B (613 bp) were obtained

TABLE 3 | Nucleotide identities of ButMV sequences obtained from Veronica sources and different methods, and compared to available GenBank accessions.

Sequence, length, type, GenBank Acc. no.	ButMV-J, 8,662 nt, RefSeq, NC_013527	ButMV-A, Veronica, NGS, 8,596 nt, MW207173	ButMV-B, Veronica, NGS, 8,033 nt, MW207174
ButMV-J, RefSeq genome, 8,662 nt, NC_013527		77.71%	68.17%[a]
ButMV-A, Veronica, NGS genome, 8,596 nt, MW207173	77.71%		67.55%[a]
ButMV-B, Veronica, NGS genome, 8,033 nt, MW207174	68.17%[a]	67.55%[a]	
ButMV-A, Veronica, 3′ PCR clones, 3,768 nt, MW246810	81.10%	99.44%	65.97%[a]
ButMV-B, Veronica, 3′ PCR clones, 3,167 nt, MW246811	65.76%[a]	65.87%[a]	99.72%
ButMV-B, partial 5′+RdRp, 6,040 nt, (MW207174)[b]	71.75%	70.77%	
ButMV-B, TGB2,TGB3, CP, NABP-3′, 1,993 nt, (MW207174)[c]	77.12%	77.76%	
ButMV-Kr, CP gene, 973 nt, MK689358[d]	78.69%	78.59%	74.52%

[a] *These figures are calculated based on alignment of the complete 8,033 nt ButMV-B sequence, which has a deletion of almost the full TGB1 ORF. Because the alignment algorithms attempt to identify matches which minimize the presence of gaps, the results underestimate the nucleotide identities between the undeleted portions. We therefore separately analyzed identities to (1) the 6,040 nt region of ButMV-B containing the partial 5′-UTR, the full RdRp reading frame, and a portion of the intergenic region corresponding to that of "complete" isolates between the RdRp and TGB1 ORFs; and (2) the 1,933 nt region corresponding to the remainder of the intergenic region of "complete" isolates, plus the complete TGB2, TGB3, CP, and NABP ORFs, and the 3′-UTR.*
[b] *Derived from MW207174, as explained above; the 6,040 nt sequence consisting of the partial 5′-UTR, the full RdRp reading frame, and a portion of the intergenic region corresponding to that of "complete" isolates between the RdRp and TGB1 ORFs was compared against the sequences of "complete" isolates ButMV-J and ButMV-A.*
[c] *Derived from MW207174, as explained above; the 1,933 nt region corresponding to the remainder of the intergenic region of "complete" isolates, plus the complete TGB2, TGB3, CP, and NABP ORFs, and the 3′-UTR was compared against the sequences of "complete" isolates ButMV-J and ButMV-A.*
[d] *ButMV isolate Kr-NS5 from butterbur (Petasites sp.; probably P. japonicus) was detected in Korea. ButMV-J was isolated in Japan from P. japonicus, the only known natural host prior to discovery of ButMV in Veronica sp. in the USA.*

with degenerate primers (ButRepFdeg/ButTGBRdeg) designed to amplify both sequences, only products representing ButMV-B sequences (with the deletion; 694 bp) or ButMV-A sequences (lacking the deletion; 1,293 bp) were recovered using the isolate-specific reverse primers (**Figure 3**), providing strong evidence that the deletion in the ButMV-B sequence was an authentic feature of this isolate, and not an artifact of cloning.

NGS Results From the MN Veronica Plant

The NGS run from the MN plant sample produced 29,056,208 total reads that assembled into 63,614 contigs. BLAST analysis initially revealed 31 contigs related to carlaviruses. An initial ButMV contig (later named ButMV-A) of 8,596 nt was derived from the CLC assemblies. Additionally, seven more contigs with similarities to ButMV were identified. Subsequent reassembly of these contigs merged to one consensus contig of 8,499 nt. This shorter contig (ButMV-B) had evident differences, and with further trimming and alignments, the consensus was established to be 8,033 nt long and had a significant deletion encompassing almost the full TGB1 gene when compared to the original ButMV (ButMV-A) assembly (**Figure 4**). Additional bioinformatics analysis produced a single HelVS contig in addition to the two distinct ButMV contigs. No significant sequences related to any other viruses were detected.

The HelVS contig was 8,668 nt (excluding the poly-A), composed of 2,001,064 reads with an average coverage of 23,072× per nt position; this sequence was designated HelVS-VR21, and represents the same HelVS isolate as the PCR-derived partial genome sequence HelVS-V. The ButMV-A isolate contig was 8,596 nt with 1,561,852 reads mapped to it and 17,773 average coverage per nt position. The isolate ButMV-B contig was 8,033 nt with 1,785,592 reads mapped to it and 21,849 average coverage per nt position.

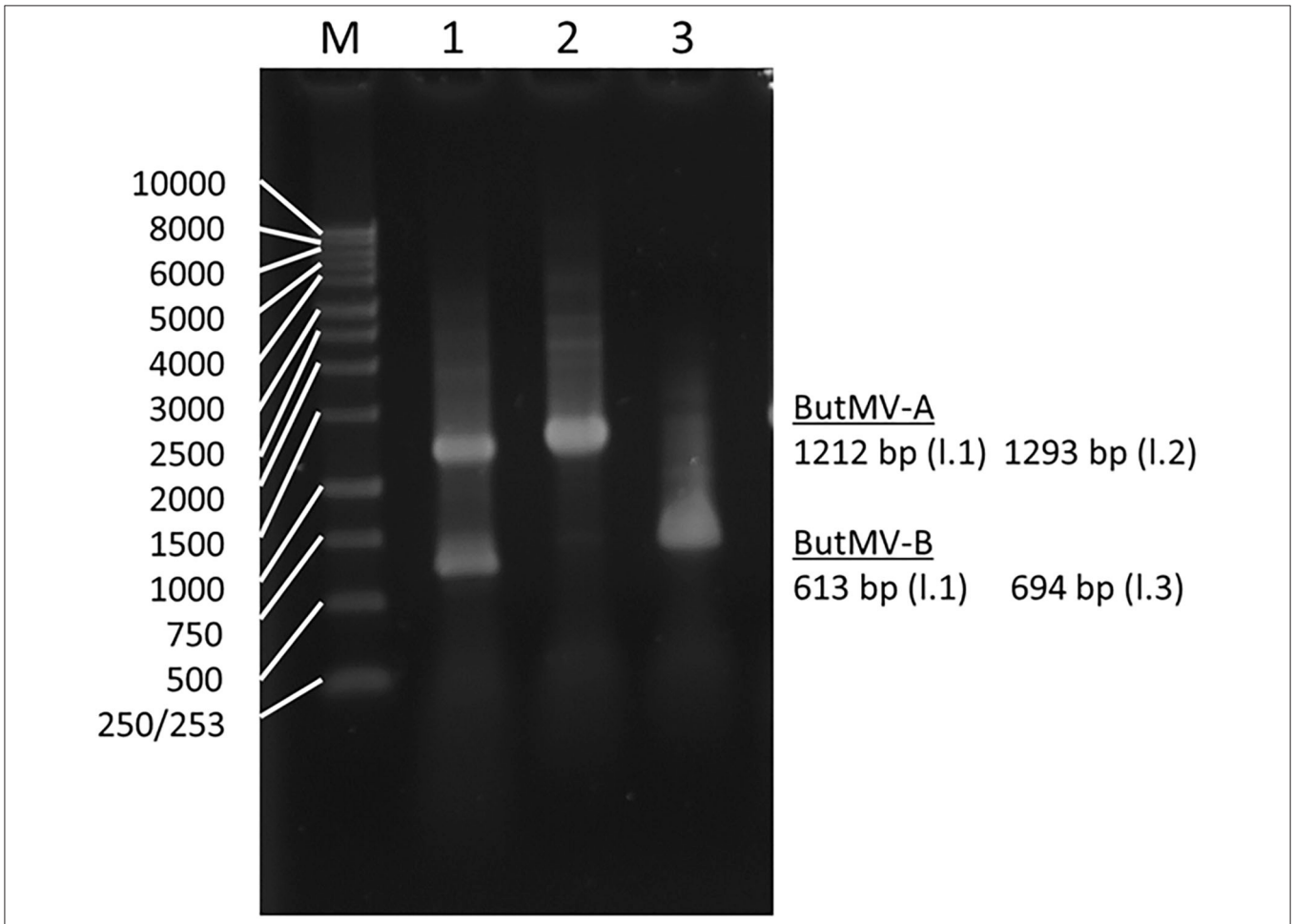


FIGURE 3 | Reverse-transcription polymerase chain reaction products of assays to confirm the deletion of most of the TGB1 reading frame in ButMV-B, with RNA extracted from the original plant with the mixed infection. Lane 1—Products of degenerate primer pair ButRepFdeg/ButTGBRdeg which amplify fragments of 1,212 bp from ButMV-A, and 613 bp from ButMV-B; Lane 2—Products of ButRepFdeg paired with ButMV-A-specific reverse primer ButA-TGB, yielding an amplicon of 1,293 bp; Lane 3—Products of ButRepFdeg paired with ButMV-B-specific reverse primer ButB-TGB, yielding an amplicon of 693 bp. Lane M = 1 kb Marker ladder.

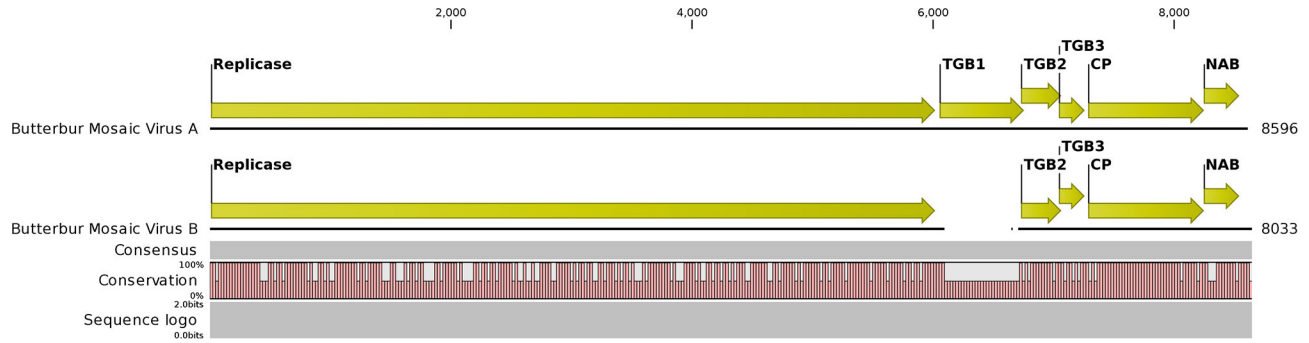


FIGURE 4 | Graphical representation of an alignment of the next-generation sequencing results for the butterbur mosaic virus analysis of the mixed infection ('Conservation' line). Note gap in the ButMV-B sequence (upper), compared to ButMV-A (lower), including almost all of the TGB1 coding sequence, and compare to the ORF diagram (Top). This gap corresponds to the shorter PCR products from ButMV-B as seen in **Figure 3**. Degenerate forward primer ButRepFdeg is located at ButMV-A nt 5,835–5,858, and at ButMV-B nt 5,838–5,861; degenerate reverse primer ButTGBRdeg is complementary to ButMV-A nt 7,024–7,046, and ButMV-B nt 6,428–6,450. ButMV-A specific reverse primer ButA-TGB is complementary to ButMV-A nt 7,103–7,127, and ButMV-B specific reverse primer ButB-TGB is complementary to ButMV-B nt 6,507–6,531 (at equivalent positions in the aligned ButMV-A/ButMV-B genomes shown, but with multiple mis-matched positions compared to the other isolate).

Comparison of Veronica Isolates ButMV-A and ButMV-B With the Type Isolate, ButMV-J, and Partial Sequence of ButMV-Kr-NS5

The NGS sequences of ButMV-A and ButMV-B were clearly consistent with data obtained by PCR amplification, showing a deletion of essentially the full TGB1 reading frame of ButMV-B (**Figure 4**). The portion of ButMV-B upstream of the deletion shared a lower degree of nt identity to the equivalent region of ButMV-A (70.77% nt) compared to 77.76% for the region including from TGB2 to the 3′-end of the NGS sequence; similar differences were observed for the same regions of ButMV-B compared to ButMV-J (**Table 3**). When sequences were compared in this manner, all of the comparisons between isolates showed >70% nt identity (**Table 3**). However, pairwise comparisons of each of the ORFs revealed values of c.65.7–94.9% aa identity for each ORF, with ButMV-B showing the lowest identity in all ORFs except for TGB2, which had 91.1% aa identity to that of ButMV-A; TGB3 showed the greatest diversity between isolates (c.65.7–75.7% aa identity) (**Table 4**).

The TGB1 ORF (missing in ButMV-B) of ButMV-A and ButMV-J each had a potential TTG alternate initiation codon in identical context (GTT.AGT.**TTG**.AAA.TAT.**ATG**.GAC) three codons upstream of the presumed native start codon, which would result in addition of three N-terminal residues (MKY). TGB2 of ButMV-A and ButMV-J each had CTG in a slightly different context (TCC.TTA.**CTG**.GAG.**ATG**.CCT for ButMV-A; TTC.TGA.**CTG**.GAG.**ATG**.CCT for ButMV-J), but in each case encoding two additional residues (ME); ButMV-B had no potential alternate initiation codon. The CP ORFs of each isolate had potential TTG starts in similar contexts, which would add different combinations of seven or eight additional residues (GTT.TGA.**TTG**.AAT.ACG.AAT.TCC.CAA.AAA.CAT.**ATG**.GGG, MNTNSAKH for ButMV-A; TTT.TGA.**TTG**.AAT.AGT.TTT.TCC.AAA.CAA.**ATG**.GGT, MNSFSKQ for ButMV-B; TTT.TGA.**TTG**.AAT.ACG.AAT.TCC.CAA.AAA.CAT.**ATG**.GGG, MNTNSAKH for ButMV-J). There was a potential CTG initiation codon immediately upstream of the native ATG

TABLE 4 | Percentage of amino acid identities between ORFs of ButMV isolates.

Sequence	ButMV-J	ButMV-A	ButMV-B
RdRp			
ButMV-J		85.93	79.53
ButMV-A	85.93		78.78
ButMV-B	79.53	78.78	
TGB1			
ButMV-J		88.05	[a]
ButMV-A	88.05		[a]
ButMV-B	[a]	[a]	
TGB2			
ButMV-J		86.61	86.61
ButMV-A	86.61		91.07
ButMV-B	86.61	91.07	
TGB3			
ButMV-J		75.71	65.71
ButMV-A	75.71		67.14
ButMV-B	65.71	67.14	
CP			
ButMV-J		91.10	88.65
ButMV-A	91.10		86.81
ButMV-B	88.65	86.81	
NABP			
ButMV-J		94.85	85.71
ButMV-A	94.85		83.51
ButMV-B	85.71	83.51	

[a] *ButMV-B has a deletion of almost the entire TGB1 reading frame, leaving only a portion of ∼14 codons, of which seven codons overlap with the TGB2 initiation codon and reading frame. There are five termination codons downstream of the ButMV-B RdRP termination codon, and in frame with the remaining fragment of the TGB1 reading frame.*

in the NABP ORFs of ButMV-B and ButMV-J in slightly different context (CTC.GGA.**CTG**.**ATG**.TCA for ButMV-B; CTC.TGA.**CTG**.**ATG**.TCG for ButMV-B) adding only an additional Met residue, but not for ButMV-A. Note that in all three isolates the native NABP ATG overlaps the termination codon of the CP ORF in the (−1) reading frame.

NGS Results From Veronica 'Sunny Border Blue'

The raw reads (7,592,256 reads of 75–250 nt) were assembled into contigs, analyzed by BLASTx, and those contigs having high identity to carlaviruses were further edited and assembled into a 8,615 nt near-complete sequence of the isolate named HelVS-Ver. The mean coverage per nt position was 14,493 reads with 892,791 total reads mapping to the consensus sequence. There was significant coverage across all regions of the genome except the extreme 5′ and 3′ regions (**Supplementary Figure 2**). The HelVS-Ver sequence lacks an estimated 41–70 nt of the 5′ UTR (by comparison to the length of HelVS-VR21, and of carnation latent virus, respectively; this report, and R. Jordan, D. Mollov, and S. Grinstead, unpublished), and 25 nt of the 3′-UTR (by comparison to previously determined HelVS sequences from veronica, including the 3′-terminal clones developed here). Primers designed from the NGS-derived genomic sequence were used to generate seven over-lapping amplicons extending to the 3′ poly(A), which were cloned and sequenced to validate the NGS-derived genome; a minimum of five clones representing each of these overlapping PCR products were Sanger sequenced in both directions, and the consensus sequences assembled to create a validated sequence representing 8,464 bp (**Supplementary Figure 3**).

Comparison of HelVS Sequences HelVS-Ver, HelVS-VR21, and Partial Sequence of Helenium and Impatiens Isolates

The NGS genomes [HelVS-Ver and HelVS-VR21 (from isolate HelVS-V)] were also compared to the 4.1 kb sequence previously obtained by PCR extension from the MN veronica isolate HelVS-V. The NGS 8,615 nt sequence of HelVS-Ver and the 8,464 nt PCR-validated sequence showed 99.95% nt identity, while the NGS sequences from the MN isolate HelVS-V and HelVS-Ver shared 98.58% identity over the length of the HelVS-Ver sequence (**Table 2**). The near-complete NGS sequence of HelVS-V (=HelVS-VR21) and the PCR extension sequence of 4,138 nt (HelVS-V) from the same plant shared 99.76% nt identity, similar to the degree of identity between the NGS and PCR-validated sequence of HelVS-Ver (**Table 2**). The full genome and 3′-UTR nt sequences of HelVS-Ver and HelVS-VR21, and their predicted aa sequences of each of the ORFs, all shared a minimum 98.58% identity, with 100% identity for the TGB3 and NABP ORFs (**Table 5**).

Interestingly, both the TGB1 and TGB3 ORFs of HelVS-Ver and HelVS-V had potential TTG alternative start sites upstream of the presumed native ATG codon, which would, if utilized, add the same N-terminal aa residues: MF in the case of TGB1; and MWPASL for TGB3. The sequence context for the potential alternate start codons for both isolates was shared for TGB1 (GTG.AAA.**TTG**.TTT. **ATG**.GAT), and also shared for TGB3 (GTG.TAG.**TTG**.TGG.CCG.GCG.TCA.TTA.**ATG**.TCA).

Pairwise Sequence Comparison (PASC) Confirms That HelVS Is a Distinct Carlavirus Species

The sequence of HelVS-Ver was compared to other carlavirus genome sequences using the Pairwise Sequence Comparison (PASC) tool at the NCBI website (Bao et al., 2014). This analysis clearly identified HelVS as a distinct carlavirus sequence, most closely related to GLV, PhlVS, and Ligustrum virus A (LVA) (**Supplementary Figure 4**).

Phylogenetic Analysis

The near-complete genome nt sequences of HelVS-VR21 and HelVS-Ver, ButMV-A and ButMV-B, plus the complete deduced aa sequences of their RdRp and CP proteins were used to generate Maximum Likelihood trees together with the equivalent sequences of multiple other ICTV-recognized carlavirus species.

Each of these trees showed broadly similar topology, with the two HelVS isolates grouped in a clade including PhlVB, PhlVS, GLV, CVB, and Atractylodes mottle virus (AtrMoV), with 100% bootstrap support; LVA and daphne virus S were also within the clade for the RdRp and CP trees, with 100% bootstrap support (**Figures 5A–C**).

BLASTn searches with the HelVS-VR21 genome yielded high identity to the HelVS CP and NABP (NC_038323), and the HelVS partial replicase/TGB1 (FJ555524); the genome of GLV showed 70.60% nt identity, but only 53% coverage, while all matches to other viruses had low identities and coverage. BLASTp searches of the HelVS-VR21 RdRp yielded no more than 55.93% aa identity to multiple carlaviruses, while BLASTp analysis of the HeVS-VR21 CP identified 96.99% aa identity to HelVS CP (YP_009505624); <72% aa identity (83% coverage) to red clover carlavirus A (AUF71594); and <66% aa identity to other carlavirus CP.

TABLE 5 | Percentage of nucleotide (nt) and amino acid (aa) identities between HelVS isolates V (HelVS-VR21) and Ver (HelVS-Ver).

	HelVS-VR21							
Region	**Genome nt**	**RdRp aa**	**TGB1 aa**	**TGB2 aa**	**TGB3 aa**	**CP aa**	**NABP aa**	**3′ UTR nt**
HelVS-Ver	98.58	98.81	99.13	99.07	100	99.0	100	98.67

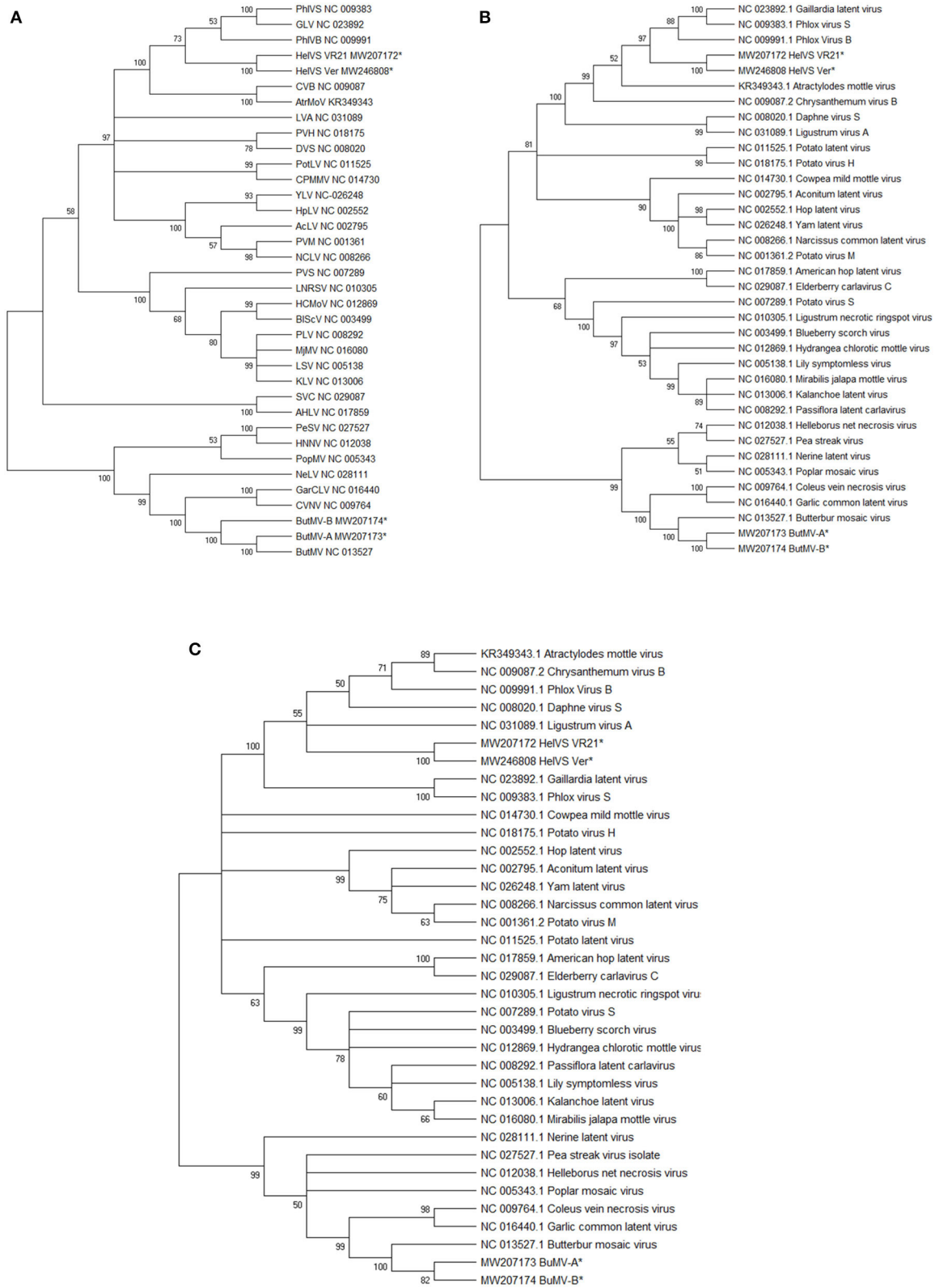


FIGURE 5 | (A–C) Maximum Likelihood phylogenetic trees with 1,000 bootstrap replications of: **(A)**, the genomic nucleotide sequences; **(B)** the Replicase (RdRp) amino acid sequences; and **(C)** the Coat protein (CP) amino acid sequences. Branches with <50% bootstrap support were collapsed. The GenBank accession

(Continued)

FIGURE 5 | number for the genomes are listed following the virus acronym in **(A)**; the RdRp and CP sequences were taken from the genome accessions. Virus abbreviations: AcLV, Aconitum latent virus; AHLV, American hop latent virus; AtrMoV, Atractylodes mottle virus; BlScV, blueberry scorch virus; ButMV, butterbur mosaic virus; CPMMV, cowpea mild mottle virus; CVB, chrysanthemum virus B; CVNV, coleus vein necrosis virus; DVS, Daphne virus S; GarCLV, garlic common latent virus; GLV, Gaillardia latent virus; HCMoV, hydrangea chlorotic mottle virus; HelVS, Helenium virus S; HNNV, Helleborus net necrosis virus; HpLV, hop latent virus; KLV, kalanchoe latent virus; LNRSV, Ligustrum necrotic ringspot virus; LSV, lily symptomless virus; LVA, Ligustrum virus A; MjMV, Mirabilis jalapa mottle virus; NCLV, narcissus common latent virus; NeLV, Nerine latent virus; PeSV, pea streak virus; PhlVB, phlox virus B; PhlVS, phlox virus S; PLV, Passiflora latent virus; PopMV, poplar mosaic virus; PotLV, potato latent virus; PVH, potato virus H; PVM, potato virus M; PVS, potato virus S; SVC, Sambucus virus C; YLV, Yam latent virus. The next-generation sequences of both isolates of HelVS, and both isolates of ButMV from the current studies are included in each tree, and each is marked with an asterisk (*).

Pairwise alignments of HelVS-VR21 to the genomes of the viruses appearing in the same clade in the phylogenetic tree revealed decreasing levels of nt identity to: PhlVS (62.32%); GLV (62.04%); LVA (60.62%); PhlVB (60.35%); AtrMoV (60.06%); and CVB (59.56%). These values are slightly higher than the BLAST-based alignments used by PASC (Bao et al., 2014; **Supplementary Figure 4**).

The ButMV-A and ButMV-B sequences were placed in all phylogenetic trees in a subclade with the type isolate of ButMV (ButMV-J), and also coleus vein necrosis virus (CVNV) and garlic common latent virus (GarCLV), with either 99 or 100% bootstrap support. An additional four viruses formed a larger clade including ButMV in all of the trees, also with either 99% or 100% bootstrap support (**Figures 5A–C**).

BLASTn analysis of the ButMV-A genome revealed no identities of >70% over the full genome, except for 78.67% nt identity to ButMV-J (AB517596) with 96% coverage, and 81.81% identity to the CP gene of ButMV-Kr-NS5 (MK689358). BLASTp analysis of ButMV-A RdRp showed 100% coverage at 85.93% aa identity to ButMV-J, and no more that 46.17% aa identity to any other viral RdRp. Similar analysis with ButMV-A CP identified 91.22% aa identity with full coverage of the ButMV-J CP.

DISCUSSION

We report here the identification and near-complete genome sequences of one isolate of HelVS, and two distinct isolates of ButMV, from a mixed infection of an unknown cultivar of *Veronica* sp. (possibly an interspecific hybrid), both first reports for this host genus. We additionally report the sequence of a second isolate of HelVS, from a single infection of a different cultivar of veronica, sold as *Veronica spicata*, or Spike Speedwell.

HelVS was found for the first time in *Veronica* sp. (family Plantaginaceae). This virus has a restricted natural host range and experimental host range, previously reported only from *Helenium amarum* (Asteraceae) in Germany and *Impatiens holstii* (Balsaminaceae) in Minnesota (USA), Yugoslavia, and the Netherlands (Kuschki et al., 1978; Koenig et al., 1983; Pleše et al., 1988). All of the natural hosts reported to date are ornamental plants, suggesting aphid-borne transmission within greenhouses as the means of host range expansion. The mild mottle or mosaic symptoms observed in veronica plants infected solely with HelVS might not be noticed by many nursery growers, which might explain detection in three additional cultivars of *Veronica* sp. by a scientist trained in viruses affecting ornamentals.

We have generated the first near-complete genome sequences of HeLVS, from two distinct sources, from commercial nurseries in Minnesota and Maryland but having a high degree of sequence identity; these are the first HelVS genome sequence beyond 3'-proximal sequences of c.1685 nt from the Helenium isolate (D10454 and S71594), and 935 nt encompassing the 3′-end of the RdRp, and part of TGB1 (FJ555524). The majority of the NGS sequence of HelVS-Ver was further validated by PCR cloning and Sanger sequencing (8,457 of 8,615 nt; **Table 2**), while 4,138 nt of the 8,668 nt HelVS-V (VR21) was also separately derived by PCR cloning and/or Sanger sequencing, with >99.7% pairwise identity for each NGS/PCR-derived pair.

The established ICTV criteria for distinguishing species with the family *Betaflexiviridae* are that isolates of different species should have less than about 72% nt identity (or 80% aa identity) between their respective CP or polymerase genes. Viruses from different genera usually have less than about 45% nt identity in these genes (ICTV 9th Report, 2011). Phylogenetic trees generated by the Maximum Likelihood (ML) method further demonstrated the close identities between HelVS-V and HelVS-Ver, grouping them into a clade with the same group of other carlavirus species, with high bootstrap support, whether trees were based on the whole genome nt sequence, or the RdRp or CP aa sequences. The highest pairwise genome identity to HeLVS-VR was 62.32% to PhlVS, fully conforming with the ICTV criteria for separating carlavirus species.

PASC analysis (Bao et al., 2014) (**Supplementary Figure 4**) further confirmed the position of HelVS as a distinct carlavirus sequence, identifying the closest relatives as GLV, PhlVS, and LVA, all of which fell within the same clade as HelVS in all of the phylogenetic trees (**Figures 5A–C**).

ButMV is also reported for the first time in Veronica, and for the first time in the USA; ButMV has previously been reported only from *Petasites japonicus* (Asteraceae), originally in Japan (Tochihara and Tamura, 1976; Hashimoto et al., 2009), and more recently in Korea (MK689358). It is interesting that both ButMV and HelVS were first reported in species of the Asteraceae, and are now found in mixed infection in a plant belonging to the taxonomically distant Plantaginaceae. Veronica is only the second natural host of ButMV reported, and only the third natural host of HelVS.

The near-complete genome sequences of two distinct isolates were identified from the same plant; isolates ButMV-A and ButMV-B were divergent from each other, and from the type isolate, ButMV-J, but still shared a sufficient degree of identity in their full genomes, and the RdRp and CP amino acid sequences,

to clearly be identified as the same carlavirus species as the type isolate, ButMV-J.

Whereas isolate ButMV-A appeared typical of carlaviruses, having the expected complement of six ORFs, encoding RdRp, TGB1, TGB2, TGB3, CP, and NABP, isolate ButMV-B had a major deletion encompassing almost the entire TGB1 ORF; TGB1 is thought to be a significant contributor toward viral movement and replication efficiency, and thus an essential gene. Absence of a functional TGB1 from ButMV-B was confirmed by RT-PCR using different combinations of degenerate primers amplifying both ButMV-A and ButMV-B, and primers specific for ButMV-A or ButMV-B, demonstrating the presence of both types of sequence in the original RNA extract. This was further confirmed by absence of most of the TGB1 ORF in the NGS sequence of ButMV-B. Interestingly, PCR from the original sample using a degenerate primer pair able to amplify across the TGB1 region of ButMV-A resulted in bands of similar intensity for the anticipated products from ButMV-A and ButMV-B, indicating that replication of ButMV-B was not considerably less efficient than for ButMV-A. The relative numbers of reads mapping to ButMV-A (1,561,852) and ButMV-B (1,785,592) sequences obtained from the NGS results from the same RNA sample extract were also quite similar, further suggesting that replication of ButMV-B was not severely debilitated, but comparable to that of ButMV-A.

That ButMV-B is successfully supported by ButMV-A over the long term, rather than being a defective replicon of recent origin, is suggested by the fact that the isolation of the original random PCR RdRp clones and the distinct 3′-terminal ButMV-A and ButMV-B clones were prepared more than 2 years apart, in each case from RNA isolated from fresh tissue. The generation of PCR products across the region of the deletion occurred several months later, so the defective ButMV-B sequence was stably maintained in the presence of ButMV-A for at least 2 years in an actively growing plant.

This appears to be the first case of a defective isolate of any carlavirus, apparently maintained long-term in Veronica as a result of complementation of the missing TGB1 functions by the fully-functional genome of ButMV-A. The essential nature of TGB1 has been demonstrated with the potexviruses white clover mosaic virus (WClMV) and potato virus X (PVX). Lough et al. (1998) showed that an infectious clone of WClMV with a deletion of a portion of TGB1 was still able to replicate in protoplasts, but was not able to spread in whole plants; however, the TGB1 deletion mutant was able to spread in transgenic plants expressing a wild-type TGB1, as a result of functional complementation. Separately, Bayne et al. (2005) generated multiple random mutants in the PVX TGB1, and showed that any mutations defective for RNA silencing suppression were also non-functional for cell-to-cell movement; a number of mutants were also recovered which retained significant RNA silencing suppression activity, but which still lacked movement capability, thus demonstrating that the RNA silencing suppression function is necessary, but not sufficient for cell-to-cell movement.

Distinct isolates of a third potexvirus, Alternanthera mosaic virus, have been shown to co-survive over multiple years and many mechanical transmission passages, interacting to produce an intermediate symptom between the most severe and mildest symptoms induced when individual sequence types were cloned and separated (Lim et al., 2010). This phenomenon was recreated by mixing infectious clones of different symptom severity; one possible advantage of the milder symptoms of the mixed infection might be a lesser effect on the host plant while still allowing high replication levels (Lim et al., 2010).

Whereas there may be a similar advantage to the presumably fully functional ButMV-A, and is clearly an advantage to the defective isolate ButMV-B, the question of the mechanism of origin of ButMV-B is less clear. One possibility is of illegitimate recombination occurring within a formerly wild-type isolate of ButMV resulted in the deletion of essentially the full TGB1 gene, with survival of the mutant due to a mixed infection with ButMV-A (or at least a progenitor sequence). However, it is interesting that the sequence of ButMV-B to the 5′ of the deletion, and the remaining sequence to the 3′ side of the deletion, share different degrees of identity to both ButMV-A and ButMV-J (c.71-72% nt identity for the RdRp region, vs. c.77–78% for the region from TGB2 to the 3′ end) (**Table 3**). This suggests an alternate possible origin in an illegitimate recombination between two isolates, each distinct from the "helper" isolate ButMV-A, possibly followed by an internal deletion. Either possible explanation suggests that mixed infection of two distinct ButMV isolates is not a recent phenomenon, but has existed for quite some time. Whether such a mixed infection initially occurred in Veronica, in butterbur, or a third and as yet unknown host, and spread to Veronica by aphid transmission is an open question. It is also possible that the presence of a HelVS infection in the same plant also contributes to maintenance of the defective ButMV-B isolate. Certainly, mixed infections and sympatric speciation of multiple carlaviruses have been documented in elderberry (Ho et al., 2016), and other hosts infected by two or more distinct carlaviruses have also been reported (e.g., Van Dijk, 1993; Eastwell and Druffel, 2012; Richert-Pöggeler et al., 2015), suggesting a fertile area for further investigations of interactions between carlavirus species.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository and accession and accession numbers can be found at:
https://www.ncbi.nlm.nih.gov/genbank/, MW207172;
https://www.ncbi.nlm.nih.gov/genbank/, MW207173;
https://www.ncbi.nlm.nih.gov/genbank/, MW207174;
https://www.ncbi.nlm.nih.gov/genbank/, MW225999;
https://www.ncbi.nlm.nih.gov/genbank/, MW226000;
https://www.ncbi.nlm.nih.gov/genbank/, MW226001;
https://www.ncbi.nlm.nih.gov/genbank/, MW226002;
https://www.ncbi.nlm.nih.gov/genbank/, MW226003;
https://www.ncbi.nlm.nih.gov/genbank/, MW226004;
https://www.ncbi.nlm.nih.gov/genbank/, MW246808;
https://www.ncbi.nlm.nih.gov/genbank/, MW246809;
https://www.ncbi.nlm.nih.gov/genbank/, MW246810;
https://www.ncbi.nlm.nih.gov/genbank/, MW246811;
https://www.ncbi.nlm.nih.gov/genbank/, MW246812.

AUTHOR CONTRIBUTIONS

JH and DM developed and directed the study. JH, DM, and RJ supervised the next-generation sequencing. MR, RJ, and SG performed the NGS analyses. BL and DM performed the initial extraction and random PCR cloning. MR performed most of the PCR and cloning. JH performed additional sequence analyses and wrote the first draft, with assistance from RJ and DM and input from MR and SG. All authors assisted in revisions and approved the manuscript for publication.

FUNDING

This work was supported by internal USDA-ARS funding under projects 8020-22000-042-00D and 8020-22000-302-00D.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.612936/full#supplementary-material

Supplementary Figure 1 | Graphical representation of a carlavirus genome and the encoded proteins, with approximate sizes and positions of PCR products derived from the initial 3′-proximal regions, and the random-PCR products from the RdRp region, aligned and colored to indicate the virus from which each originates, with ButMV-related products in blue, and HelVS-related products in red. The approximate overlaps of RdRp-d and RdRp-b, and of RdRp-b and HelVS-I (GenBank acc. no. FJ555524) are also shown. The 3′-proximal PCR products and associated random PCR RdRp products are shown in line with each other (see also **Table 1**).

Supplementary Figure 2 | Genome coverage of the Helenium virus S genome by the next-generation sequencing reads, which averaged 14,493 reads per nucleotide.

Supplementary Figure 3 | Graphical illustration of the strategy for PCR amplification and Sanger sequencing of a series of overlapping PCR product covering 8,457 nt of the 8,615 nt HelVS-Ver genome obtained by next-generation sequencing.

Supplementary Figure 4 | Pairwise sequence comparison (PASC) (Bao et al., 2014) result from analysis at the NCBI website, showing that HelVS is distinct from other characterized carlavirus species.

Supplementary Table 1 | Primers utilized in this study; unless otherwise annoted, all primers were designed for this study.

REFERENCES

Allen, T. C. (1975). Viruses on lilies and their control. *Acta Hortic*. 47, 69–75. doi: 10.17660/ActaHortic.1975.47.11

Bao, Y., 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

Bayne, E. H., Rakitina, D. V., Morozov, S. Y., and Baulcombe, D. C. (2005). Cell-to-cell movement of *Potato Potexvirus X* is dependent on suppression of RNA silencing. *Plant J*. 44, 471–482. doi: 10.1111/j.1365-313X.2005.02539.x

Brierley, P., and Smith, F. F. (1944). Studies on lily virus diseases: the necrotic-fleck complex in *Lilium longiflorum*. *Phytopathology* 34, 529–555.

Eastwell, K. C., and Druffel, K. L. (2012). Complete genome organization of American hop latent virus and its relationship to carlaviruses. *Arch. Virol.* 157, 1403–1406. doi: 10.1007/s00705-012-1312-0

Fletcher, J. D. (1989). Additional hosts of *Alfalfa mosaic* virus, *Cucumber mosaic* virus and *Tobacco mosaic* virus in New Zealand. *N. Z. J. Crop Hortic. Sci.* 17, 361–362. doi: 10.1080/01140671.1989.10428057

Foster, G. D., Millar, A. W., Meehan, B. M., and Mills, P. R. (1990). Nucleotide sequence of the 3'-terminal region of Helenium virus. S RNA. *J. Gen. Virol*. 71, 1877–1880. doi: 10.1099/0022-1317-71-8-1877

Froussard, P. (1992). A random PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucl. Acids Res*. 20:2900. doi: 10.1093/nar/20.11.2900

Hammond, J., and Reinsel, M. (2011). Mixed infections and novel viruses in various species of Phlox. *Acta Hort*. 901, 119–126. doi: 10.17660/ActaHortic.2011.901.15

Hammond, J., Reinsel, M. D., and Maroon-Lango, C. J. (2006). Identification and full sequence of an isolate of Alternanthera mosaic potexvirus infecting Phlox stolonifera. *Arch. Virol*. 151, 477–493. doi: 10.1007/s00705-005-0646-2

Hashimoto, M., Komatsu, K., Maejima, K., Yamaji, Y., Okano, Y., Shiraishi, T., et al. (2009). Complete nucleotide sequence and genome organization of butterbur mosaic virus. *Arch. Virol*. 154, 1955–1958. doi: 10.1007/s00705-009-0529-z

Hausbeck, M. K., Welliver, R. A., Derr, M. A., and Gildow, F. E. (1992). Tomato spotted wilt virus survey among greenhouse ornamentals in Pennsylvania. *Plant Dis*. 76, 795–800. doi: 10.1094/PD-76-0795

Henderson, D. C., and Hammond, J. (2013). CKC: Isolation of nucleic acids from a diversity of plants using CTAB and silica columns. *Mol. Biotechnol*. 53, 109–117. doi: 10.1007/s12033-012-9494-y

Ho, T., Quito-Avila, D., Keller, K. E., Postman, J. D., Martin, R. R., and Tzanetakis, I. E. (2016). Evidence of sympatric speciation of elderberry carlaviruses. *Virus Res*. 215, 72–75. doi: 10.1016/j.virusres.2016.01.017

ICTV 9th Report (2011). Available online at: https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/241/betaflexiviridae (accessed December 2, 2020).

Koenig, R., Lesemann, D.-E., Lockhart, B., Betzold, J. A., and Weidemann, H. L. (1983), Natural occurrence of *Helenium* virus S in *Impatiens holstii*. *J. Phytopathol*. 106, 133–140. doi: 10.1111/j.1439-0434.1983.tb00036.x

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

Kuschki, G. H., Koenig, R., Düvel, D., and Kühne, H. (1978). Helenium virus S and Y – two new viruses from commercially grown *Helenium hybrids*. *Phytopathology* 68:1407. doi: 10.1094/Phyto-68-1407

Lim, H. S., Vaira, A. M., Reinsel, M. D., Bae, H., Bailey, B. A., Domier, L. L., et al. (2010). Pathogenicity of Alternanthera mosaic virus is affected by determinants in RNA-dependent RNA polymerase and by reduced efficacy of silencing suppression in a movement-competent TGB1. *J. Genrl. Virol.* 91, 277–287. doi: 10.1099/vir.0.014977-0

Lough, T. J., Shash, K., Xoconostle-Cázares, B., Hofstra, K. R., Beck, D. L., Balmori, E., et al. (1998). Molecular dissection of the mechanism by which potexvirus triple gene block proteins mediate cell-to-cell transport of infectious RNA. *Mol. Plant Microbe Inter.* 11, 801–814. doi: 10.1094/MPMI.1998.11.8.801

Massart, S., Chiumenti, M., De Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., et al. (2019). Virus detection by high-throughput sequencing of small RNAs: large-scale performance testing of sequence analysis strategies. *Phytopathology* 109, 488–497. doi: 10.1094/PHYTO-02-18-0067-R

Matthews, R. E. F. (1982). Classification and nomenclature of viruses. fourth report of the international committee on nomenclature of viruses. *Intervirology* 17, 1–199. doi: 10.1159/000149278

Milusheva, S., and Rankova, Z. (2002). Plum pox potyvirus detection in weed species under field conditions. *Acta Hortic*. 577, 283–287. doi: 10.17660/ActaHortic.2002.577.48

Mollov, D., Lockhart, B., and Zlesak, D. (2013). Complete nucleotide sequence of rose yellow mosaic virus, a novel member of the family *Potyviridae*. *Arch. Virol.* 158, 1917–1923. doi: 10.1007/s00705-013-1686-7

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

Pleše, N., Erić, Ž., and Krajačić, M. (1988). Further information on infection of *Impatiens holstii* with Helenium virus S. *Acta Hortic*. 234, 477–484. doi: 10.17660/ActaHortic.1988.234.58

Richert-Pöggeler, K. R., Turhal, A.-K., Schuhmann, S., Maaß, C., Blockus, S., Zimmermann, E., et al. (2015). C*arlavirus* biodiversity in horticultural host plants: efficient virus detection and identification combining electron microscopy and molecular biology tools. *Acta Hortic*. 1072, 37–45. doi: 10.17660/ActaHortic.2015.1072.3

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol*. 10, 512–526.

Thomas, P. R. (1970). Host status of some plants for Xiphinema diversicaudatum (Micol.) and their susceptibility to viruses transmitted by this species. *Ann. Appl. Biol*. 65, 169–178. doi: 10.1111/j.1744-7348.1970.tb04575.x

Tochihara, H., and Tamura, M. (1976). Viruses in Japanese butterbur (*Petasites japonicum* Miq.). *Ann. Phytopath. Soc. Jpn.* 42, 533–539. doi: 10.3186/jjphytopath.42.533

Turner, R. L., Mills, P. R., and Foster, G. D. (1993). Nucleotide sequence of the. 7 K gene of Helenium virus S. *Acta Virol*. 37, 523–528.

Valverde, R. A., Sabanadzovic, S., and Hammond, J. (2012). Viruses that enhance the aesthetics of some ornamental plants: beauty or beast? *Plant Dis*. 96, 600–611. doi: 10.1094/PDIS-11-11-0928-FE

Van Dijk, P. (1993). Carlavirus isolates from cultivated Allium species represent three viruses. *Netherlands J. Plant Pathol.* 99, 233–257. doi: 10.1007/BF01974306

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109, 716–725. doi: 10.1094/PHYTO-07-18-0257-RVW

Wetter, C., and Milne, R. G. (1981). "Chapter 22: Carlaviruses," in *Handbook of Plant Virus Infections*, ed E. Kurstak (Amsterdam: Elsevier/North-Holland Biomedical Press), 695–730.

Zitter, T. A. (2001). *Vegetable Crops: A Checklist of Major Weeds and Crops as Natural Hosts for Plant Viruses in the Northeast*. Available onine at: http://vegetablemdonline.ppath.cornell.edu/Tables/WeedHostTable.html (accessed December 2, 2020).

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.

Copyright © 2020 Hammond, Reinsel, Grinstead, Lockhart, Jordan and Mollov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fmicb.2020.621179

Next-Generation Sequencing Reveals a Novel Emaravirus in Diseased Maple Trees From a German Urban Forest

Artemis Rumbou[1], Thierry Candresse[2], Susanne von Bargen[1] and Carmen Büttner[1]*

[1] Faculty of Life Sciences, Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences, Humboldt-Universität zu Berlin, Berlin, Germany, [2] UMR 1332 Biologie du Fruit et Pathologie, INRAE, University of Bordeaux, UMR BFP, Villenave-d'Ornon, France

While the focus of plant virology has been mainly on horticultural and field crops as well as fruit trees, little information is available on viruses that infect forest trees. Utilization of next-generation sequencing (NGS) methodologies has revealed a significant number of viruses in forest trees and urban parks. In the present study, the full-length genome of a novel *Emaravirus* has been identified and characterized from sycamore maple (*Acer pseudoplatanus*) – a tree species of significant importance in urban and forest areas – showing leaf mottle symptoms. RNA-Seq was performed on the Illumina HiSeq2500 system using RNA preparations from a symptomatic and a symptomless maple tree. The sequence assembly and analysis revealed the presence of six genomic RNA segments in the symptomatic sample (RNA1: 7,074 nt-long encoding the viral replicase; RNA2: 2,289 nt-long encoding the glycoprotein precursor; RNA3: 1,525 nt-long encoding the nucleocapsid protein; RNA4: 1,533 nt-long encoding the putative movement protein; RNA5: 1,825 nt-long encoding a hypothetical protein P5; RNA6: 1,179 nt-long encoding a hypothetical protein P6). Two independent NGS sequencing runs from the same symptomatic maple tree detected the same genome segments. For one of these sequencing runs the cDNA library was prepared using a primer targeting the conserved genome terminal region, known to be shared between emaraviruses genome segments. We suggest, therefore, that the six identified genome segments represent the complete genome of a novel emaravirus from maple, which we tentatively name maple mottle-associated virus (MaMaV). Phylogenetic and sequence homology analyses place this virus on the distinct "subgroup a" clade within the *Emaravirus* genus along with – among others – rose rosette virus, *Actinidia* emaravirus 2, and fig mosaic virus. Validation RT-PCR assays performed on symptomatic and non-symptomatic trees suggest that MaMaV may be the symptom-inducing virus in the diseased trees. To our knowledge, this is the first time an *Emaravirus* is described from maple and is fully genetically characterized. With the discovery of MaMaV, the genus *Emaravirus* comprising negative-sense single-stranded viruses with very divergent genomes – that were until recently overlooked – has substantially increased counting 22 established and putative members.

Keywords: emaravirus, maple mottle-associated virus, maple virome, RNA-Seq, forest disease

OPEN ACCESS

Edited by:
Ahmed Hadidi,
Agricultural Research Service, United States Department of Agriculture, United States

Reviewed by:
Beatriz Navarro,
Istituto per la Protezione Sostenibile delle Piante, Italy
Satyanarayana Tatineni,
Agricultural Research Service, United States Department of Agriculture, United States

****Correspondence:***
Artemis Rumbou
artemis.rumbou@agrar.hu-berlin.de

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

Received: *25 October 2020*
Accepted: *09 December 2020*
Published: *08 January 2021*

Citation:
Rumbou A, Candresse T, von Bargen S and Büttner C (2021) Next-Generation Sequencing Reveals a Novel Emaravirus in Diseased Maple Trees From a German Urban Forest. Front. Microbiol. 11:621179. doi: 10.3389/fmicb.2020.621179

INTRODUCTION

Viruses of forest trees have been, until recently, only slightly characterized due to two main reasons; (a) biased sampling based on a restricted focus on agricultural crops and fruit trees viruses (Büttner et al., 2013) and (b) a general bias against the identification of the most divergent genomes (Zhang et al., 2018). Due to the utilization of next generation sequencing (NGS), however, forest virology has gained a significant momentum in identifying viruses infecting forest trees. Recently, a birch virome was unraveled revealing a complex of novel and known viruses (Rumbou et al., 2020), while a novel badnavirus associated with the birch leaf-roll disease was identified and genetically characterized (Rumbou et al., 2018). In mosaic-diseased Eurasian aspen (*Populus tremula*) a novel emaravirus has been identified (von Bargen et al., 2020a). European mountain ash ringspot-associated virus has been detected in new hosts like *Karpatiosorbus* × *hybrida* in Finland (von Bargen et al., 2020b), *Sorbus intermedia* (von Bargen et al., 2019), and *Amelanchier* spp. (von Bargen et al., 2018). It is apparent, that the application of NGS tools has substantially increased the rate of virus discovery in forest and urban green ecosystems.

Viral diseases of different maple species (*Acer spp.*) have been reported by plant virologists since long (Cooper, 1979). Atanasoff (1935) was probably the first to describe a "yellow mottle-mosaic" symptom in *Acer negundo* and *A. pseudoplatanus* in Japan and Europe, possibly related to virus presence. Brierley (1944) observed chlorotic or ring mottle in *Acer saccharum* in North America, and a similar symptom was reported by Ploiaie and Macovei (1968) in Romania. Szirmai (1972) described a "mosaic mottling with chlorotic – tending to yellow ochre – spots" in *A. negundo* and *A. pseudoplatanus* in Hungary, while Cooper (1979) confirmed this symptomatology in maples in the United Kingdom. At the same time, mechanical transmission of the so-called "maple leaf perforation virus" from naturally infected maples to beans (*Phaseolus vulgaris*) was described by Subikova (1973), but without visualizing virus particles by electron microscopy (EM).

The graft-transmissibility of an agent causing leaf mottle in maples was earlier reported (Führling and Büttner, 1998), while rod-shaped virus particles approx. 300 nm-long and genomic material from tobamoviruses were observed by EM in symptomatic leaves. Rod-shaped virus particles were also detected in young *A. saccharum* seedlings with chlorotic spots and mottle symptoms (Lana et al., 1980), which were attributed to tobacco mosaic virus based on their serological and biological properties. Isometric particles of 26–30 nm diameter were detected in maple trees from Turkish urban areas (*A. negundo, A. pseudoplatanus, Acer campestre*) exhibiting mottling, mosaic, leaf deformation and lateral shoot formation (Erdiller, 1986). These particles were attributed to arabis mosaic virus, cucumber mosaic virus and soybean mosaic virus. However, to date, *Acer sp.* has never been unambiguously described as a host for any well-characterized viral agent and data of the Virus-Host database also confirm that (Mihara et al., 2016).

Maples are abundantly found in European forests and urban parks, with the most common species *A. pseudoplatanus* (sycamore) and *Acer platanoides* (Norway maple) representing a natural component of birch (*Betula* sp.) and fir (*Abies* sp.) forests (Gibbs and Chen, 2009). Several *Acer* species provide valuable timber and are the main sources of maple sugar and maple syrup (Binggeli, 1993). Damages on the trees have been regularly attributed to fungi with most harmful being *Verticillium* wilt, sooty bark disease caused by *Cryptostroma* species (Wulf and Kehr, 2009) as well as *Eutypella parasitica* causing trunk infections (Brglez et al., 2020). Phytoplasma-associated witches-broom disease has been reported in *A. negundo* (Kaminska and Suwa, 2006) as well as in Japanese maple, *Acer palmatum* (Li et al., 2012). Maple trees exhibiting virus-like symptoms were regularly observed in forests around Berlin as well as in other regions of Germany during the last 40 years (Bandte et al., 2008). However, viral agent(s) affecting maples have not been adequately characterized by conventional methods. Considering the ecological and economical importance of maples, we aimed to fill in the gap in maple's pathology by employing an RNA-Seq methodology to identify viruses possibly affecting maple trees. As a result, the genome of a novel viral agent is fully identified and characterized, while its association with the observed symptomatology is strongly suggested.

MATERIALS AND METHODS

Plant Materials

In 2014, leaf samples exhibiting virus-like symptoms, including mottle and leaf deformation (**Figure 1**), were collected from an *A. pseudoplatanus* tree in the Berlin-Grunewald urban forest [Acer+ (2014): symptomatic tree], where such symptoms have been monitored for at least two decades. Randomly selected leaf parts were pooled together and used for RNA extraction. A similar pool of leaves was obtained from a symptomless seedling [Acer− (2014): non-symptomatic tree]. In 2015, the same symptomatic tree was re-sampled and RNA was extracted from pooled leaves exhibiting symptoms [Acer+ (2015)].

For the investigation of the virus presence in the urban Berlin-Grunewald forest, leaves were collected from symptomatic and non-symptomatic maples in order to be used for RNA isolation and RT-PCR diagnostic assays. In total, 26 sycamore maple trees exhibiting symptoms similar to those of the Acer+ (2014) tree as well as six trees without symptoms were tested in 2015, 2016 or 2019 (**Supplementary Table 1**).

Next Generation Sequencing and Sequence Assembly

Total RNAs were isolated from 100 mg leaf tissue using the InviTrap Spin Plant RNA Mini Kit (STRATEC Molecular, Germany), followed by removal of remaining DNA with rDNase according to the supplier protocol (Macherey-Nagel, Germany) and RNA purification using NucleoSpin RNA Clean-up (Macherey-Nagel, Germany). Ribosomal RNA depletion was performed using the RiboMinus Plant Kit for RNA-Seq (Invitrogen). One to two micrograms of RiboMinus RNA of each sample were used for cDNA synthesis with the Maxima H Minus double-stranded cDNA synthesis Kit (Thermo Scientific)

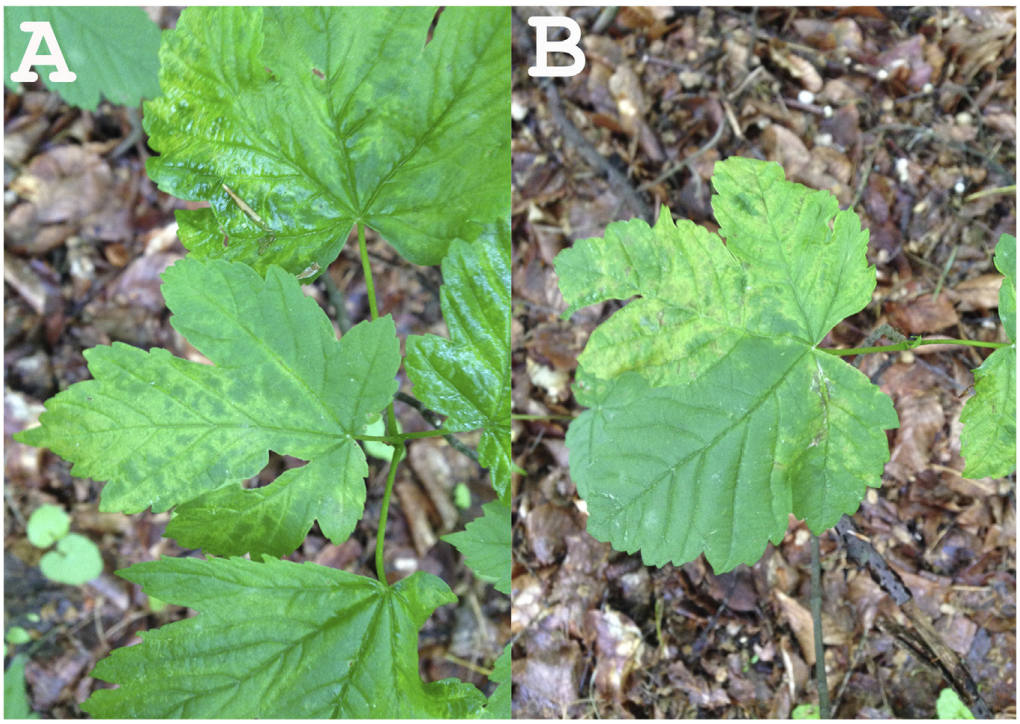


FIGURE 1 | Leaf symptoms of the tested sycamore tree subjected to NGS (**A:** leaf mottle; **B:** leaf mottle and deformation).

primed with random hexamers for samples Acer+ (2014) and Acer− (2014). For sample Acer+ (2015), the precipitated RNA was reverse transcribed into cDNA using the generic terminal primer PDAP213 (Di Bello and Tzanetakis, 2013).

Two micrograms purified double-stranded cDNA from each sample were sent to BaseClear (Netherlands) for RNA-Seq analysis on the Illumina HiSeq2500 system. FASTQ sequence files were generated using the Illumina Casava pipeline version 1.8.3. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing adapters and/or PhiX control were removed using an in-house filtering protocol. The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.10.0. NGS data processing and analysis were carried out either using CLC Genomics Workbench version 7.0.4. or the VirAnnot pipeline (Lefebvre et al., 2019). The quality of the FASTQ sequences was enhanced by trimming off low quality bases using the "Trim sequences" option of the CLC Genomics Workbench version 7.0.4. The quality-filtered reads were *de novo* assembled into contig sequences using CLC Genomics Workbench. Contigs annotation was carried out using BLASTn against the NCBI-GenBank databases. When needed, virus-related contigs were manually assembled into larger scaffolds and scaffolds polished by re-mapping of reads on the scaffolds using CLC Genomics Workbench.

Taxonomic Analysis of the Metagenome

The taxonomic content of the obtained datasets, as provided by the BLAST annotations, was visualized using MEGAN (Huson et al., 2016), in which the BLAST results were parsed to assign the best hits to appropriate taxa in the NCBI taxonomy. As a result, the taxonomical content ("species profile") of the sample from which the reads were collected was estimated, with a particular focus on viral species.

Validation of the Presence of Novel Virus in Maples

Samples from 32 symptomatic and non-symptomatic maple trees were collected from two different locations in Berlin-Grunewald and from grafted scions of that origin propagated in the experimental garden of Humboldt University of Berlin since 2017 (**Supplementary Table 1**). Symptomatic trees exhibited most often mottle, or mottle in combination with a mixture of other symptoms like flecking, chlorotic ringspots, vein banding, chlorotic line pattern, mosaic and/or leaf deformation. Pooled samples of 100 mg leaf tissue from three to five leaves from different twigs of each tree were used. Total RNAs were isolated according to Boom et al. (1990).

To confirm the presence of the identified segments in the samples, specific RT-PCR assays were performed using segment-specific primer pairs (**Table 1**). The first-strand cDNAs were synthesized from 1 μg of total RNA in a 20 μl reaction volume of 1x RT buffer (Thermo Scientific) containing 1 μM dNTPs mix, 100 U RevertAid Premium reverse transcriptase (Thermo Scientific), 20 U RiboLock RNase inhibitor (Thermo Scientific), and 100 pmol of random hexamer-oligonucleotides

TABLE 1 | Features of primers used for the specific RT-PCR detection of MaMaV.

Primer name	Primer sequence (5′ – 3′)	Product length (bp)
RNA1aF	AACCAATGCTGTCACTTAAGC	346
RNA1aR	GATATAACTACCATCTAACATCC	
RNA1bF	GGATGTTAGATGGTAGTTATATC	290
RNA1bR	CCACTTATAGTTATTGCTTCACC	
RNA2F	GCAAGATTTTGATGTGGCTGG	149
RNA2R	AACCATCATGGCCATCACAAC	
RNA3F	TGTGCTATAATGGCAGCTGG	289
RNA3R	CATCAGTCATGCTATCTGGTATG	
RNA4F	TTGGACACCAACATCTACAAG	470
RNA4R	GCAATTCCTTCCTCTCATTGT	
RNA5F	GAACTATGTCTTACCAACACTG	221
RNA5R	CTAATTCCCTAAGTTTGATAGTAAC	
RNA6F	CAGATAACATATTCTCTTCTGG	300
RNA6R	AAGCGAGATATATGCTATGGCT	

(Biolegio). Subsequent PCR amplifications were conducted in a 50 μl volume of 1x DreamTaq Buffer (Thermo Scientific) containing 0.2 μM dNTP mix, 0.25 U of DreamTaq DNA polymerase and 1 μM of each forward and reverse primer. The thermal cycles were as follows: 2 min at 94°C followed by 35 cycles at 94°C for 30 s, 55°C for 30 s, 72°C for 30 s, with a final extension step of 72°C for 5 min. The product lengths amplified for the different RNA segments are shown in **Table 1**.

Sequence Analysis and Phylogenetic Comparison of Novel Sequences

Multiple nucleotide or amino acid sequence alignments as well as pairwise sequence identity calculations were performed using AliView version 1.17.1 (Larsson, 2015). ORF finder at NCBI[1] was used to identify open reading frames (ORFs) on assembled genome segments and identify the encoded proteins. All ORFs with 300 or more nucleotides (nt) were considered. For the phylogenetic comparisons of complete coding regions, the 21 established and tentative emaravirus species identified to date and represented in GenBank were used. Maximum likelihood (ML) trees were constructed with MEGA6 (Tamura et al., 2013) applying the Jones-Taylor-Thornton (JTT) substitution model for amino acids. Robustness of nodes of the phylogenetic tree was assessed from 1,000 bootstrap replications and values >70% were displayed for trees' internal nodes.

To perform preliminary genetic divergence analysis, 40 RT-PCR products from 10 tested samples were directly submitted for Sanger sequencing (Macrogen) without previous cloning. They were amplified from three RNA segments; RNA1, RNA3, and RNA4. For RNA1, RT-PCR products were amplified in two different genome regions; the primer-pair RNA1aF/R generated PCR products of 311 nt length (nt positions 3,006 – 2,756), while the primer pair RNA1bF/R generated 274 nt-long RT-PCR products located at nt positions 3,317 – 3,042. Sequences of RT-PCR products for the RNA3 segment were 274 nt-long (nt positions 734 – 468), while in the RNA4 segment RT-PCR products of 431 nt length were amplified (nt positions 835 – 405). Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013). Genetic distance between sequences was assessed applying the Maximum Composite Likelihood model (Tamura et al., 2004), where the number of base substitutions per site between sequences was calculated.

[1]https://www.ncbi.nlm.nih.gov/orffinder/

RESULTS

Quality Analysis of FASTQ Sequence Reads, *de novo* Assembly and Taxonomic Analysis of the Metagenome

RNA-Seq was performed in 2014 using RNA preparations from a symptomatic and a symptomless maple tree. 124 and 14 MB data/sample with average quality of approx. 35 Phred were generated for Acer+ (2014) and Acer− (2014), respectively [620,460 FASTQ reads for Acer+ (2014); 69,353 FASTQ reads for Acer− (2014)] (**Table 2**). For the sample Acer+ (2014) the *de novo* assembly of quality-filtered paired-end reads resulted in 532 assembled contigs. Analysis identified 14 contigs exhibiting significant nt identities to RNA sequences of several emaraviruses as assessed by BLASTn (seven contigs with identities to emaravirus-RNA1 assembling a 6,005 bp-long scaffold with missing genome parts within the sequence and at the ends; three contigs with identities to emaravirus-RNA2 assembling a 1,761 bp-long scaffold; one contig 1,085 bp-long with identities to emaravirus-RNA3; one 1,256 bp-long contig with identities to emaravirus-RNA4; one 1,277 bp-long contig with identities to RNA5 segment of some emaravirus species; one 928 bp-long contig with identities to RNA6 segment of some emaravirus species). All assembled contigs/scaffolds were missing sequences at the 3′ and 5′ ends. In the negative control sample Acer− (2014) none of the 30 contigs assembled showed any significant identities to emaraviruses or any other plant viruses. No other contigs from the Acer+ (2014) sample were identified as viral besides the 14 ones showing affinities with emaraviruses.

The contigs resulting from the *de novo* assembly of the reads for each sample were used for the MEGAN analysis. For the symptomatic sample Acer+ (2014), out of the 532 contigs assembled, 474 belong to Eucaryota, most of them to clade *Euphylophyta* (Phylum: Streptophyta; Kingdom: Viridiplantae) – where *Acer spp*. is classified – and three belong to Bacteria. The 11 viral contigs identified by MEGAN are attributed to the *Fimoviridae* family and ten are attributed to pigeon pea sterility mosaic virus (*Emaravirus*, *Fimoviridae*). In the case of the non-symptomatic sample Acer− (2014), none of the 30 assembled contigs is attributed to a viral species. The taxonomic analysis performed by MEGAN show a high degree of consistency with the results delivered by BLASTn annotation and clearly suggest the presence of an emaravirus in the tested samples.

To confirm the RNA-Seq results from the Acer+ (2014) sample and to complete the missing parts of the detected genome

TABLE 2 | Quality statistics of FASTQ sequence reads and *de novo* statistics for the three maple samples.

Plant	FastQ sequence reads	Sample Yield (in MB)	*De novo* assembly				
			Total reads	Matched reads	Total contigs	Average contig length (bp)	Emara-specific contigs
Acer+ (2014)	620,460	124	1,116,554	924,998	532	677	14
Acer– (2014)	69,353	14	123,424	115,795	30	984	0
Acer+ (2015)	850,283	198	1,700,566	1,496,745	2,206	646	8

segments, the whole NGS process was repeated in 2015 with leaves from the same tree. A ds-cDNA library was generated using the emara-specific terminal primer PDAP213, aiming to detect the missing genome ends. The RNA-Seq analysis provided a higher amount of sequence data (198 MB), delivering 850,283 FASTQ sequence reads (av. quality score: ∼35 Phred) (**Table 2**). The *de novo* assembly resulted in a total of 2,206 contigs with an average length of 646 bp. BLASTn analysis identified eight long contigs exhibiting significant identities to emaraviruses (three with identities to emaravirus-RNA1, assembling a 6,907 bp-long scaffold missing the 5′ end; one 2,289 bp-long contig with identities to emaravirus-RNA2; one 965 bp-long contig with identities to emaravirus-RNA3; one short 235 bp-long contig with identities to emaravirus-RNA4, one 1,851 bp-long contig with identities to RNA5 segment of some emaraviruses, and one 1,100 bp-long contig with identities to RNA6 segment of some emaravirus). The new contigs showed 100% nucleotide identity with the ones generated from the same tree in 2014. As such, they were assembled together with the 2014 contigs, providing the full-length sequence of the various RNA segments. Finally, a polishing step was performed to ensure genome completion and the absence of errors and to confirm genome ends. As a result, complete sequence of six genomic RNA segments were obtained (RNA1: 7,074 nt; RNA2: 2,289 nt; RNA3: 1,525 nt; RNA4: 1,533 nt, RNA5: 1,825 nt; RNA6: 1,179 nt) (**Figure 2**). The full-length genomic sequences of the maple emaravirus RNA segments are deposited in GenBank under accession numbers MT879190–MT879195.

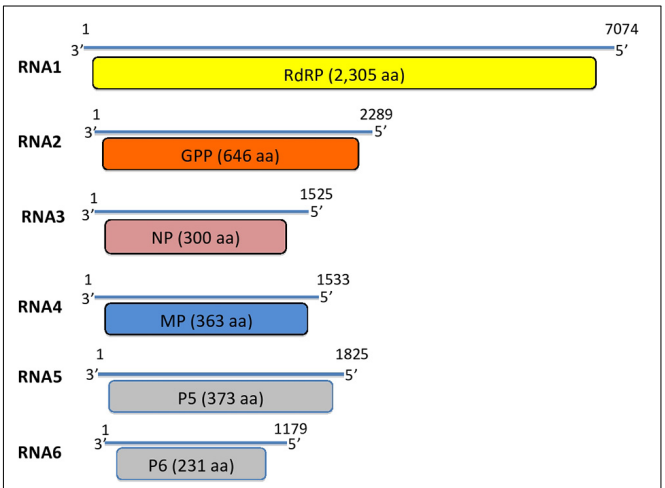


FIGURE 2 | Genome segments of the novel maple mottle-associated virus (MaMaV). RdRP, RNA-dependent RNA polymerase; GPP, glycoprotein precursor; NP, nucleocapsid protein; MP, movement protein; P5, hypothetical protein; P6, hypothetical protein. Numbers represent the nucleotide positions of the RNA segments.

Genome Structure and Encoded Proteins

To identify ORFs on assembled genome segments, ORF Finder from NCBI was employed and identified six ORFs, one corresponding to each RNA segment. BLASTp annotation of the amino acid (aa) sequences derived from the assembled scaffolds from the symptomatic maple tree revealed high BLAST scores with RNA segments from viruses of the genus *Emaravirus* (*Fimoviridae, Bunyavirales*) (**Table 3**).

ORF1 predicted on RNA1 is 2,305 aa-long and the encoded protein shows in BLASTp analysis significant aa identity of 32–75% with the viral replicase of 21 known emaraviruses (ORF1: nt positions 6,966 – 49). The putative RNA polymerase exhibits highest aa identity with that of rose rosette virus (Accession number: QHZ99251.1; 74.4% aa identity) (**Figure 3**).

ORF2 encoded on RNA2 is predicted to encode a protein of 646 aa (nt positions 1,996 – 56). In BLASTp comparisons the encoded protein shows 26–57% aa identity with 21 other emaraviruses. The highest aa identity is with the glycoprotein precursor (GPP) of rose rosette virus (Accession number: QID76023.1; aa identity: 56.7%).

ORF3 encoded on RNA3 is predicted to encode a protein of 300 aa (nt positions 104 – 1,006) with highest BLASTp aa identity with the nucleocapsid proteins (NP) of the 21 known emaraviruses (24–56%). The ORF3-encoded protein shows highest aa identity with the NP protein of *Actinidia* emaravirus 2 (Accession number: QEE82888.1; aa identity: 55.6%). In **Table 3** the recently identified NP protein from pea-associated virus (Y.Z.A Gaafar and Ziebell, 2019, H; GenBank: QJX15716.1.) is also included, because ORF3 shows relatively high identity with it (50.5%). There are, however, no other proteins identified from this tentative emaravirus.

ORF4 encoded on RNA4 is predicted to encode a protein of 363 aa (nt positions 1,179 – 88) which shows very variable BLASTp aa identity levels with the movement proteins identified from 20 emaraviruses (19–65%). It exhibits highest aa sequence identity with the movement protein of fig mosaic virus (Accession number: BAM13817.1; aa identity: 65.2%).

The fifth predicted ORF is encoded on RNA5 and is predicted to encode a protein of 373 aa (nt positions 1,526 – 93). The hypothetical protein encoded by RNA5 shows sequence

TABLE 3 | Pairwise comparison of sequence identities at the amino acid level of putative proteins (RdRP, GPP, NP, MP, P5, P6) encoded by RNA1-RNA6 of the novel *Emaravirus* from *Acer pseudoplatanus* with the homolog proteins of the established and putative members of the genus *Emaravirus*.

		RdRP	GPP	NP	MP	P5	P6
1	Rose rosette virus	**74.40%**	**56.65%**	51.77%	64.84%	**44.81%**	42.04%
		QHZ99251.1	QID76023.1	QIB97971.1	QIB98058.1	QIB98219.1	QJR96770.1
2	Aspen mosaic-associated virus	70.23%	53.11%	51.04%	57.73%	–	40.39%
		CAA0079389.1	CAA0079597.1	CAA0079646.1	CAA0079685.1		CAA0079719.1
3	*Actinidia* emaravirus 2	69.66%	47.78%	**55.63%**	63.31%	43.41%	**44.16%**
		QEE82886.1	QEE82887.1	QEE82888.1	QEE82889.1	QEE82890.1	QEE82891.1
4	Pistacia emaravirus	69.16%	52.07%	51.60%	62.33%	40.46%	32.43%
		QAR18002.1	QAR18003.1	QAR18004.1	QAR18005.1	QAR18006.1	QAR18008.1
5	Fig mosaic emaravirus	68.67%	52.15%	54.04%	**65.19%**	33.98%	41.71%
		QBH72675.1	QBK46595.1	BAM13809.1	BAM13817.1	YP_009237273.1	BAM13854.1
6	Pigeon pea sterility mosaic emaravirus 1	53.58%	46.56%	40.29%	62.64%	42.92%	37.70%
		ANQ90714.1	QBA83603.1	CDX09880.1	ANQ90777.1	CUR49054.1	ANQ90719.1
7	Pigeon pea sterility mosaic emaravirus 2	68.14%	51.32%	53.29%	64.00%	43.89%	37.70%
		QBA83607.1	YP_009268865.1	ALU34071.1	ANQ90759.1	QBA83611.1	ANQ90763.1
8	Blackberry leaf mottle-associated virus	66.09%	52.22%	50.17%	56.32%	–	35.57%
		AQX45473.1	AQX45474.1	AQX45475.1	QBM15152.1		AQX45477.1
9	European mountain ash ringspot-associated emaravirus	48.70%	40.67%	36.69%	35.24%	–	–
		VFU05375.1	YP_003104765.1	SPN63240.1	VFU05382.1		
10	Actinidia chlorotic ringspot-associated virus	47.82%	41.53%	38.43%	33.70%	–	–
		YP_009507925.1	YP_009507926.1	YP_009507928.1	YP_009507927.1		
11	Lilac chlorotic ringspot-associated virus	48.92% QIN85945.1	42.61%	40.69%	36.56%	–	–
			QIN85946.1	QIN85947.1	QIN85948.1		
12	Redbud yellow ringspot-associated emaravirus	47.27%	40.96%	37.91%	35.85%	–	–
		YP_009508083.1	YP_009508087.1	YP_009508085.1	YP_009508084.1		
13	Ti ringspot-associated emaravirus	37.26%	26.55%	30.82%	23.10%	–	–
		QAB47307.1	QAB47308.1	QAB47309.1	QAB47310.1		
14	Raspberry leaf blotch emaravirus	36.87%	26.98%	28.57%	24.08%	31.91%	–
		YP_009237274.1	YP_009237265.1	YP_009237266.1	YP_009237267.1	YP_009237268.1	
15	Palo verde broom virus	35.24%	26.42%	24.42%	27.00%	–	–
		AWH90165.1	AWH90170.1	AWH90176.1	AWH90182.1		
16	Jujube yellow mottle-associated virus	37.60%	28.39%	28.23%	24.53%	–	–
		QDM38999.1	QDM39000.1	QDM39001.1	QDM39002.1		

(Continued)

TABLE 3 | Continued

		RdRP	GPP	NP	MP	P5	P6
17	High Plains wheat mosaic emaravirus	34.97%	28.17%	26.55%	24.14%	33.33%	–
		YP_009237277.1	QGT41075.1	YP_009237257.1	QGT41077.1	QGT41070.1	
18	Camellia japonica-associated emaravirus 1	32.61%	26.60%	29.73%	19.93%	–	–
		QGX73503.1	QGX73504.1	QGX73505.1	QGX73506.1		
19	Camellia japonica-associated emaravirus 2	32.89%	25.93%	25.85%	21.07%	–	–
		QGX73507.1	QGX73508.1	QGX73509.1	QGX73510.1		
20	Perilla mosaic virus	32.07%	23.75%	32.58%	21.36%	–	–
		BBM96177.1	BBM96178.1	BBM96180.1	BBM96181.1		
21	Common oak ringspot-associated emaravirus	35.73%	28.04%	27.27%	–	–	–
		LR828198	LR828199	LR828200			
22	Pea-associated emaravirus	–	–	50.52%	–	–	–
				QJX15716.1			

homology (supported by *E*-values < 0.01) with the corresponding protein detected from eight emaraviruses (30–43% aa identity). This putative protein shares highest identity with the P5 protein of rose rosette virus (Accession number: QIB98219. 1; aa identity: 44.8%), a protein of unknown function. The rest of the emaraviruses do not have a homologous protein or such has not been identified yet.

Finally, the sixth ORF, encoded on RNA6 (nt positions 766 – 71), produces a 231 aa-long protein with 35–45% aa identity with that of other emaraviruses. The highest aa identity is with the P6 protein from *Actinidia* emaravirus 2 (Accession number: QEE82891.1; aa identity: 44.2%). The novel virus is found to have a homologous protein to the one encoded by RNA6 segment of only eight from the already known emaraviruses.

Elbeaino et al. (2018) in the 10th report of the International Committee on Taxonomy of Viruses regarding the taxonomy of *Fimoviridae* describe that the virus genome of the family members comprises from four negative-sense ssRNA segments [like the type species European mountain ash ringspot-associated emaravirus (EMARaV) from the host European mountain ash] to eight RNA segments [High Plains wheat mosaic virus (Tatineni et al., 2014)]. Later, two novel emaravirus with five RNA segments were identified; lilac chlorotic ringspot-associated virus from lilac leaves with yellow mottle symptoms in China (Wang et al., 2020) and ti ringspot-associated virus from the ti plant (*Cordyline fruticosa*) in Hawaii (Olmedo-Velarde et al., 2019). Six genomic RNA segments were identified in jujube yellow mottle-associated virus causing the same named disease in jujube (*Ziziphus jujuba*) in China (Yang et al., 2019). Very recently, a novel emaravirus in Shiso named *Perilla mosaic virus* (PerMV) was found to consist of 10 RNA segments, each encoding a single protein in the negative-sense orientation (Kubota et al., 2020). In the case of the maple virus two independent NGS samples from the same maple tree detected the same number of genome segments. Additionally, for the Acer+ (2015) sample, the cDNA library was prepared using a primer targeting the conserved genome terminal region, known to be shared among emaraviruses genome segments. A screen for further genome segments containing these terminal regions did not reveal further sequences and we therefore consider that it is unlikely to have missed additional RNA segments of the viral genome. Concluding, we suggest that the segments RNA1-RNA6 represent the complete genome of the novel virus constituting a new member of the family *Fimoviridae*.

Phylogenetic Analysis of the Novel Emaravirus

Phylogenetic relationships between the maple virus and the sequences of 21 emaraviruses known to date were estimated, based on amino acid sequences comparisons. Irrespective of the RNA segment investigated, the maple viral agent clusters consistently in "subgroup a" (according to Elbeaino et al., 2018), together with – among others – rose rosette virus, fig mosaic virus, the novel aspen mosaic-associated virus, the two pigeon pea sterility mosaic emaraviruses and *Actinidia* emaravirus 2. The phylogenetic analysis performed here confirms the overall structure within the genus *Emaravirus* described by von Bargen

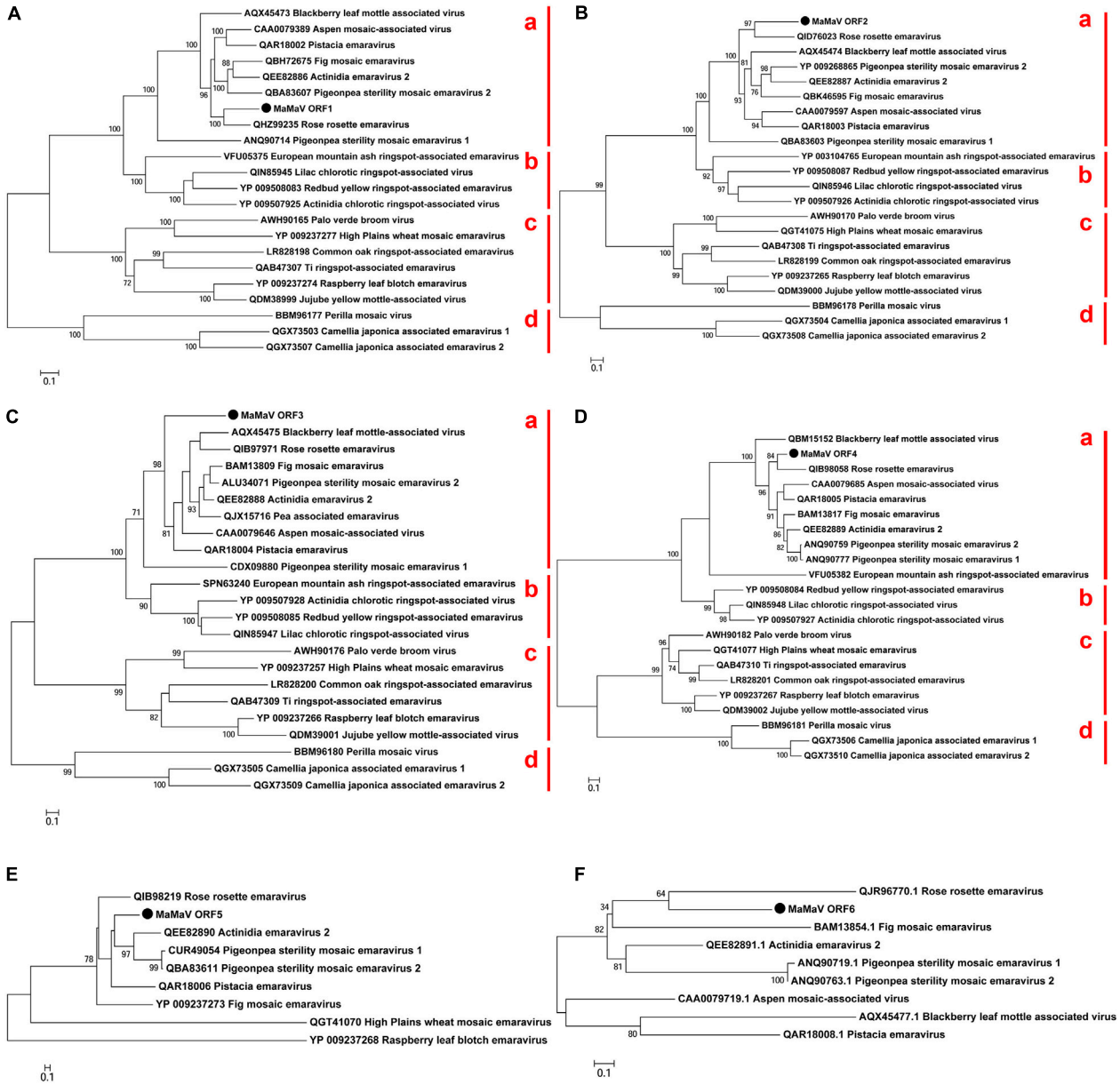


FIGURE 3 | Phylogenetic tree reconstructed using the amino acid sequences of the RdRP **(A)**, GPP **(B)**, NP **(C)**, MP **(D)**, P5 **(E)**, and P6 **(F)** of the MaMaV. The analysis involved 22 amino acid sequences of established and putative members of the genus *Emaravirus* (the GenBank accession numbers are also indicated). The tree was constructed using the Maximum Likelihood method and the statistical significance of branches was evaluated by bootstrap analysis (1,000 replicates). Only bootstrap values above 70% are indicated. The scale bar represents 10% amino acid divergence. Subgroups a and b described by Elbeaino et al. (2018) as well as subgroups c and d within the genus *Emaravirus* are indicated at the right side of the trees.

et al. (2020a) and enriches the tree with seven additional emaraviruses. **Figures 3A–D** shows representative ML trees obtained using the putative RdRP, the GPP, the NP and the MP encoded by the genome segments RNA1, RNA2, RNA3, and RNA4, respectively. The maple emaravirus proteins P5 and P6 are homologous to those of eight different emaraviruses (**Figures 3E,F**). All predicted proteins cluster together with the respective ones from rose rosette virus. The only exception is the NP protein, which still clusters with the "subgroup a," but with no close relation to any of its members. Additionally, a fourth subgroup is constructed ("subgroup d") based on the phylogenetic relationships inferred from ORF1 – ORF4 amino acid sequences, which consists of three recently discovered viruses, the perilla mosaic virus and the camellia japonica-associated emaraviruses 1 and 2.

The novel aa sequences, as it is shown from the phylogenetic analysis, are clearly genetically differentiated from all encoded proteins identified to date and cluster together with the ones from the "subgroup a." This result was already drawn by the taxonomic analysis with MEGAN, where the novel virus was shown to be

close related to the "subgroup a" member pigeon pea sterility mosaic emaravirus (PPSMV) (without differentiation between PPSMV-1 and PPSMV-2). At the same time, all predicted proteins of the maple emaravirus exhibit significant aa identities only with the ones from other emaraviruses. Taken together, these results demonstrate that the virus identified represents a new species in the genus *Emaravirus* and it is, therefore, tentatively named maple mottle-associated virus (MaMaV).

Validation of the RNA-Seq Results and Association of Virus Presence With Symptom Appearance

From the 32 maple trees tested by specific RT-PCR, MaMaV was only detected in the symptomatic ones and all six RNAs were generally simultaneously detected (**Supplementary Table 1**). For a few samples, one of the six RNA segments failed to be amplified [i.e., RNA2 in sample E54934; RNA1 (using primers RNA1bF/R) in samples E54936 and E54942 (although primers RNA1aF/R did produce amplicons with these samples); RNA5 in sample E54946]. Because these cases were rare and all other RNA segments in those samples were successfully amplified, we suggest the negative results were due to the low concentration of respective RNA segments beyond detection limit of the RT-PCR assay. Non-symptomatic samples were consistently negative for all MaMaV segments. These results suggest that the 26 positive-tested symptomatic maples were MaMaV-infected and that MaMaV could be the virus responsible for the identified virus-like symptoms.

It should be, however, underlined that the tested trees showed a variability of symptoms, where mottle was mainly observed, but other -weaker or stronger- symptoms were also exhibited. Although mottle was the main and common symptom exhibited in the majority of infected trees, occasionally other virus-like symptoms, like chlorotic ringspots, line pattern or flecking and in some rare cases also leaf deformations were observed. More concretely, apart from mottle, in 2015 also flecking was recorded, while in 2016 chlorotic ringspots and line pattern were observed, although in both years samples were collected at the same time (1st of July; see **Supplementary Table 1**).

Preliminary Results on Genetic Variability of Three RNA Segments

The evolutionary divergence between variants from different trees was estimated for three RNA segments. In these RNA segments low genetic diversity was estimated; maximal divergence among RT-PCR products amplified with primer-pairs RNA1aF/R and RNA1bF/R was 2% and 1.1%, respectively. Sequences were identical in two and in 13 of the pairwise comparisons for the RNA1aF/R and RNA1bF/R RT-PCR products, respectively. However, all sequences differed from the corresponding variant generated from the original sequence from Acer + (2014) by 0.3–1.6% and by 0.4–1.1% in the cases of the RT-PCR products amplified by primer-pair RNA1aF/R and RNA1bF/R, respectively (**Supplementary Table 2**). Regarding variability among RT-PCR products for the RNA3 and among RT-PCR products from the RNA4 segment, no genetic diversity was shown among the respective sequences and all were identical with the corresponding segment from the original sample Acer+ (2014).

Based on these results, evidence is provided that the diversity in the local Berlin-Grunewald MaMaV population is generally low and that different level of genetic diversity may characterize its different genome segments.

DISCUSSION

The six newly identified RNA segments are attributed to a novel *Emaravirus* species based on the following: (a) The multipartite genome is composed of six single-stranded RNA molecules; (b) All six RNAs share a fully conserved stretch of 13 nt at their 5′ and 3′ termini; (c) Each segment of the genome encodes a single protein, which shows sequence identity with homologous proteins of other emaraviruses; (d) In all phylogenetic trees generated with amino acid sequences, MaMaV is only distantly related phylogenetically to the emaraviruses currently represented in the GenBank fulfilling the current species demarcation criteria of emaraviruses to show more than 25% aa divergence of RNA1-RNA3 encoded proteins (Elbeaino et al., 2018). To our knowledge, for first time an emaravirus is described from maple and is fully genetically characterized.

The NGS method applied not only led to the discovery and genetic characterization of the novel emaravirus; it unraveled at the same time the maple virome, in terms of identifying the exhaustive collection of nucleic acid sequences deriving from viral agents. The maple virome of the symptomatic tree tested is found to be very simple, as it includes a single variant from a single virus. The lack of virome complexity is rather surprising, when we consider obtained NGS results from other wild as well as cultivated woody hosts. A complex virome was revealed in birch, where up to five virus variants were identified in the transcriptome of individual trees (Rumbou et al., 2020). In birch, it has been demonstrated that not only multiple viral species but also diverse variants of the same virus may accumulate in single trees (Rumbou et al., 2016, 2020). Similarly, in single peach trees multi-viral infection of up to six viruses and viroids was detected (Jo et al., 2018). A multiplicity of viral infections was also shown in metagenome samples from single plum trees (Jo et al., 2020). A possible explanation for the single viral infection in the maple tree could be the age of the tree; it was very young (approx. 3 years-old) when sampled, thus it was exposed for only a short time to viral pathogens.

The novel virus was only present in the tested symptomatic maple trees in Berlin, while it was not detected in non-symptomatic trees. This suggests that MaMaV presence is associated with the leaf symptoms identified in maples. Maple seedlings scions from Berlin-Grunewald were grafted to non-symptomatic 2-year old maple rootstocks in March 2017 (data not shown). In 2018 the grafted seedlings exhibited mottle, while MaMaV could be detected in the scions by RT-PCR in 2019 and in 2020. Detection of MaMaV in rootstocks was not yet possible, as either the grafted seedlings did not develop shoots from rootstocks, or non-symptomatic shoots from rootstocks

were not tested. Although the main observed symptom was mottle, other -weaker or stronger- symptoms were also exhibited. Whether MaMaV is the only causal agent for the different kind of symptoms or whether other viruses are also involved in the symptomatology needs to be further investigated. Further efforts are needed to satisfy Koch's postulates and firmly establish its causal role.

In maple trees, virus-like symptoms other than mottle have been observed and reported earlier not only in Germany (Bandte et al., 2008; Büttner et al., 2013), but also in Romania (Ploiaie and Macovei, 1968), Hungary (Szirmai, 1972), Turkey (Erdiller, 1986), North America (Brierley, 1944), United Kingdom (Cooper, 1979). Symptom descriptions such as "mosaic," "chlorotic mottle," "yellow-mottle," "ring-mottle," "mosaic mottling with chlorotic spots" were used by earlier researchers. Whether the newly identified virus might be associated with the symptoms described in these earlier studies remains to be determined, given in particular that electron-dense structures known as double-membrane-bound bodies with diameters differing within the range of 80–200 nm and typical of emaraviruses (Mielke-Ehret and Mühlbach, 2012) have never been reported in earlier electron microscopy samples from symptomatic maples, while more typical plant virus particles sometimes have. Furthermore, infection of maples by tobamoviruses reported by Führling and Büttner (1998) could not be confirmed by the present study. In that case, we suggest either that the samples observed by EM were contaminated during sample processing or that indeed tobamoviruses may also affect maples. It remains to be confirmed by future NGS analyses whether other viral agents may be present in symptomatic maples.

From the discovery of the first viral agent in maple thanks to the development of NGS's tools to its full biological characterization there is a long way to go (Massart et al., 2017). To investigate its potential for emergence, possible vectors and mode(s) of dispersal need to be determined. Several members of the genus *Emaravirus* are known to be transmitted by eriophyid mites (*Acari*: *Eriophyidae*) (Mielke-Ehret and Mühlbach, 2012; Hassan et al., 2017; Elbeaino et al., 2018). In the present study, in several of the sampled trees damages from gall mites (*Aceria macrophylla*, *Eriophyes psilomerus*) as well as from leafhoppers were found. The gall mites can be considered as putative vectors, but the hypothesis that they may be involved in the emaravirus transmission needs to be studied. Regarding MaMaV's impact on infected trees, we assume that the mottle symptoms exhibited in infected leaves may lead to reduced photosynthetic capacity and, consequently, to trees' health deterioration. Based on the available diagnostic RT-PCR assays designed in the frame of the present study, the investigation of the agents' distribution and its impact on the trees health can be estimated.

During the last decade a significant number of novel DNA and RNA viruses (Villamor et al., 2019) and viroids (Hadidi, 2019) have been uncovered in herbaceous as well as in woody hosts by applying a wide range of NGS methods for virus detection and discovery (Hadidi et al., 2016; Olmos et al., 2018; Villamor et al., 2019). In comparison to other viruses, emaraviruses were until recently overlooked, not only when applying conventional methods, but also by metagenomic studies (Bejerman et al., 2020). The application of NGS resulted in the discovery of 13 emaraviruses during the last 5 years (Bejerman et al., 2020). By employing NGS methodologies a significant presence of emaraviruses in forests has been revealed. In the last 3 years four emaraviruses have been discovered in six forest or urban tree species; aspen mosaic-associated virus (AsMaV) in *Populus tremula* (von Bargen et al., 2020a), European mountain ash ringspot-associated virus (EMARaV) in *Sorbus intermedia* (von Bargen et al., 2019), *Karpatiosorbus* × *hybrid* (von Bargen et al., 2020b), and *Amelanchier* sp. (von Bargen et al., 2018), common oak ringspot-associated virus (CORaV) in *Quercus robur* (Bandte et al., 2020), and maple mottle-associated virus (MaMaV) in *Acer pseudoplatanus*. Based on existing knowledge on viral disease of fruit trees (Hadidi et al., 2011) we suggest that viral agents might be a considerable stress factor for forests trees, they may predispose affected trees to other more harmful stress factors and, as a consequence, lead to deterioration of forests and other natural ecosystems (Büttner et al., 2013, 2015). Additional efforts in the field of forest virology are needed to provide data on the magnitude of the tree damage due to viral infections and to prevent future viral outbreaks.

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/, MT879190–MT879195.

AUTHOR CONTRIBUTIONS

AR and CB: conceptualization. AR and TC: data curation, formal analysis, and software. CB: funding acquisition and project administration. AR, TC, and SB: investigation and validation. AR and SB: methodology and visualization. AR: writing – original draft. TC, CB, and SB: Writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the German Research Foundation (DFG) (project BU890/27-1).

ACKNOWLEDGMENTS

This work is partly accomplished in the frame of the COST Action FA1407-DIVAS (Deep Investigation of Virus Associated Sequences) – "Application of next-generation sequencing for the study and diagnosis of plant viral diseases in agriculture." We deeply thank all colleagues from this COST action for their support to acquire the results – directly or indirectly – and for the fruitful and open scientific exchange. We also acknowledge support by the German Research Foundation

(DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.621179/full#supplementary-material

Supplementary Table 1 | Samples used for RT-PCR validation and genetic divergence analysis. Sample label, collection date, tree location, symptoms exhibited, damage caused by pests and the RT-PCR results (+/−) for each of the RT-PCR assays applied with seven different primer-pairs in 6 RNA segments of the novel emaravirus are shown.

Supplementary Table 2 | Estimates of evolutionary divergence among RT-PCR products from 10 maple trees and the original Acer+ (2014). **(A)** RT-PCR products amplified with primer-pair RNA1aF/R. **(B)** RT-PCR products amplified with primer-pair RNAbF/R.

REFERENCES

Atanasoff, D. (1935). Old and new virus diseases of trees and shrubs. *Phytopathol. Z.* 8, 197–223.

Bandte, M., Hamacher, J., and Büttner, C. (2008). "Maple tree (*Acer sp.*) infected by viruses and phytoplasma," in *Jahrbuch der Baumpflege: Thalacker-Medien*, eds H. D. Dujesiefken and P. Kockerbeck (Braunschweig: Deutsche Baumpflegetage), 208–212.

Bandte, M., Rehanek, M., Leder, B., von Bargen, S., and Buettner, C. (2020). Identification of an emaravirus in a common Oak (Quercus robur L.) Conservation seed orchard in germany: implications for Oak health. *Forests* 11:1174. doi: 10.3390/f11111174

Bejerman, N., Debat, H., and Dietzgen, R. G. (2020). The plant negative-sense RNA virosphere: virus discovery through new eyes. *Front. Microbiol.* 11:588427. doi: 10.3389/fmicb.2020.588427

Binggeli, P. (1993). Sycamore lore. *Plant Lore Notes News* 29, 131–133.

Boom, R., Sol, C. J. A., Salians, M. M. M., Jansen, C. L., Wertheim-van Dillen, P. M. E., and van der Noordaa, J. (1990). Rapid and simple method for puri cation of nucleic acids. *J. Clin. Microbiol.* 28, 495–503.

Brglez, A., Piškur, B., and Ogris, N. (2020). *Eutypella parasitica* and other frequently isolated fungi in wood of dead branches of young sycamore maple (*Acer pseudoplatanus*) in slovenia. *Forests* 11:467. doi: 10.3390/f11040467

Brierley, P. (1944). Viruses described primarily on ornamental or miscellaneous hosts. *Pl. Dis. Reptr.* 150, 410–482. (mimeographed).

Büttner, C., von Bargen, S., and Bandte, M. (2015). "Chapter 13: phytopathogenic viruses," in *Principles of Plant Microbe Interactions: Role of Microbes in Sustainable Agriculture*, ed. Lugtenberg (Heidelberg: Springer), 115–122. doi: 10.1007/978-3-319-08575-3_13

Büttner, C., von Bargen, S., Bandte, M., and Muehlbach, H. P. (2013). "Forests diseases caused by viruses," in *Infectious Forest Diseases*, eds P. Gonthier and G. Nicolotti (CABI), 50–75.

Cooper, J. I. (1979). *Virus Disease of Trees and Shrubs*. Oxford: Institute of Terrestrial Ecology, Natural Environment Research Council.

Di Bello, P. L., and Tzanetakis, I. E. (2013). *Rose rosette virus* is the causal agent of rose rosette disease. *Phytopathology* 103:S1.

Elbeaino, T., Digiaro, M., Mielke-Ehret, N., Muehlbach, H. P., and Martelli, G. P. (2018). ICTV report consortium, ICTV virus taxonomy profile: fimoviridae. *J. Gen. Virol.* 99, 1478–1479. doi: 10.1099/jgv.0.001143

Erdiller, G. (1986). Acer virus diseases in Turkey. *J. Turkish Phytopathol.* 15, 46–59.

Führling, M., and Büttner, C. (1998). Nachweis von Tobamo-Viren in Bergahorn (*Acer pseudoplatanus L.*) mit Scheckung und Blattdeformation. *Forstw. Cbl.* 117, 92–97. doi: 10.1007/bf02832962

Gibbs, D., and Chen, Y. (2009). *The Red List of Maples. Archived 2019-05-28 at theWayback Machine*. Richmond: Botanic Gardens Conservation International (BGCI).

Hadidi, A. (2019). Next-generation sequencing and CRISPR/Cas13 editing in viroid research and molecular diagnostics. *Viruses* 11:120. doi: 10.3390/v11020120

Hadidi, A., Barba, M., Candresse, T., and Jelkmann, W. (2011). *Virus and Virus-Like Diseases of Pome and Stone Fruits*. St. Paul, MN: APS Press.

Hadidi, A., Flores, R., Candresse, T., and Barba, M. (2016). Next-generation sequencing and genome editing in plant virology. *Front. Microbiol.* 7:1325.

Hassan, M., Di Bello, P. L., Keller, K. E., Martin, R. R., Sabanadzovic, S., and Tzanetakis, I. E. (2017). A new, widespread emaravirus discovered in blackberry. *Virus Res.* 235, 1–5. doi: 10.1016/j.virusres.2017.04.006

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12:e1004957. doi: 10.1371/journal.pcbi.1004957

Jo, Y., Choi, H., Lian, S., Cho, J. K., Chu, H., and Cho, W. K. (2020). Identification of viruses infecting six plum cultivars in Korea by RNA-sequencing. *PeerJ* 8:e9588. doi: 10.7717/peerj.9588

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

Kaminska, M., and Suwa, H. (2006). First report of a decline of ash leaf maple (*Acer negundo*) in Poland, associated with *Candidatus* Phytoplasma asteris. *Plant Pathol.* 55:293. doi: 10.1111/j.1365-3059.2005.01304.x

Kubota, K., Usugi, T., Tomitaka, Y., Shimomoto, Y., Takeuchi, S., Kadono, F., et al. (2020). Perilla mosaic virus is a highly divergent emaravirus transmitted by *Shevtchenkella* sp. (Acari: Eriophyidae). *Phytopathology* 7, 1352–1361. doi: 10.1094/PHYTO-01-20-0013-R

Lana, A. E., Thomas, O. T., and Peterson, J. E. (1980). A virus isolated from sugar maple. *Phytopathol. Z* 97, 214–218. doi: 10.1111/j.1439-0434.1980.tb03689.x

Larsson, A. (2015). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531

Lefebvre, M., Theil, S., Ma, Y., and Candresse, T. (2019). The VirAnnot pipeline: a resource for automated viral diversity estimation and operational taxonomy units (OTU) assignation for virome sequencing data. *Phytobiom. J.* 3, 256–259. doi: 10.1094/PBIOMES-07-19-0037-A

Li, Z.-N., Zhang, L., Zhao, L., and Wu, Y.-F. (2012). A new phytoplasma associated with witches-broom on Japanese maple in China. *For. Path.* 42, 371–376. doi: 10.1111/j.1439-0329.2012.00769.x

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

Mielke-Ehret, N., and Mühlbach, H.-P. (2012). Emaravirus: a novel genus of multipartite, negative strand RNA plant viruses. *Viruses* 4, 1515–1536. doi: 10.3390/v4091515

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., et al. (2016). Linking virus genomes with host taxonomy. *Viruses* 8:66. doi: 10.3390/v8030066

Olmedo-Velarde, A., Park, A. C., Sugano, J., Uchida, J. Y., Kawate, M., Borth, W. B., et al. (2019). Characterization of ti ringspot-associated virus, a novel emaravirus associated with an emerging ringspot disease of cordyline fruticosa. *Plant Dis.* 103:9. doi: 10.1094/PDIS-09-18-1513-RE

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

Ploiaie, P. G., and Macovei, A. (1968). New plant virosis recorded in Roumania. *Revue. Roum. Biol. (Ser. Bot.)* 13, 269–274.

Rumbou, A., Candresse, T., Marais, A., Svanella-Dumas, L., Landgraf, M., von Bargen, S., et al. (2020). Unravelling the virome in birch: RNA-Seq reveals a complex of known and novel viruses. *PLoS One* 15:e0221834. doi: 10.1371/journal.pone.0221834

Rumbou, A., Candresse, T., Marais, A., Theil, S., Langer, J., Jalkanen, R., et al. (2018). A novel badnavirus discovered from *Betula sp*. affected by birch leaf-roll disease. *PLoS One* 13:e0193888. doi: 10.1371/journal.pone.0193888

Rumbou, A., von Bargen, S., Demiral, R., Langer, J., Rott, M., Jalkanen, R., et al. (2016). High genetic diversity at the inter-/intra-host level of *Cherry leaf roll virus* population associated with the birch leaf-roll disease in Fennoscandia. *Scand. J. For. Res.* 31, 546–560. doi: 10.1080/02827581.2016.1165283

Subikova, V. (1973). The mechanical transmission of Euonymus mosaic virus, maple leaf perforation by leaf extracts or leaf nucleic acid to herbaceous plants. *Biologia Plant.* 15, 166–170. doi: 10.1007/bf02922389

Szirmai, J. (1972). An *Acer* virus disease in maple trees planted in Avenues. *Acta Phytopathol. Acad. Sci. Hung.* 7, 197–207.

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11030–11035. doi: 10.1073/pnas.0404206101

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

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

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109, 716–725. doi: 10.1094/phyto-07-18-0257-rvw

von Bargen, S., Al Kubrusli, R., Gaskin, T., Fürl, S., Hüttner, F., Blystad, D.-R., et al. (2020a). Characterisation of a novel Emaravirus identified in mosaic-diseased Eurasian aspen (*Populus tremula*). *Ann. Appl. Biol.* 176, 210–222. doi: 10.1111/aab.12576

von Bargen, S., Bandte, M., Al Kubrusli, R., Jalkanen, R., and Büttner, C. (2020b). First report of European mountain ash ringspot-associated virus in *Karpatiosorbus* × hybrida in Finland. *New Dis. Rep.* 42:1.

von Bargen, S., Dieckmann, H.-L., Candresse, T., Mühlbach, H.-P., Roßbach, J., and Büttner, C. (2019). Determination of the complete genome sequence of European mountain ash ringspot-associated emaravirus from *Sorbus intermedia* reveals two additional genome segments. *Arch Virol.* 164, 1937–1941. doi: 10.1007/s00705-019-04275-0

von Bargen, S., Tischendorf, M., and Büttner, C. (2018). First report of *European mountain* ash ringspot-associated virus in serviceberry (*Amelanchier* spp.) in Germany. *New Dis. Rep.* 37, 19–19. doi: 10.5197/j.2044-0588.2018.037.019

Wang, Y., Song, Y., Cao, M., Cheng, Q., Wu, J., and Hu, T. (2020). Identification of a novel emaravirus infecting lilac through next-generation sequencing. *J. Integr. Agricult.* 19, 2064–2071. doi: 10.1016/S2095-3119(19)628066

Wulf, A., and Kehr, R. (2009). "Diseases, disorders and pests of selected valuable broadleaved tree species," in *Valuable Broadleaved Forests in Europe*, eds H. Spiecker, S. Hein, K. Makkonen-Spiecker, and M. Thies (Leiden: Brill), 61–84.

Yang, C., Zhang, S., Han, T., Fu, J., Di Serio, F., and Cao, M. (2019). Identification and characterization of a novel emaravirus associated with jujube (*Ziziphus jujuba* Mill.) yellow mottle disease. *Front. Microbiol.* 10:1417. doi: 10.3389/fmicb.2019.01417

Zhang, Y. Z., Shi, M., and Holmes, E. C. (2018). Using metagenomics to characterize an expanding virosphere. *Cell* 172, 1168–1172. doi: 10.1016/j.cell.2018.02.043

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.

Copyright © 2021 Rumbou, Candresse, von Bargen and Büttner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

REVIEW
published: 12 January 2021
doi: 10.3389/fmicb.2020.609376

Next-Generation Sequencing and the CRISPR-Cas Nexus: A Molecular Plant Virology Perspective

Muhammad Shafiq Shahid[1], Muhammad Naeem Sattar[2], Zafar Iqbal[2], Amir Raza[1] and Abdullah M. Al-Sadi[1]*

[1] *Department of Plant Sciences, College of Agricultural and Marine Sciences, Sultan Qaboos University, Muscat, Oman,*
[2] *Central Laboratories, King Faisal University, Hofuf, Saudi Arabia*

In recent years, next-generation sequencing (NGS) and contemporary Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR-associated (Cas) technologies have revolutionized the life sciences and the field of plant virology. Both these technologies offer an unparalleled platform for sequencing and deciphering viral metagenomes promptly. Over the past two decades, NGS technologies have improved enormously and have impacted plant virology. NGS has enabled the detection of plant viruses that were previously undetectable by conventional approaches, such as quarantine and archeological plant samples, and has helped to track the evolutionary footprints of viral pathogens. The CRISPR-Cas-based genome editing (GE) and detection techniques have enabled the development of effective approaches to virus resistance. Different versions of CRISPR-Cas have been employed to successfully confer resistance against diverse plant viruses by directly targeting the virus genome or indirectly editing certain host susceptibility factors. Applications of CRISPR-Cas systems include targeted insertion and/or deletion, site-directed mutagenesis, induction/expression/repression of the gene(s), epigenome re-modeling, and SNPs detection. The CRISPR-Cas toolbox has been equipped with precision GE tools to engineer the target genome with and without double-stranded (ds) breaks or donor templates. This technique has also enabled the generation of transgene-free genetically engineered plants, DNA repair, base substitution, prime editing, detection of small molecules, and biosensing in plant virology. This review discusses the utilities, advantages, applications, bottlenecks of NGS, and CRISPR-Cas in plant virology.

Keywords: CRISPR, CRISPR associated (Cas) proteins, genome editing, next generation sequencing (NGS), plant viruses

OPEN ACCESS

Edited by:
John Wesley Randles, University of Adelaide, Australia

Reviewed by:
C. Martin Lawrence, Montana State University, United States
Hernan Garcia-Ruiz, University of Nebraska-Lincoln, United States

****Correspondence:***
Muhammad Shafiq Shahid
mshahid@squ.edu.om; shafiqinayat@gmail.com
orcid.org/0000-0002-3550-0000

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

Received: *23 September 2020*
Accepted: *14 December 2020*
Published: *12 January 2021*

Citation:
Shahid MS, Sattar MN, Iqbal Z, Raza A and Al-Sadi AM (2021) Next-Generation Sequencing and the CRISPR-Cas Nexus: A Molecular Plant Virology Perspective. Front. Microbiol. 11:609376. doi: 10.3389/fmicb.2020.609376

INTRODUCTION

Plant viruses infect diverse plant species across the globe. During successful infections, viruses trigger an array of interactions with insect vectors or the plant hosts. Some plant viruses have acquired extra-viral components, *viz.*, DNA-satellites, RNA-satellites, and satellite viruses (Mansoor et al., 1999; Briddon et al., 2001; Palukaitis, 2016). Plant viruses have caused a significant reduction in crop productivity across Asia, Africa, Europe, and South America, resulting in losses of approximately 30 billion US$ annually (Sastry and Zitter, 2014). For example, in the last decade,

cassava mosaic disease caused an approximate 25 million ton reduction in cassava production worldwide (Legg and Thresh, 2000; Thresh and Cooter, 2005). Millions of citrus plants have been destroyed annually by the Citrus Tristeza virus (CTV) (Moreno et al., 2008; Harper, 2013). The potato leafroll virus has led to a loss of approximately US$100 million in the United States and around £50 million in the United Kingdom (Wale et al., 2008; Sastry and Zitter, 2014). Similarly, during the period between 1992 and 9197, cotton leaf curl disease caused a loss of around US$5 billion (Briddon et al., 2001) to Pakistan's economy.

An effective approach to virus control requires efficient detection methods and subsequent insights into the genomic architecture of the target viruses. Various approaches, including enzyme-linked immunosorbent assay (ELISA), restriction enzyme analysis, polymerase chain reaction (PCR), and reverse transcriptase PCR (RT-PCR), have been regularly used as an initial screening tool. Most of these diagnostic techniques rely on prior knowledge of viral genomes so that unknown viruses may remain undetected. Next-generation sequencing (NGS) technologies have revolutionized the field of molecular biology, especially plant virology, by comprehensively unearthing the genomic data at a level that was not possible before. Contemporary NGS technologies can sequence all types of nucleic acid molecules, concurrently. NGS technologies have enabled the detection of novel pathogenic viruses that have remained undetected due to low viral titer or detection threshold levels (Villamor et al., 2019). These NGS technologies facilitate the discovery of overlooked plant virus species and help broaden our understanding of phytoviromes.

Recombination has been used as a tool to modify prokaryotic genomes, but this approach was least specific and low yielding. The discovery of four sequence-specific endonucleases such as meganucleases, Zinc Finger Nuclease (ZFN), Transcription Activator like Effector Nuclease (TALEN), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR-associated protein 9 (CRISPR-Cas9) (Zhang et al., 2013) substantially improved the genome editing (GE) in higher organisms (Wiedenheft et al., 2011; Jinek et al., 2012, 2013; Gaj et al., 2013). Among them, the CRISPR-Cas system is the simplest, most efficient, and versatile GE tool that allows site-directed mutagenesis at the desired genomic position (Gaj et al., 2013; Bortesi and Fischer, 2015).

CRISPR-Cas systems are derived from prokaryotic immune systems that provide indigenous immunity against invading nucleic acids (**Figure 1**). CRISPR-Cas systems are diverse and can be divided into two major classes, six different types, and multiple types (Makarova et al., 2020). The main components of the most widely used CRISPR-Cas systems are Cas9 endonucleases, which are derived from different microorganisms, such as *Streptococcus pyogenes*, *Staphylococcus aureus*, and *Francisella novicida*, and are part of class II type II systems (Makarova and Koonin, 2015; Makarova et al., 2020). These Cas9 proteins are accompanied by CRISPR-RNA (crRNA) and trans-activating CRISPR-RNA (tracrRNA) for sequence-dependent cleavage of foreign nucleic acids (Hille et al., 2018; Makarova et al., 2020). The invention of a single-guide RNA (sgRNA) increased the potential applications of CRISPR-Cas systems (Jinek et al., 2012). As an initial step in CRISPR-Cas-based prokaryotic immunity, nucleic acid fragments of the invading pathogens are integrated into the CRISPR locus during infection. The subsequent infections thus activate the transcription of these smaller fragments as part of the CRISPR array they coordinate with CRISPR-associated (Cas) protein machinery to recognize, bind, and cleave to the foreign DNA/RNA elements.

In addition to the well-known CRISPR-Cas9 systems, the biotech applications for a number of the other CRISPR types and subtypes are also under active development. CRISPR-Cas systems are classified into two major classes based on interference type: Class I and Class II. Class I and II are further categorized into six types based on the type of nucleic acid they target (Makarova and Koonin, 2015). The evolutionary classification of CRISPR-Cas systems, especially class II and its variants, has been appraised by Makarova et al. (2020). Among the class II single Cas protein systems, these include Cas12 (type V), Cas9 (type II), Cas13a-d (type VI), and Cas14a-c (type V-F) (Burstein et al., 2017; Shmakov et al., 2017). Like Cas9, Cas12 targets double-stranded (ds) DNA, however, Cas13 (type VI) effector proteins: Cas13a (C2c2) (Abudayyeh et al., 2016), Cas13b (C2c6) (Cox et al., 2017), Cas13c (C2c7) (Shmakov et al., 2017), and Cas13d (Yan et al., 2018) are characterized as RNA-guided ribonucleases in microbial genomes.

In contrast to the class 2 systems, class 1 systems are composed of multiple Cas proteins, in different combinations, depending on the type and sub-type. Specifically, type I complexes are generally composed of Cas5, Cas6, Cas7, Cas8, and Cas11, and target dsDNA, while the type III complexes are composed of Cas5, Cas6, Cas8, and Cas10 and these are widespread in the immune system of nearly a quarter of bacterial species (Koonin et al., 2017). Type III CRISPR is further divided into two subtypes, i.e., type III-A and type III-B, based on two main Cas effectors (Cas10-Csm and Cas10-Cmr). Type II are composed of Cas5, Cas6, Cas7, and Csf1. Interestingly, these type III systems do not require a PAM, and instead, target nascent mRNAs and the corresponding DNA in transcriptionally active complexes.

In the last few years, the applications of CRISPR-Cas technology have been extended to all the fields of bioscience, including animal and human cell lines (Belhaj et al., 2015), as well as human viruses (both RNA and DNA) (Hadidi et al., 2016). CRISPR technology has extensive applications encompassing the insertion and/or deletion of a particular segment of DNA, introducing site-directed mutagenesis, expression and/or repression of genes, and epigenome remodeling. CRISPR-Cas systems offer great advantages including ease of cloning, low cost, and multiplexing, where multiple sites in the genome can be targeted simultaneously (Shin et al., 2017; Manghwar et al., 2019; Lee et al., 2020). Several studies have demonstrated the effectiveness of CRISPR-Cas technology due to its fast, easy to use applications in recalcitrant species, which can be effectively used to introduce or remove different genes (at a time) and do not require many manipulating tools. Since the first application of CRISPR-Cas-based GE in plants in 2013, this technique has been used continuously to engineer resistance against a variety of plant viruses (Scheben et al., 2017; Vats et al., 2019; El-Mounadi et al., 2020).

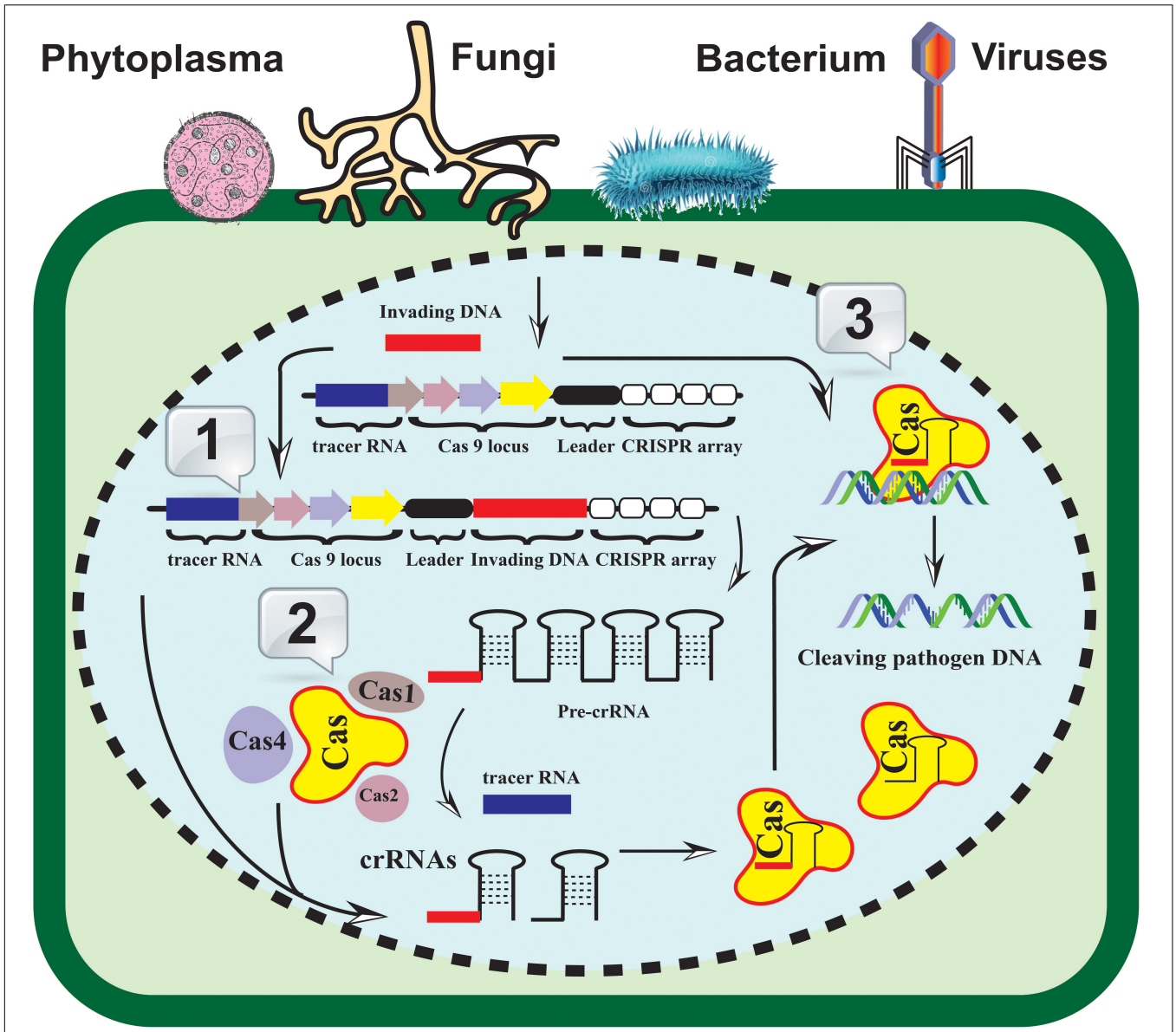


FIGURE 1 | A generalized CRISPR-Cas-based resistance induction model. The identification, recruitment, and cleavage of the invading pathogen's nucleic acid (fungi, bacteria, viruses, and/or phytoplasma) are achieved in three fundamental steps. 1. Acquisition: the invaded DNA (bar in red color) is integrated into the CRISPR-array (white rectangles) as new spacers; 2. Expression: the pre-CRISPR and crRNAs machinery is triggered and expressed in the invaded cell; 3. Interference: the mature crRNA hybridized to the invading pathogen genome and subsequently recognized by the Cas proteins. The coordination of Cas helicase and nuclease with CRISPR RNA machinery results in the invader's nucleic acid cleavage.

Three general approaches to engineering anti-viral mechanisms using CRISPR-Cas have been suggested: (1) the Cas9/sgRNA complex recruits viral genetic elements such as the origin of replication (ori) and averts the binding of replication-associated proteins, (2) the Cas9/sgRNA complex directly dissects the viral nucleic acid to prevent viral DNA replication, and (3) the Cas9/sgRNA complex mutates the viral genome at certain positions through non-homologous end joining (NHEJ) (Chaparro-Garcia et al., 2015). In addition to Cas9, a recently discovered endonuclease Cas12a (formerly referred to as Cpf1) offers dual nuclease activity as an endoribonuclease for crRNA processing and as endo-deoxyribonuclease to dissect the nucleic acid and produce ds breaks (DSB), respectively (Alok et al., 2020). Another milestone in CRISPR-Cas-based GE is the recent discovery of another effector protein, Cas13, an RNAi analog in eukaryotes for directly targeting the RNA genomes of plant viruses (Abudayyeh et al., 2017). This system has not been exploited yet against DNA viruses. However, it can be used to target the mRNA of DNA viruses during infection (Loriato et al., 2020). In geminiviruses, the CRISPR-Cas system can be employed to effectively inhibit virus accumulation by targeting any genomic region and stacking different sgRNAs against

a single virus or multiple viruses and their associated DNA-satellites (Iqbal et al., 2016; Rahman et al., 2017; Dahan−Meir et al., 2018; Ali et al., 2019; Roy et al., 2019).

Advances in NGS and GE technologies are revolutionizing the fields of genetics, genomics, molecular biology, and others, including plant virology. NGS technologies helped in the emergence and evolution of modern gene silencing and GE technologies, such as RNA interference (RNAi) and CRISPR-Cas, respectively. In this review article, we discuss the role of NGS and CRISPR-Cas technologies in plant virology.

NEXT-GENERATION SEQUENCING TECHNOLOGIES

A huge effort was made to sequence the entire human genome, which cost about US$ 3 billion and took nearly a decade to complete using the Sanger sequencing platform. NGS technologies have empowered the processing of large genomic datasets that generate an enormous amount of sequence reads and reduce costs to less than US$ 1000 per genome. Other advantages of NGS include the quantification of gene expression, the discovery of new RNA species such as microRNA (miRNA) regulating gene expression, metagenomic analysis, identification of latent viruses, multiplexing, and the ability to decipher specific species or strain of virus(es) without prior knowledge (Mardis, 2011). The development of new software with improved bioinformatics algorithms has also contributed to revolutionizing NGS techniques. NGS techniques have been categorized from first-to-fourth generations and are briefly discussed in the subsequent sections.

EVOLUTION OF NGS TECHNOLOGIES AND THEIR ROLE IN VIRUS DISCOVERY

Part of the first generation of sequencing technologies, Sanger and Maxam-Gilbert's invention of the first DNA sequencers revolutionized the field of molecular biology (Sanger et al., 1977).

First-Generation Sequencing Platforms

The first automated Sanger sequencing platform with single capillary electrophoresis (ABI Prism 310) was made commercially available in the mid-1990s (Watts and MacBeath, 2001). It was subsequently upgraded to a pre-NGS DNA technology with 96 lanes capillary electrophoresis (Paegel et al., 2002) to fix the issues associated with cost, time, and quality of the sequencing output (Metzker, 2005). For the last four decades, the Sanger sequencing technique is the preferred method for researchers dealing with the smaller DNA genomes of plant viruses. The ABI 3730xl DNA sequencer can generate reads from a genome as small as 1.9 kb to as long as 84 kb with a maximum precision of 99.99% and an average read length of 400–900 bp in about 3 h run-time (Liu et al., 2012). The impediments related to the Sanger sequencing platforms, including low scalability, tediousness, and cost ineffectiveness, have been developed in present-day cutting-edge sequencing technologies.

Second-Generation Sequencing Platforms

To address the shortcomings of the Sanger sequencing technique, second-generation sequencing technologies were released in 2005. They were later owned and commercially launched by Roche in 2007. These sequencing platforms not only reduced sequencing costs but produced millions of short reads (Kchouk et al., 2017); the Roche 454 platform worked on the principle of emulsion polymerase-mediated nucleotide extension and utilized sequencing-by-synthesis chemistry, which captures a template molecule, loads it into the well, and sequences the template genome on the basis of parallel pyrosequencing (Rothberg et al., 2011). The latest 454 FLX Titanium XL + framework can deliver around one million reads in a single run spanning about 23 h and having read lengths of about 700 bp with a maximum accuracy of 99.997% (**Table 1**).

The Solexa Genome Analyzer, introduced in 2006 and later acquired by Illumina in 2007, was another addition to the second-generation sequencing platform (Balasubramanian, 2015). Improvements in high-throughput sequencing, a less than 1% error rate, and substantial reductions in sequencing costs made Illumina sequencers the most suitable and frequently used NGS platforms (Fox et al., 2014; Pfeiffer et al., 2018). Apple stem grooving virus, blackberry chlorotic ringspot virus, and prunus necrotic ringspot virus, in addition to a previously unknown rose-leaf rosette-associated virus, were identified as infecting rose plants using the Illumina-Solexa sequencing platform (He et al., 2015). Recently, a mixed infection of tomato yellow leaf cur virus (TYLCV), mungbean yellow mosaic virus (MYMIV), and an associated tomato leaf curl betasatellite was identified in tomato plants using the Illumina HighSeq 4000PE101 platform (Shahid et al., 2019). Liang et al. (2015) employed the Illumina HiSeq 2000 sequencing platform to detect a novel apple geminivirus (AGV) infecting apple trees in China. Sequencing by Oligo Ligation Detection (SOLiD) is another second-generation platform with a throughput of around 9 gigabytes (GB) per run and an average read length of 50 bp. The output of the available SOLiD platforms can also be increased using a quadrant and/or octet slide platform through multiplexing. The high-end SOLiD 5500xl system offers 180 GB output data per run and an improved read length of 2×60 bp and produces ca. 3 billion paired-end reads in 24 h with 99.99% accuracy. Another second-generation sequencing platform, the Ion Torrent Semiconductor system, is based on the principle of detecting the released H^+ ions and is the most suitable platform for microbiome and exome sequencing (Rothberg et al., 2011).

Third-Generation Sequencing Platforms

The development of third-generation sequencing technologies has led to many problems associated with second-generation sequencing systems, including intensive sample preparation, PCR-based amplification, and sequencing time. These technologies use the single molecule real time (SMRT) sequencing method, which uses fluorescently labeled nucleotide bases that fluoresce upon incorporation into the growing DNA read. The PacBio sequencing system is a third-generation SMRT

TABLE 1 | Next-generation sequencing (NGS)-mediated discovery of selected different plant infecting viruses.

Sequencing platform	Virus identified*	Genome	Virus taxonomy	Host plant	Region	References
Illumina	CLCuGV, ToLCSDV, TYLCV-OM	ssDNA	Begomovirus (*Geminiviridae*)	Tomato and okra	Saudi Arabia	Idris et al., 2014
	SqMV	(+)ssRNA	Comovirus (*Secoviridae*)	Squash	Spain	Li et al., 2015
	SbBMV, ToYSV, ToYVSV, ToDfLV, SiGMBRV, PepBLV	ssDNA	Begomovirus (*Geminiviridae*)	Pepper	Argentina	Bornancini et al., 2020
	ToMV	(+)ssRNA	Tobamovirus (*Virgaviridae*)	Chickpea	Netherlands	Pirovano et al., 2015
	AGV	ssDNA	*Geminiviridae*	Apple	China	Liang et al., 2015
	MMDaV, CCDaV	ssDNA	Mastrevirus and Becurtovirus (*Geminiviridae*)	Mulberry	China	Ma et al., 2015
	CuLCrV	ssDNA	Begomovirus (*Geminiviridae*)	*B. tabaci*	Florida	Ng et al., 2011
	TPNRBV, TPLPV	(+)ssRNA	Blunervirus and Ilarvirus (*Bromoviridae*)	Tea	China	Hao et al., 2018
	NSPaV	(+)ssRNA	Luteovirus (*Luteoviridae*)	Nectarine	California	Bag et al., 2015
	CYLV	(+)ssRNA	Closterovirus (*Closteroviridae*)	Carrot	United Kingdom	Adams et al., 2014
	CCDaV	ssDNA	*Geminiviridae*	Citrus	Italy	Loconsole et al., 2012b
	PPSMV	(-)ssRNA	Emaravirus (*Luteoviridae*)	Pigeon pea	Italy	Elbeaino et al., 2014
	GRLaV	ssDNA	Grablovirus (*Geminiviridae*)	Grape	California	Poojari et al., 2013
	CVEV	(-)ssRNA	Emaravirus (*Luteoviridae*)	Citrus	Spain	Vives et al., 2013
	MYMIV	ssDNA	Begomovirus (*Geminiviridae*)	Tomato	Oman	Shahid et al., 2019
	SsHADV-1	ssDNA	Gemycircularvirus (*Genomoviridae*)	Dragonfly	Arizona/Oklahoma	Dayaram et al., 2015
	LCV, TCV, CpMMV	(+)ssRNA	Criniviruses (*Closteroviridae*) and Carlavirus (*Betaflexiviridae*)	*B. tabaci*	Florida	Rosario et al., 2014
Roche 454	BVF	dsDNA	Badnavirus (*Caulimoviridae*)	Blackberry	Arkansas	Shahid et al., 2017
	LChV-1	(+)ssRNA	Velarivirus (*Closteroviridae*)	Cherry	France	Candresse et al., 2013
	GMMV	(+)ssRNA	Cucumovirus (*Bromoviridae*)	Tomato	United Kingdom	Adams et al., 2009
	RLBV	(-)ssRNA	Emaravirus (*Luteoviridae*)	raspberry	United Kingdom	McGavin et al., 2012
	SWSV	ssDNA	Mastrevirus (*Geminiviridae*)	Sugarcane	Egypt	Candresse et al., 2014
	RRV	(-)ssRNA	Emaravirus (*Luteoviridae*)	Rose	Arkansas	Laney et al., 2011
Oxford Nanopore [MinION]	PVS, PVX, PVY, PLRV	(+)ssRNA	Carlavirus (*Betaflexiviridae*), Potexvirus (*Alphaflexiviridae*), Potyvirus (*Potyviridae*), Polerovirus (*Luteoviridae*)	Potato	Ireland	Della Bartola et al., 2020
	CMV	ssDNA	Begomovirus (*Geminiviridae*)	Cassava	Tanzania, Uganda, and Kenya	Boykin et al., 2019
	WSMV, TriMV, BYDV	(+)ssRNA	Poacevirus and Tritimovirus (*Potyviridae*), Luteovirus (*Luteoviridae*)	Wheat	United States	Fellers et al., 2019
	PPV	(+)ssRNA	Potyvirus (*Potyviridae*)	Plant tissue	United States	Bronzato Badial et al., 2018

**Selected viruses are included in each case and acronyms used are: apple geminivirus (AGV), barley yellow dwarf virus (BYDV), blackberry virus F (BVF), carrot yellow leaf virus (CYLV), cassava mosaic viruses (CMV), citrus vein enation virus (CVEV), citrus yellow vein clearing virus (CYVCV), citrus chlorotic dwarf-associated (CCDaV), cucurbit leaf crumple virus (CuLCrV), cotton leaf curl Gezira virus (CLCuGV), cowpea mild mottle virus (CpMMV), gayfeather mild mosaic virus (GMMV), grapevine red leaf-associated virus (GRLaV), lettuce chlorosis virus (LCV), little cherry virus 1(LChV-1), mulberry mosaic dwarf-associated virus (MMDaV), mungbean yellow mosaic India virus (MYMIV), nectarine stem pitting-associated virus (NSPaV), pepper blistering leaf virus (PepBLV), pigeon pea sterility mosaic virus (PPSMV), plum pox virus (PPV), potato virus Y (PVY), potato virus X (PVX), potato virus S (PVS), potato leafroll virus (LRV), raspberry leaf blotch virus (RLBV), riticum mosaic virus (TriMV), rose rosette virus (RRV), sclerotinia sclerotiorum hypovirulence-associated DNA virus-1 (SsHADV-1), Sida golden mosaic Brazil virus (SiGMBRV), Soybean blistering mosaic virus (SbBMV), squash mosaic virus (SqMV), sugarcane white streak virus (SWSV), tea plant necrotic ring blotch virus (TPNRBV), tea plant line pattern virus (TPLPV), tomato leaf curl Sudan virus (ToLCSDV), tomato yellow leaf cur virus-Oman (TYLCV-OM), tomato yellow spot virus (ToYSV), tomato yellow vein streak virus (ToYVSV), tomato dwarf leaf virus (ToDfLV), tomato mottle mosaic virus (ToMV), tomato chlorosis virus (TCV), and wheat streak mosaic virus (WSMV).*

sequencing system, which requires less than 5 h for sample preparation with a much-reduced cost, and produces an average read length of around 10 kb (Liu et al., 2012; Chin et al., 2016; Kchouk et al., 2017). The only drawback is the high error rate (14%) in the PacBio system, which has been resolved by recurrent sequencing of a single DNA molecule using hairpin adaptors to generate a circular ds DNA template and producing 30x consensus sequences with a superior accuracy of about 99.99%. Oxford Nanopore (*viz.*, MinION and PromethION) are single-molecule sequencing technologies developed by Oxford Nanopore Technologies (ONT). This technology provides not only better resolution but also long reads of superior quality and MinION is also suitable for single-nucleotide polymorphisms (Greig et al., 2019). The PromethION system can give an

∼8.5 TB and produces high-quality long reads of around 10 kb at a very low cost. Although the error rate is very high (∼10–15%), it is still being implemented in various research areas (Rang et al., 2018). ONT was successfully exploited for the identification of begomoviruses causing cassava mosaic disease in cassava plants in sub-Saharan African countries (Boykin et al., 2019) (**Table 1**).

Fourth-Generation Sequencing Platforms

Subsequent advances in NGS technologies resolved the shortcomings of preceding platforms. The recently developed, *in situ* sequencing (ISS) is a fourth-generation sequencing platform that directly sequences the nucleic acids by spatially resolving the transcriptomics of cells and tissues (Mignardi and Nilsson, 2014). ISS offers a great advantage by revealing distinctive variations, even at the single nucleotide level (Ke et al., 2016). Despite the limitations associated with each platform, there is an appropriate platform for each experimental requirement. Fourth-generation sequencing platforms have not yet been employed for plant virus detection but show potential and it is anticipated that future studies will continue to develop the potentials of this technology in this area.

Application of NGS to Deciphering the Role of miRNAs in Plant–Virus Interactions

Non-coding small RNAs (sRNAs), including miRNAs and short interfering RNAs (siRNAs), comprise 20–30 nt long molecules that play regulatory roles in plant–virus interactions. For instance, the expressions of miR156, miR158, miR160, miR164, and miR1885 were elicited during turnip mosaic virus (TuMV) infection (He et al., 2008) and miR162 was upregulated during cotton leafroll dwarf polerovirus (CLRDV) infection (Silva et al., 2011). The upregulation of miR444 and downregulation of miR528 and miR396 in rice, wheat, barley, sorghum, and sugarcane during rice stripe virus (RSV) infection are some striking examples (Zhang et al., 2019). A common miRNA, miR168, is upregulated by tobacco mosaic virus (TMV), potato virus X (PVX), and tobacco etch virus (TEV) infection (Várallyay et al., 2010). Exploration of virus-responsive sRNA profiles is crucial for unraveling plant–virus interactions.

Many sRNA molecules, including transacting siRNA, phased siRNA, and the repeat-associated siRNA, are substantial contributors in mediating plant host–virus interactions. NGS revealed the changes in the miRNAs expression profile of miR156, miR159, miR160, miR166, miR398, miR1511, miR1514, and miR2118 upon MYMIV infection in *Vigna mungo* plants, while four novel miRNAs, *viz*., vmu-miRn7, vmu-miRn8, vmu-miRn13, and vmu-miRn14, were also identified (Kundu et al., 2017). The NGS-based analysis of the miRNA profiles of two tomato varieties, Pusa Ruby and LA1777, upon tomato leaf curl virus infection, led to the identification of 53 novel miRNAs, 15 novel homologs, and 91 already known miRNAs (Tripathi et al., 2018). Another class of RNAs, long non-coding RNAs (lncRNAs), also play a pivotal role in host–virus interactions and can originate from either virus, plant, or both (Fortes and Morris, 2016). The functions of many sRNA species have not been explored yet but may contribute to better plant protection strategies. A complete virome analysis is a prerequisite to fully explore the sRNA profile of a plant with a mixed virus infection. Without the assistance of modern NGS technology, the analysis of complete virome would be a time-consuming, tedious, and chaotic task.

Grapevine samples collected from a vineyard in South Africa were analyzed by deep sequencing of total RNAs using the Illumina Genome Analyzer platform. This process successfully detected grapevine leafroll-associated virus 3 (GLRaV-3), grapevine rupestris stem pitting associated virus, and grapevine virus A (Coetzee et al., 2010). In another study, the *de novo* genome assembly of virus enriched sRNAs using Illumina Genome Analyzer IIx platform led to the discovery of a novel potyvirus tomato necrotic stunt virus (ToNSV) from tomato (Li et al., 2012). A complete virome of pepper (*Capsicum* species) plants comprising aphid lethal paralysis virus, bell pepper endornavirus, chilli leaf curl virus (ChiLCV), pea streak virus, pepper leaf curl Bangladesh virus, tobacco vein clearing virus, and a novel pepper virus A was reported using Illumina's HiSeq 2000 platform (Ng et al., 2011; Candresse et al., 2014; Dayaram et al., 2015; Liang et al., 2015; Jo et al., 2017; Hao et al., 2018; Bornancini et al., 2020). The development of sequencing platforms together with the application of bioinformatics tools proved to be a robust approach for the precise detection of viruses (Fonseca et al., 2018).

A metagenomics study unraveled the presence of sida golden mosaic Brazil virus, soybean blistering mosaic virus, tomato dwarf leaf virus, tomato yellow spot virus, tomato yellow vein streak virus, and a novel pepper blistering leaf virus in a single infection of pepper plants in Argentina (Bornancini et al., 2020). A vine plant exhibiting grapevine viral disease symptoms was subjected to NGS analysis, revealing a mixed infection involving different RNA viruses (Al Rwahnih et al., 2009). The study assessed the genomic and biological properties of a potyvirus, Bean yellow mosaic virus, isolated from *Lupinus angustifolius* plants with mild to severe symptoms, and two other plant species using the Illumina HiSeq2000 platform. The results showed the presence of one new virus and 23 new BYMV sequences. Based on these newly identified sequences, the phylogenetic evolutionary relationship was inferred, and a new nomenclature was proposed (Kehoe et al., 2014). Over the last few years, hundreds of new plant virus species have been reported through NGS technology (Barba et al., 2014; Ho and Tzanetakis, 2014; Roossinck et al., 2015; Wu et al., 2015). NGS has also been extremely useful in determining the host range and genetic diversity of plant viruses and in understanding their evolution (Roossinck, 2017). NGS has been employed to determine the mutational landscapes, particularly SNPs in the genomes of plant viruses (Kutnjak et al., 2017; Katsiani et al., 2020). Studying the genetic diversity of plant viruses helps devise robust strategies to circumvent viral infections and differentiate prevalent viral strains and discover new viral isolates (Marais et al., 2014).

NGS: IMPACT ON QUARANTINE PLANT AND VIRUS CHARACTERIZATION

The discovery and detection of new/unknown viruses have been augmented with the advent of NGS technologies. Prior to the execution of NGS, pistachio rosette virus was the only virus known to infect pistachio trees in Russia and Iran (Kreutzberg, 1940), but the NGS analysis revealed the presence of a new virus together with a virus-like agent provisionally named "pistachio ampelovirus A" and citrus bark cracking viroid-pistachio, respectively (Al Rwahnih et al., 2018).

It has been predicted that NGS will be used to monitor or control the dissemination of plant viruses and/or infected plant material across borders through quarantine measures (**Figure 2**). The increasing volume of trade, exchange of germplasm, and the diversity of plant material are significant threats due to the movement of plant viruses and phytopathogens across the globe. Plant quarantine and certification programs have been introduced to control the introduction of new viruses. In the past, the Plum pox virus spread across Europe and North America in 1990, and then an extensive mitigation program was introduced to control the virus (Welliver, 2012). Two very recent examples are cassava mosaic disease and groundnut rosette disease in peanut, occurring after cassava and groundnut were introduced during the 16th century into Africa from South America (Carter et al., 1997; Naidu et al., 1999). A novel mastrevirus, sugarcane white streak virus (family *Geminiviridae*), was detected in quarantined sugarcane plants in France via NGS (Candresse et al., 2014).

A novel marafivirus and luteovirus were also detected in quarantined nectarine plants when they were probed using NGS-mediated detection (Bag et al., 2015; Villamor et al., 2016). The etiology of two separate citrus diseases has been updated through NGS. The first disease was found to be associated with a new alphaflexivirus, citrus yellow vein clearing virus, following siRNA analysis in Turkish accessions (Loconsole et al., 2012a). Moreover, siRNAs and complete DNA sequences were also analyzed to discover a highly divergent monopartite geminivirus, citrus chlorotic dwarf-associated virus (Loconsole et al., 2012b). Using NGS platforms, quarantine and certification programs have become more sophisticated, enabling the quick and disease-free dissemination of plant material across the globe (**Figure 2**).

NGS AND CHARACTERIZATION OF PLANT VIRUSES FROM ARCHEOLOGICAL PLANT MATERIAL

The domestication of plants creates a new environment for the co-evolution of many pathogens. Most viral diseases have emerged recently and clues about their evolutionary pathway tracks have been lost in antiquity. The advent of NGS has enabled researchers to track the evolutionary footprints of such viral pathogens at the molecular level by sequencing ancient genomes from archeological materials. The re-construction of an archeological virus genome, that of barley stripe mosaic virus (BSMV), from a 750-year-old barley grain revealed that the divergence between the BSMV and its closest relative had taken place about 2000 years ago (Smith et al., 2014). Similarly, archeological material discovered in an approximately 50–168-year-old herbarium was found to contain peach latent mosaic viroid disease (Guy, 2013). Ancient maize cobs, dating from approximately 1000 CE from Antelope House in Arizona were found to contain the novel ds RNA plant virus genome of *Zea mays* chrysovirus 1 (ZMCV1), belonging to the plant and fungi infecting family, *Chrysoviridae* (Peyambari et al., 2019). These NGS analyses offer substantial new perspectives and knowledge on the evolutionary history of plant viruses and have helped to establish several new genera of plant viruses, such as the genera *Bymovirus*, *Macluravirus*, *Ipomovirus*, *Rymovirus*, and *Tritimovirus* (Gibbs et al., 2008).

CRISPR-CAS-MEDIATED RESISTANCE TO PLANT DNA VIRUSES

After successful detection, the development of broad-spectrum resistance can potentially limit infection by prevalent virus species and their variants. CRISPR-mediated viral immunity can be conferred directly by explicitly designing gRNA against the target virus(es) or indirectly by editing the host-susceptibility or resistance genes (**Figure 3**).

CRISPR-mediated resistance has successfully been executed against single-stranded (ss) DNA geminiviruses that replicate through dsDNA intermediates in the nuclei of host cells. The CRISPR-Cas system was first employed against two different mastreviruses, bean yellow dwarf virus (BYDV) (Baltes et al., 2015) and beet severe curly top virus (BSCTV) (Ji et al., 2015) in *Nicotiana benthamiana* and *Arabidopsis* plants, respectively. Both studies showed an up to 87% reduction in virus accumulation. A virus-inducible CRISPR-Cas system was later developed that transiently inhibited the BSCTV accumulation in *N. benthamiana* plants and the transgenic *Arabidopsis* plants with no off-target activity (Ji et al., 2018). In another study, tobacco rattle virus (TRV)-based vectors were used to express the CRISPR-Cas system into *N. benthamiana* plants to confer resistance against TYLCV by targeting three different regions of the TYLCV genome, including Rep protein, CP, and a non-coding intergenic region (IR). The construct targeting the IR region produced better results and more strongly inhibited virus replication than the other two constructs (Ali et al., 2015a). A similar approach targeting the IR region of cotton leaf curl disease-associated begomoviruses (CABs) has been proposed (Iqbal et al., 2016). A broad-spectrum resistance was achieved by targeting the IR region of three distinct begomoviruses, including cotton leaf curl Kokhran virus (CLCuKoV), TYLCV, and merremia mosaic virus (MeMV) (Ali et al., 2016). In another study, the simultaneous targeting of CLCuKoV-encoded Rep and βC1 of cotton leaf curl Multan betasatellite (CLCuMuB) via CRISPR-Cas9 system led to symptom attenuation and manyfold reduction in the viral titer (Khan et al., 2020). The simultaneous targeting of two regions, Rep and IR, of cotton leaf curl Multan virus (CLCuMuV) similarly produced almost complete resistance to CLCuMuV in *N. benthamiana* plants (Yin et al., 2019).

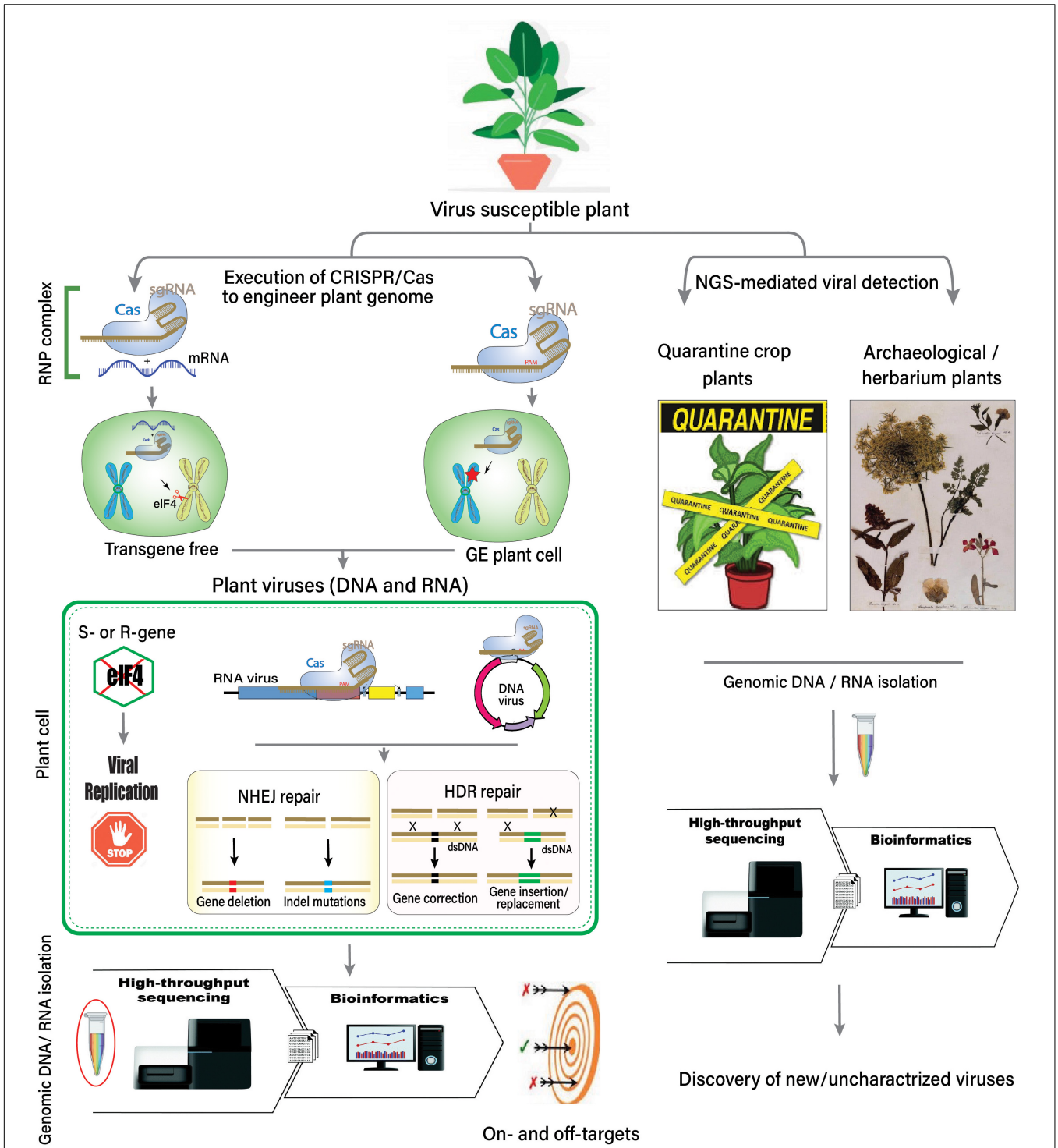


FIGURE 2 | Overview of the host–virus interaction, the role of HTS, CRISPR-Cas system, and the discovery of unknown viruses.

Recently, a new Cas protein, Cas14a, to target ssDNA has been described (Khan et al., 2019); although no practical applications of this protein have been demonstrated in plants, it could potentially confer comprehensive resistance to geminiviruses and nanoviruses. Nonetheless, CRISPR-mediated resistance is not always successful. The AC2 and AC3 genes of African cassava mosaic virus (ACMV) were targeted but this did not result in resistance. The authors claimed that ACMV variants emerged after NHEJ repair and those variants triggered the rapid evolution of the virus (Mehta et al., 2019). The application of CRISPR-Cas9

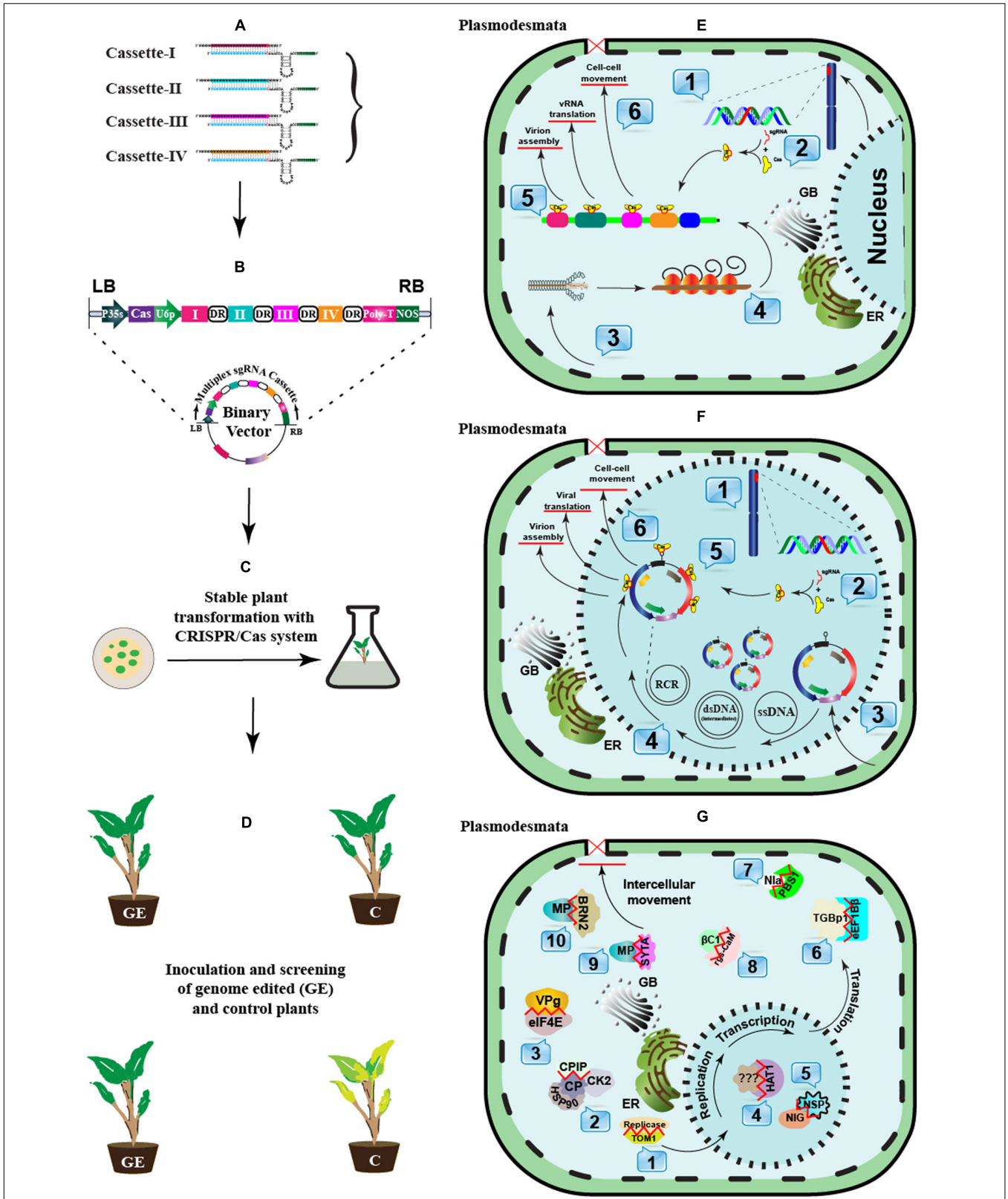


FIGURE 3 | A schematic approach to engineer CRISPR-Cas-based antiviral resistance in crop plants. **(A)** Designing of sgRNA to target the virus-encoded genes or the host susceptibility factors. **(B)** Construction of a cassette to express multiple sgRNAs and Cas protein under suitable promoters. The Cas protein can be

(Continued)

FIGURE 3 | Continued
expressed under CaMV-35S promoter and NOS terminator sequences, while sgRNAs can be expressed under the RNA polymerase-III promoter. Direct repeat (DR) sequences can follow the cloning of individual sgRNAss. **(C)** A suitable recombinant binary vector carrying the multiplexed sgRNAs cassette can be employed for the stable *in vitro* genetic transformations in plants. **(D)** The successful GE plants are tested for resistance against single or multiple plant viruses through mechanical, *Agrobacterium*- or vector-mediated plant inoculation techniques. The plants expressing effective CRISPR-Cas machinery would show resistance against the invading viruses. **(E)** *In planta* CRISPR-Cas-based genetic resistance model against potato virus Y (PVY): 1. CRISPR-Cas components are transcribed in the successfully transformed plants. 2. The Cas protein and respective sgRNAs are assembled to make a complex. 3. Mechanical or vector-mediated transmission of PVY in the primary plant cell. 4. Uncoating and subsequent host-mediated translation of the viral RNA genome. Multiple viral proteins help in viral replication at the periphery of the endoplasmic reticulum. 5. The Cas:sgRNA complex recruits and subsequently binds to the targeted PVY genome. 6. The activation of the Cas:sgRNA complex disrupts the genes involved either in viral translation, virion assembly, or long-distance movement through plasmodesmata. **(F)** *In planta* CRISPR-Cas-based genetic resistance model against an ssDNA plant virus, cotton leaf curl Kokhran virus (CLCuKoV), and it associated DNA-satellites. 1. CRISPR-Cas components are transcribed in the successfully transformed plants. 2. The Cas protein and respective sgRNAs are assembled to form a complex. 3. *Agrobacterium*- or vector-mediated transmission of CLCuKoV and associated DNA-satellites in the primary plant cell. 4. The ssDNA viral genome replication in the cellular nuclei following rolling circle replication (RCR) via dsRNA intermediates. The Replication associated protein (Rep) initiates RCR and generates free 3'-hydroxyl end to prime ssDNA synthesis by nicking dsDNA molecules. 5. The Cas:sgRNAs complex recruits and subsequently binds to the targeted genes encoded by CLCuKoV and/or DNA-satellites genomes, respectively. 6. The activation of the Cas:sgRNA complex cleaves the viral genes involved either in viral translation, virion assembly, or long-distance movement through plasmodesmata. **(G)** A recessive resistance model mediated by host factors against plant viruses (RNA or DNA). During the infection cycle, plant viruses interact with several host factors to endure their successful infection. Many host factors are known as promising candidates in antiviral breeding, which do not perturb the plant development or growth if mutated. These host factors aid virus replication (1–3), transcription (4–5), translation (6–8), or intercellular long-distance movement (9–10). These host factors can be targeted through CRISPR-Cas based GE to break their interaction with viral proteins.

has been extended to the dsDNA genome of cauliflower mosaic virus (CaMV) and a significantly lower viral titer was achieved in *Arabidopsis* plants by multiplexed targeting of the CP region (Liu et al., 2018).

CRISPR-CAS-MEDIATED RESISTANCE TO PLANT RNA VIRUSES

RNA viruses are the most diverse entity in the biosphere, and plant RNA viruses cause enormous economic losses to crop productivity worldwide. The advent of FnCas9, Cas13, and CasRx paved the way to restrict RNA virus infection in plants. The first-ever study to use two RNA viruses, TMV and CMV, expressed FnCas9 in *N. benthamiana* and *Arabidopsis* plants. Multiplexed GE was commenced by designing multiple gRNAs to target multiple regions of both viruses simultaneously. The results showed a 40–80% reduction in viral accumulation (Zhang et al., 2018). A theoretical approach to confer broad-spectrum resistance against PVY has been proposed recently (Hameed et al., 2019). Cas13 has successfully conferred immunity against plant-infecting RNA viruses. The first example of Cas13 utility was reported in 2018 against TuMV by transiently and stably expressing Cas13a in *N. benthamiana* plants (Aman et al., 2018a) but promising results could not be achieved. Later, the same researchers employed different Cas13 versions, including Cas13a, Cas13b, and CasRx (Cas13d), and the results revealed the robustness of CasRx against RNA viruses (Mahas et al., 2019a). CRISPR-Cas13 system was engineered in *N. benthamiana* plants to harness resistance against TMV and a significant attenuation in infection was associated with reduced viral titer (Zhang et al., 2020). Cas13 has also been successfully employed to engineer resistance in rice plants against two members of the negative-sense ss RNA rhabdoviruses, southern rice black-streaked dwarf virus (SRBSDV; genus *Fijivirus*) and rice stripe mosaic viruses (genus *Cytorhabdovirus*) (Zhang et al., 2019). In addition, Cas13-mediated resistance against a positive-sense ssRNA virus, potato virus Y (PVY, genus *Potyvirus*, family *Potyviridae*) has been achieved (Zhan et al., 2019).

CRISPR-CAS-MEDIATED HOST GENOME EDITING TO ENGINEER VIRAL RESISTANCE

CRISPR-Cas-mediated viral resistance can also be achieved by either mutating/inactivating the host susceptibility factors (S-genes) or expressing/activating resistance factors (R-genes). This approach involves commencing CRISPR-Cas-mediated modifications in the S- or R-genes. The transgenic sequences are then segregated from the progeny to yield non-transgenic plants. R-genes are usually linked to undesired traits such as poor flavor, low yield, or developmental abnormalities, meaning that they have the least potential in virus resistance. S-genes such as the *Arabidopsis ssi2* can confer resistance to CMV but exhibit similar problems including growth-related abnormalities. Despite this, most S-genes have promising potential for antiviral engineering. Inactivation of these susceptibility factors could lead to resistance without compromising plant general health due to the functional redundancy of the isoforms. Cao et al. (2020) categorized such S-genes into four groups (Cao et al., 2020). The first group comprised of negative regulators of plant defense like rgs-CaM (Li et al., 2018) and homeodomain leucine zipper protein 1 (HAT1) (Zou et al., 2016). The second group constitutes S-genes involved in different stages (such as replication/translation/movement) of the viral life cycle, for example, tobamovirus multiplication 1 (TOM1) and its homologs (Yamanaka et al., 2000), eEF1A and eEF4s, and Sec24a (a COPII coatomer) involved in replication, translation, and the movement of different viruses, respectively (Wang, 2015). Third group members interact and modify (phosphorylate) the viral proteins and include shaggy-related protein kinases (SK4-1, NsAK) (Lozano-Durán et al., 2011) and cellular

CK2 (Lõhmus et al., 2017). The candidate host factors in the fourth group positively affect the virus behavior, such as secondary cell wall synthesis factor (Bearskin2B, BRN2) and phenylpropanoid metabolism factor (4-coumarate:CoA ligase1, 4CL1; Lozano-Durán et al., 2011).

Translation initiation factor *eIF4E*, and its isoforms, are among the most exploited susceptibility factors by the RNA viruses to induce a successful *in planta* infection (Sanfaçon, 2015). By opting for this technique, eif4 mutant plants were developed by transient expression of the CRISPR system and, subsequently, mutant plants demonstrated resistance to multiple potyviruses, including the zucchini mosaic virus, cucumber vein yellowing virus, and papaya ringspot virus (Chandrasekaran et al., 2016). The same *eIF4E* susceptibility factor of *Arabidopsis* was knocked out to confer resistance against TuMV (Pyott et al., 2016) and *eIF4G* was mutated in rice to curb rice tungro spherical virus (RTSV; family *Sequiviridae*) (Macovei et al., 2018). SK4-1 interacts with phosphorylate geminiviral C4 proteins to suppress disease symptoms (Lozano-Durán et al., 2011). Cassava-encoded *eIF4E* genes, nCBP 1 and 2, interact with the VPG protein of the cassava brown streak virus, and these two genes were mutated via the CRISPR-Cas system in cassava plants. The edited plants showed delayed and attenuated symptoms (Gomez et al., 2019). HAT1 nutation led to a higher accumulation of salicylic acid and jasmonic acid and resulted in resistance to CMV (Zou et al., 2016).

ROLE OF DIFFERENT CAS VARIANTS IN GE AGAINST PLANT VIRUSES

Several versions of CRISPR-Cas systems have been employed to engineer the genomes of different plant species (**Table 2**) after executing either DNA (SpCas9, SaCas9, Cas12a, and Csm1) or RNA (C2c2/Cas13a) targeting endonucleases (**Figure 4**). Some recently discovered endonucleases for DNA (Cas14a, Cas12e, and C2c1) (Harrington et al., 2018) or RNA (Cas13b) genomes could also be good candidates to study GE in plants. As most of the destructive plant viruses have ssDNA or ssRNA genomes, Cas14a and Cas13b can be of prime interest due to their ability to target ssDNA and ssRNA genomes, respectively. The average size of most of the endonucleases ranges between 400 and 1368 amino acids. Nonetheless, employing the smaller sized endonucleases such as SaCas9, Csm1, and Cas12e may improve the efficiency of the delivery of CRISPR-Cas components to the plant genome. Moreover, these miniature endonucleases can be stand-alone CRISPR effectors for virus-mediated GE against plant viruses (details in the preceding section). Additionally, Cas12a, Csm1, C2c1, and Cas12e endonucleases are specific for dsDNA genomes and thus produce staggered ends, making them useful for the homology-dependent repair (HDR) pathway.

Cas9 Nuclease

Cas9 (earlier known as Cas5, Csn1, or Csx12) plays a significant role in the defense of certain bacteria against viruses and plasmids. Cas9 is an RNA-guided DNA endonuclease enzyme associated with the CRISPR adaptive immune system in bacteria. Despite its original function as a defense protein, Cas9 has been greatly exploited as a genome engineering tool to induce site-directed double-strand breaks in DNA. Cas9 induces dsDNA breaks in a sequence-specific manner and is directed to a target locus by a gRNA (Jinek et al., 2012; Gilbert et al., 2013; Maeder et al., 2013). The CRISPR-based GE cleavage complex requires tracrRNA besides Cas9 and crRNA, where crRNA and tracrRNA are mutually expressed as a sgRNA molecule (Jinek et al., 2012). CRISPR-Cas9 could be reprogrammed for another target site within the genome by changing the sequence of its crRNA.

Cas12a Nuclease

The Cas12a (Cpf1) nuclease belongs to the class-II type V CRISPR system. Contrary to the Cas9 nuclease, which requires two separate catalytic domains, *viz*., HNH and RuvC, Cas12a only need one catalytic domain (RuvC) for creating DSB in DNA. Cas12a only needs crRNA to be fully operational instead of crRNA and tracrRNA duo, which is a must for Cas9. The efficiency of Cas12a further improves as it produces sticky ends by generating 5′ overhangs during DSBs compared to the blunt end, a characteristic of the Cas9 system (Zetsche et al., 2015; Fonfara et al., 2016). The idea of the generation of 5' overhangs is further strengthened by the fact that HDR mechanisms prefer sticky ends (Vu et al., 2017; Schmidt et al., 2019).

Cas13 Nuclease

Cas13 proteins have recently been re-programmed to target ssRNA viral genomes infecting plants (Mahas et al., 2019b) and humans (Freije et al., 2019) without any known off-targets. The DNA-nuclease activity of Cas13 orthologs (Cas13a, Cas13b, and Cas13d) is replaced with two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains. The HEPN domains have exclusively transformed Cas13 as an RNA-guided RNA endonuclease to edit ssRNA targets specifically. Cas13 is explicitly a single effector nuclease, which acts as a dual-effector nuclease mediating the processing and maturation of precursor-CRISPR (pre-CRISPR) RNA into CRISPR RNA (crRNA), respectively. Thus, a single crRNA of Cas13 proteins mediates the recognition of target ssRNAs following an RNA–RNA hybridization (Wessels et al., 2020). The two HEPN domains mediate the nuclease function of Cas13a, preferably at U-rich targets. Empirically, a single point mutation at any HEPN domain can abolish the nuclease activity of Cas13a, resulting in a catalytically inactive dead Cas13a (dCas13a). Interestingly, dCas13a can retain the RNA-binding ability with higher sensitivity. Contrarily, the sequence-specific RNA cleavage activity of Cas13b depends upon double-sided protospacer flanking sequence (PFS) sites and ssRNA templates (Hameed et al., 2019). Cas13 nucleases may commit some non-specific nuclease activity; however, their off-target mutations are greatly reduced compared to RNAi (Abudayyeh et al., 2017). Although the PFS constraints are minimal in the Cas13 gRNAs design, the spacer sequences are mostly intolerant for any mismatches between 15 and 21 nt (Wessels et al., 2020). Thus, the applications of catalytically active Cas13a have been extended into the plant genomes and targeting ssRNA viruses such as potyviruses (family *Potyviridae*).

TABLE 2 | Examples of CRISPR-Cas-mediated genome editing into different plant hosts against different plant infecting viruses.

CRISPR-Cas system	Targeted virus(s)*	Genome	Genus	Family	Experimental host	References
CRISPR-Cas9	BSCTV	ssDNA	Curtovirus	*Geminiviridae*	*A. thaliana and N. benthamiana*	Ji et al., 2015
	CaLCuV	ssDNA	Begomovirus	*Geminiviridae*	Cabbage	Yin et al., 2015
	BSCTV	ssDNA	Curtovirus	*Geminiviridae*	*A. thaliana* and *N. benthamiana*	Ji et al., 2015
	BeYDV	ssDNA	Mastrevirus	*Geminiviridae*	*N. benthamiana*	Baltes et al., 2015
	TYLCV	ssDNA	Begomovirus	*Geminiviridae*	*N. benthamiana*	Ali et al., 2015a
	CVYV, ZYMV and PRSV-W	ssRNA (+)	Ipomovirus, Potyvirus	*Potyviridae*	*Cucumis sativus*	Chandrasekaran et al., 2016
	TuMV	ssRNA (+)	Potyvirus	*Potyviridae*	A. thaliana	Pyott et al., 2016
	TYLCV, MMV, BCTSV and BYDV	ssDNA	Begomovirus, Curtovirus, Mastrevirus	*Geminiviridae*	*A. thaliana* and *N. benthamiana*	Hadidi et al., 2016
	TRV and PEBV	ssRNA (+)	Tobravirus	*Virgaviridae*	*A. thaliana* and *N. benthamiana*	Ali et al., 2018
	TYLCV	ssDNA	Begomovirus	*Geminiviridae*	Tomato	Ghorbani et al., 2020
	Middle East-Asia Minor I (MEAM1)				*B. tabaci*	Heu et al., 2020
	ACMV	ssDNA	Begomovirus	*Geminiviridae*	*N. benthamiana*	Mehta et al., 2019
	CMV and TMV	ssRNA (+)	Cucumovirus Tobamovirus	*Bromoviridae, Virgaviridae*	*A. thaliana* and *N. benthamiana*	Zhang et al., 2018
	CaMV	dsDNA	Caulimovirus	*Caulimoviridae*	*A. thaliana*	Liu et al., 2018
	RTSV	ssRNA (+)	Waikavirus	*Sequiviridae*	*Oryza sativa*	Macovei et al., 2018
	TYLCV	ssDNA	Begomovirus	*Geminiviridae*	*N. benthamiana* and tomato	Tashkandi et al., 2018
	SMV	ssRNA (+)	Potyvirus	*Potyviridae*	Soya bean	Zhang et al., 2019
	TRV	ssRNA (+)	Tobravirus	*Virgaviridae*	*N. benthamiana*	Mahas et al., 2019a
	WDV	ssDNA	Mastrevirus	*Geminiviridae*	Barley	Kis et al., 2019
	ChiLCV	ssDNA	Begomovirus	*Geminiviridae*	*N. benthamiana*	Roy et al., 2019
CRISPR-Cas13/a	PVY	ssRNA (+)	Potyvirus	*Potyviridae*	Potato	Zhan et al., 2019
	TuMV	ssRNA (+)	Potyvirus	*Potyviridae*	*A. thaliana*	Aman et al., 2018b
	TuMV	ssRNA (+)	Potyvirus	*Potyviridae*	*N. benthamiana*	Aman et al., 2018a
	PVY	ssRNA (+)	Potyvirus	*Potyviridae*		Hameed et al., 2019
	RNA viruses	ssRNA (+)	Potyvirus	*Potyviridae*	*N. benthamiana*	Chaudhary, 2018
	Plant pathogens/pest				Soybeans	Abudayyeh et al., 2019
CRISPR-Cas14/a	Geminiviruses	ssDNA	Begomovirus	*Geminiviridae*		Khan et al., 2019
	Geminiviruses	ssDNA	Begomovirus	*Geminiviridae*		Aquino-Jarquin, 2019
	Geminiviruses	ssDNA	Begomovirus	*Geminiviridae*		Harrington et al., 2018
CRISPR-CasRx	RNA viruses	ssRNA (+)	Potyvirus	*Potyviridae*	*N. benthamiana*	Mahas et al., 2019a

**Viruses acronyms used are: African cassava mosaic virus (ACMV), bean yellow dwarf virus (BeYDV), beet curly top virus (BCTV), beet severe curly top virus (BSCTV), cabbage leaf curl virus (CaLCuV), chilli leaf curl virus (ChiLCV), cauliflower mosaic virus (CaMV), cucumber vein yellowing virus (CVYV), cucumber mosaic virus (CMV), merremia mosaic virus (MMV), papaya ring spot mosaic virus-W (PRSV-W), pea early browning virus (PEBV), potato virus Y (PVY), rice tungro spherical virus (RTSV), soybean mosaic virus (SMV), tobacco mosaic virus (TMV), tobacco rattle virus (TRV), tomato yellow leaf curl virus (TYLCV), turnip mosaic virus (TuMV), wheat dwarf virus (WDV), and zucchini yellow mosaic virus (ZYMV). Virus genomes are denoted as single-stranded (ss) DNA and a positive sense (+) ssRNA.*

These studies showed that Cas13a conferred target-specific resistance against TuMV in *Arabidopsis* and *N. benthamiana* plants (Aman et al., 2018a; Mahas et al., 2019a). Both Cas13a and Cas13b showed robust and highly specific interference against multiple RNA viruses (Mahas et al., 2019a), whereas Cas13d conferred the most robust interference against RNA viruses in transient assays as well as *in planta*, respectively (Mahas et al., 2019b). Moreover, Cas13d has been shown to target two RNA viruses simultaneously and could be a good choice in enhancing broad-spectrum resistance against single and multiple viruses in plants.

Cas14 Nuclease

It was recently discovered that Cas14 nucleases are effector proteins in extremophile archaea, which cleave only ssDNA instead of dsDNA or ssRNA (Harrington et al., 2018). These class 2, type V nucleases are notably smaller in size (400–700 amino acids) compared to other class 2 counterparts. Unlike other class 2 Cas effectors, PAM binding is not an absolute requirement for Cas14 nucleases for their cleavage activity (Savage, 2019). Nevertheless, the gRNA must have a 20 bp sequence complementary to the target ssDNA. Cas14 nucleases require certain sequence specificity in the central

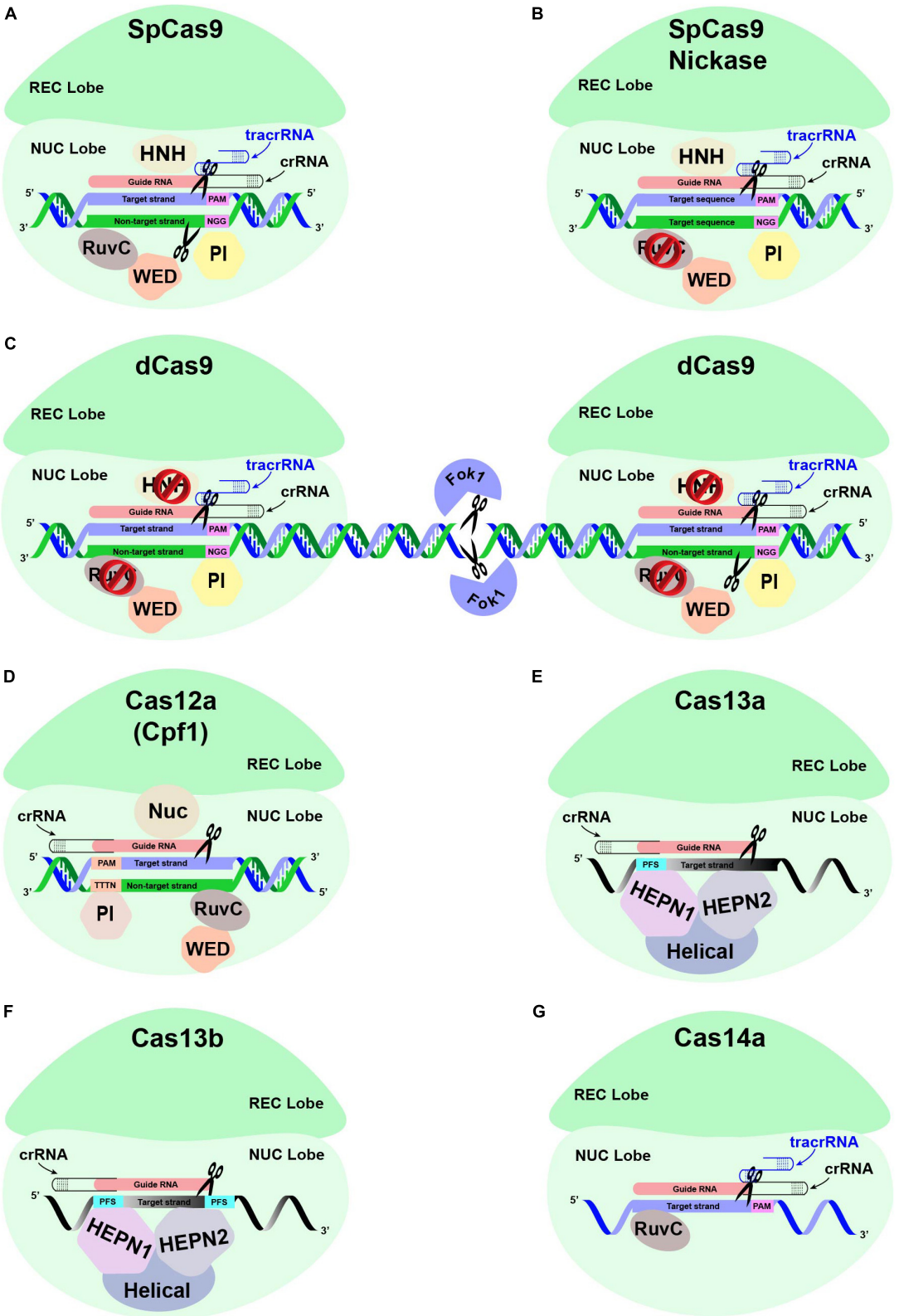


FIGURE 4 | Different Cas proteins that have been opted widely to engineer antiviral resistance in plants. **(A)** The conventional CRISPR-Cas9-based genome editing is mediated through a single effector protein Cas9, the crRNA, and a tracrRNA. The crRNA and tracrRNA associated with Cas9 endonuclease are hybridized and

(Continued)

FIGURE 4 | Continued
subsequently bind to the target region with the help of ∼20 nucleotide (nt) guide RNA (gRNA) sequence upstream of the PAM sequence, respectively. The recognition lobe (REC) is responsible for recognizing this crRNA:tracrRNA:target DNA complex. The PAM interacting domain (PI) recognizes the PAM sequence. The HNH and RuvC domains in the NUC lobe cleave the target strand and the non-target strand through a blunt-ended double-stranded break (DSB) at the upstream of the PAM sequence, respectively. **(B)** The conventional Cas9 has been modified to reduce off-target mutations. SpCas9 is generated by introducing point mutations in one of the two nuclease domains, RuvC, which produces single-stranded breaks (SSB) rather than DSB. **(C)** The specificity of Cas9 protein is also enhanced by fusing catalytically inactive "dead-Cas9" (dCas9) to an RNA-guided Fok1 nuclease. **(D)** Unlike CRISPR-Cas9, the CRISPR-Cas12a-based GE is mediated through a single effector protein (Cas12a) associated with a single crRNA. The REC lobe recognizes the Cas12a:crRNA complex, which subsequently binds to the target strand specifically at the target region downstream of the PAM sequence. The PI domain recognizes the PAM sequence of ∼23–25 nt long gRNA, thus helping specific DNA-binding. A DSB is introduced at the target region, with the Nuc domain responsible for cleaving the target strand 18 nt downstream of the PAM, while the non-target strand is cleaved by RuvC domain 23 nt downstream of the PAM, respectively. **(E)** The CRISPR-Cas13-based GE is mediated via a single effector protein (Cas13) associated with single crRNA. The REC lobe is responsible for recognizing the Cas13 and crRNA complex, which binds to the recognition site based on sequence complementarity with the ssRNA substrate directed by the gRNA sequence of crRNA. The sequence-specific cleavage of the ssRNA substrate is mediated through HEPN1 and HEPN2 domains. Instead of PAM sequences, Cas13a protein is directed toward the ssRNA target via a single protospacer flanking sequence (PFS). **(F)** The CRISPR-Cas13b is distinct from Cas13a due to the presence of a suppressor and enhancer Cas genes directed by two PFS. **(G)** The CRISPR-Cas14a-based ssDNA GE is mediated via single-effector Cas14a in association with crRNA and a 130-bp-long tracrRNA. The crRNA:tracrRNA complex does not require the presence of the PAM sequence in the ssDNA substrates. Nevertheless, ssDNA cleavage by the RuvC domain requires sequence-specific complementarity of 20 nt in the crRNA guide sequence to the ssDNA substrate.

6 bp stretch of the gRNA to be activated. Moreover, RNase-III endonuclease activity is missing in the Cas14 system (Harrington et al., 2018). The Cas14-mediated ssDNA cleavage can help to confer resistance against viruses with an ssDNA genome or mobile genetic elements (MGEs) such as viruses, plasmids, and transposons (Aquino-Jarquin, 2019). These characteristics make Cas14 nucleases uniquely suitable for nucleic acid detection. More specifically, Cas14 is capable of recognizing SNPs with high-fidelity in the targeted region.

ROLE OF NGS IN THE SCREENING OF GE EVENTS AND THE EVALUATION OF GE EFFICIENCY

Next-generation sequencing and CRISPR-Cas systems can be executed one after the other while conferring resistance to plant viruses. After successfully editing the susceptibility factor, eIF4E4, of the *Arabidopsis* genome, Pyott et al. (2016) performed extensive screening to yield transgene-free plants followed by developing homozygous mutations into T2 and T3 generations. Similarly, Macovei et al. (2018) carried out extensive screening to produce the transgene-free GE rice plants.

Although CRISPR-mediated plant genome engineering shares unparalleled advantages, it still poses two major pitfalls that need to be addressed in every successful event and are a major constraint in commercializing GE crops. The first challenge is the presence of a transgene, the second is unwanted (off-targets) mutations, and both pose a high risk to biosecurity. In plants with a shorter life span, it is easy to remove most of the CRISPR-Cas cargo via backcrossing and/or screening of segregating populations, but this approach is impractical for vegetatively propagated and longer life-spans plants. Additionally, achieving transgene-free plants is trickier as it may take years to get transgene-free seeds. One effective way to reduce the transgene integration is to express CRISPR-Cas system transiently. However, complete elimination is highly unlikely and the resulting degraded DNA fragments can still be integrated into the plant genome (Zhang et al., 2016). Therefore, developing transgene-free plants without any off-targets is significant, as this will persuade common customers and policymakers that the genome of GE plants is comparable to the native plants. To address these notions, NGS offers CRISPR-induced mutations (both on- and off-target), with a quick selection of transgene-free GE plants. Thus, it is unbiased, direct, and potentially can identify SNPs, indels, and different kinds of structural variants such as genome shuffling, inversions, and duplication (Manghwar et al., 2020).

The NGS platform has been exploited to detect on- and off-target changes in a variety of plant species including rice (Zhang et al., 2014; Tang et al., 2018), *Arabidopsis* (Feng et al., 2014; Xu et al., 2019), cotton (Li et al., 2019b), and tomato (Nekrasov et al., 2017). An assay compared the editing efficiency of T7E1, Indel Detection by Amplicon Analysis (IDAA), Tracking of Indels by Decomposition (TIDE), and targeted NGS in mammalian cells, where T7E1 often does not produce precise results but NGS does. TIDE and IDAA can also skip the alleles in edited clones compared to NGS (Sentmanat et al., 2018). Based on Illumina sequencing followed by a high-resolution melting (HRM) analysis, a cost-effective, rapid, and high-throughput mutant screening protocol was developed in tetraploid tobacco plants. In this study, CRISPR-mediated GE of *Phytoene desaturase* (PDS) gene was achieved, followed by regeneration and sequencing. The results of one study revealed that 17.2% of plants were non-transgenic, so the established method was unprecedented in the development of non-transgenic GE plants without segregating out the transgenes via sexual reproduction (Chen et al., 2018).

To improve the role of host genetic resources to confer resistance against plant viruses, the screening of host factors is essential. However, traditional approaches such as random mutagenesis and introgression through crossings are labor-intensive, costly, and difficult to execute due to functional redundancy. Moreover, these host factors have an essential role in plant viability and can lead to lethal phenotypes (Nicaise et al., 2007; Patrick et al., 2014). Nevertheless, NGS and CRISPR-Cas have opened new avenues for identifying, improving, and executing genetic resources for host-factors-mediated viral resistance. Such loci have been identified via

NGS from *Arabidopsis* and *Gossypium hirsutum* (Schneeberger et al., 2009; Austin et al., 2011; Hashimoto et al., 2016). Most of the S-genes (discussed in section "CRISPR-Cas Mediated Host Genome Editing to Engineer Viral Resistance" and **Figure 3G**) are excellent candidates for antiviral engineering due to their functional redundancy and isoforms.

APPLICATIONS OF CRISPR-CAS-BASED BIOSENSING TECHNOLOGY IN PLANT VIROLOGY

The CRISPR-Cas-based system can be harnessed to develop biosensing systems. Cas proteins can be fused with a split enzyme or a split fluorescent protein to build a biosensing system. Various CRISPR orthologs have been optimized to develop a cost-effective, highly sensitive, and pathogen-specific diagnosis of infectious and non-infectious diseases (Li L. et al., 2019). These biosensing platforms are DNA- or RNA-based CRISPR-Cas effectors. A remarkable collateral activity of Cas12a is the complete cleavage of ssDNA molecules during its binding to the target dsDNA. This non-specific cleavage activity is the basic principle of DNA endonuclease targeted CRISPR trans-reporter (DETECTR) development, where Cas protein (Cas12a or Cas14) is combined with the isothermal amplification technique such as recombination polymerase amplification to detect target DNA that subsequently cleaves ssDNA sequences coupled to fluorescent reports. DETECTR has been successfully employed in clinical diagnostics (Ai et al., 2019). Other applications of Cas12 have led to the development of 1-h low-cost multipurpose highly efficient system (HOLMES) and its derivative HOLMESv2 for rapid detection of target DNA/RNA and SNPs, respectively (Li L. et al., 2019). HOLMESv2 has been further improved for accurate quantification of DNA methylation in combination with Cas12b nuclease. The high-throughput application of Cas12 nuclease has been further expanded to develop CaT-SMelor (CRISPR-Cas12a- and a transcription factor-mediated small molecule detector) to detect small molecules and to discriminate them from their structural analogs (Liang et al., 2015). Furthermore, Cas14 has expanded promising high-fidelity SNPs genotyping tools to detect the ssDNA viruses infecting eukaryotes (Aquino-Jarquin, 2019). More interestingly, the ability of Cas14 to detect ssDNA targets independent of PAM recognition makes it an excellent candidate for detecting ss- and dsDNA. The non-specific DNase activity of Cas14 has been utilized to develop the Cas14-DETECTR method, which is more specific and active than its counterpart (Cas12a-DETECTR) as a high-fidelity system for DNA SNPs (Harrington et al., 2018). Cas14-DETECTR can be exploited for high-fidelity SNPs genotyping and detection of ssDNA viruses. Finding low abundance sequences by hybridization (FLASH) is another next-generation diagnostic tool based upon the specificity of Cas9 to efficiently enrich specific target sequences (Quan et al., 2019).

The discovery of Cas13 and its derivative nuclease has led to a variety of RNA-based applications in different systems due to its dual enzymatic activity (i.e., pre-crRNA processing and signature HEPN domain). These include various biosensing applications such as the detection of viral RNA (Gootenberg et al., 2017; Mahas et al., 2019b), precise RNA editing (Cox et al., 2017), site-directed mRNA demethylation (Li et al., 2019a), dynamic real-time RNA imaging (Wang et al., 2019), and site-specific polyadenylation in eukaryotic cells (Anderson et al., 2019). During the binding and cleavage of target RNA, the activation of Cas13 triggers random collateral trans-ssRNA cleavage in its vicinity. East-Seletsky et al. (2016) used the collateral ssRNA cleavage activity of Cas13 to detect the presence of specific nucleic acids by constructing reporter RNAs and the release of fluorescent signals upon recognition and cleavage (East-Seletsky et al., 2016). Further refinement in this technique led to the SHERLOCK development (specific high-sensitivity enzymatic reporter unlocking) detection method (Gootenberg et al., 2017; Kellner et al., 2019). The SHERLOCK method has been successfully utilized to detect specific strains of Dengue and Zika viruses. A more refined version of SHERLOCK is heating unextracted diagnostic samples to obliterate nucleases (HUDSON) which allows the detection of a low titer of a pathogen in a biological sample (Myhrvold et al., 2018). Further improvement in the SHERLOCK system resulted in SHERLOCKv2, which combined different Cas proteins (such as Cas12 and Cas13) in a single reaction, enabling multiplexed nucleic acid detection. The sensitivity of SHERLOCKv2 was further improved by joining Csm6 nuclease to amplify the signal of Cas13 collateral cleavage and the development of a FAM-biotin reporter kit (Gootenberg et al., 2018). The application of SHERLOCKv2 has been further expanded for various agricultural applications such as genotyping and quantification of genes related to pathogen resistance (Abudayyeh et al., 2019). The SHERLOCKv2 can be optimized exponentially for the detection of vital traits, surveillance of insect pests and disease, or other agricultural applications (Aman et al., 2020). The amenability of CRISPR-Cas-based multiplexing systems can enable biosensing techniques to identify multiple target nucleic acids (even at low concentrations) through a single diagnostic test kit. Further refinements may lead to cost-effective, super-sensitive, highly accurate, and on-field diagnostic kits with a range of applications in agriculture.

PLANT VIRUSES AS CRISPR DELIVERY VECTORS

The targeting efficiency of CRISPR-Cas depends on the efficient delivery of the CRISPR-cassette. This is usually achieved via biolistic or *Agrobacterium*-mediated plant transformation. In either case, only a limited number of cells are transformed and extra cargo like promoters and the presence of degrading DNA molecules give rise to GMO concerns. To this end, more efficient delivery methods are required to integrate the GE reagents effectively into economically important crops. The use of virus-based vectors with autonomous replication is popular for integrating GE cassettes into target plants with improved efficiency (Zaidi and Mansoor, 2017). Most importantly, these

viral vectors with either a DNA or RNA genome have been effectively employed to harness GE in model plants and commercial crops. Plant viruses are an attractive option for gene delivery to the host genome due to their intrinsic proximity with the host cells, their autonomous mode of replication, and smaller structure (Pasin et al., 2019).

Plant viruses are being used as vectors to express foreign proteins and RNAs (Cody and Scholthof, 2019). SSRNA viruses such as TRV and TMV, and ssDNA viruses of the family *Geminiviridae* such as BYDV, cabbage leaf curl virus, and wheat dwarf virus (WDV) have been employed for efficient GE. Geminiviruses have characteristics such as a wide host range, a single replication associated protein (Rep) for *in planta* replication, and the production of many replicons inside host cells (Hanley-Bowdoin et al., 2013). To overcome their limited cargo capacity, the geminiviruses were engineered as non-infectious but replicating systems. The movement protein (MP) and coat protein (CP) were substituted with the Cas protein and the sgRNA sequences (Čermák et al., 2015) to achieve hyper-expression of the CRISPR-Cas system; however, the absence of CP and MP limits their application for only transient expression at localized tissues. The MP can be expressed transiently using another vector or a bipartite genome of a begomovirus (which encodes MP on a separate genomic component) to overcome this problem. Another possibility is to engineer geminivirus associated DNA-satellites to deliver the GE components. These DNA-satellites have been successfully modified as virus-induced gene silencing (VIGS) vectors for several crops and can be a good choice as cargo vectors for GE reagents.

Tobacco rattle virus-based vectors are simple, versatile, and efficient GE tools that surpass long transformation and tissue culture procedures. Such vectors are equally effective for cloning, multiplexing, library construction, and agroinfiltration. The most important feature of TRV-based vectors is the possibility of producing transgene-free GE plants because their RNA genome is not integrated into the plant genome (Mahas et al., 2019a). However, the TRV-based CRISPR-Cas systems are limited to the transgenic lines of those crops in which Cas9 protein is stably expressed. TMV-based ssRNA vectors have also been used for the transient expression of certain genes and offer flexible gene expression and GE in a variety of crops (Cody et al., 2017). Kaya et al. (2017) developed a modified version of a TMV-based expression vector to express the target region of tobacco microRNA 398 (miR398). After plant transformation, during shoot development in the regenerated plants, the miR398 expression eliminated the viral RNA. Moreover, a split-protein approach has been used to transiently express the Cas9 protein from two fragments via TMV and *Agrobacterium*. The active Cas9 protein was re-assembled and ultimately induced a targeted mutation producing virus-free GE plants via tissue culture (Kaya et al., 2017). The utility of viral vectors for GE has opened a new era of functional genomics and applications in agricultural biotechnology. However, disposing of the viral genome from the GE plants may pose a serious challenge at a later stage. The use of meristematic tissue culturing can be a possible solution to eliminate the virus and its remnants from GE plants.

BOTTLENECKS IN ADOPTION OF NGS AND GE TOOLS IN PLANT VIROLOGY

Despite these innovative applications, CRISPR-Cas-based GE still has some limitations in terms of translational research applications, especially in engineering resistance to plant viruses. The efficient delivery of recombinant plasmids into the host genome followed by the successful regeneration of plants is a challenging task, particularly in the vegetatively propagated plants. Stable plant transformation is the key to regenerating transgenic plants with GE events and heritable mutations. Plasmid transgenes are usually segregated out of the developing progenies of the transformed plants at a later stage to make them transgene-free. However, this strategy does not apply to all crop plants. Moreover, if the designed sgRNAs are based on the potential plant virus genome, these cannot be segregated out from the developing plant progenies. Alternatively, the application of DNA-free GE techniques can be used. The popular biolistic and *Agrobacterium*-mediated plant transformation is not competent for many crop plants in terms of (1) low frequency of transformation, (2) prolonged tissue culturing procedures, (3) impaired tissues during biolistic delivery, (4) limited application of *Agrobacterium*-mediated plant transformation, (5) tissue browning and necrosis due to *Agrobacterium*, (6) somaclonal variations in the regenerated plants, and (7) insufficient DNA-delivery to induce HDR. The optimization of certain explants and culturing media are required to bypass these problems (Altpeter et al., 2016).

A major setback of GE in plants is the primary mode of DNA-repair via NHEJ, which produces many unwanted genetic mutations compared to HDR. The most significant cellular DNA-repair pathway during GE is HDR, which requires high titers of nucleases and repair templates delivered into the targeted plant tissues or cells. Enhancement of the efficacy of HDR in plants during a GE event is required. Plant viral-mediated transformation strategies can be a good alternative to deliver nucleases and sgRNAs into the host plant genome for increased HDR frequency during GE. TRV-based plant transformation vectors have been developed recently for successful GE in many plant species (Ali et al., 2015b). TRV-based transformations are crucial because the RNA genome of TRV vectors could not be integrated into the plant genome and could lead to transgene-free plant transformations (Mahas et al., 2019a). The use of a marker-free transformation technique using regeneration promoting factors (cycD3, auxin, and cytokinin-related genes) can also be used to apply GE to a wide range of plant species (Vats et al., 2019).

CRITICAL ASSESSMENT OF CRISPR-CAS AND NGS TECHNIQUES IN PLANT VIROLOGY

Despite the popularity and tremendous success of the CRISPR-Cas technology, there are still some limitations to its application in crop plants. These include but are not limited to, off

target effects, the difficulty of execution in woody plants, low mutagenesis, inefficient delivery approaches, dependence on *in vitro* regeneration, persistent activity in subsequent generations, potential risk of transgene transfer to wild type population, risk of reversion of induced mutations via cross-pollination, and stringent GMO regulations.

Most of the resistance strategies targeting the virus genome have failed to provide laconic control due to their inherent inability to target multiple and synergistically interacting viruses. Other reasons may include the evolution of the viral genomes through mutations and plant viruses, which form sub-genomic components to evade host resistance. To circumvent such problems, CRISPR-Cas technology holds great potential as an effective antiviral technology that can not only target the viral genome at multiple sites but can also simultaneously target different, related, or unrelated virus genomes. Aside from directly targeting the viral genome, the CRISPR-Cas can be executed to nullify the host susceptibility factors and circumvent the problem of generating resistance to viral mutants.

Despite these numerous successful applications, the execution of CRISPR-Cas still faces many challenges. It requires a meticulous and critical approach to avoid erroneous design. Various strategies have been proposed to address off target activity, such as GC contents should be ideally between 40 and 60% to form stable DNA:RNA duplex in the gRNA and to enhance the on-target activity and destabilize off-target binding (Wang et al., 2014). Higher GC contents (65–80%) can lead to off-target activity. The length of gRNA can have a profound effect on GE efficiency and can lead to off-target/unwanted mutations; for example, the results of studies appraising the 16-to-20 nucleotides long gRNA effect on GE efficiency and off-target mutations showed higher GE efficiency when 18–20 bp long sgRNAs were used (Fu et al., 2014; Sugano et al., 2018). Furthermore, the dead RNA off-target suppression (dOTS) technique employs dead truncated gRNA, which can guide Cas9 while suppressing cleavage, reducing off-target activity, and improving on-target activity by 40-fold (Rose et al., 2020). The chemical modification of gRNA, by incorporating 20-*O*-methyl-30-phosphonoacetate into the gRNA ribose-phosphate backbone, improves the on-target efficiency by 40–120-fold (Ryan et al., 2017).

Using a low concentration of Cas protein/gRNA is another potential way to minimize off-target effects. The expression of Cas9 under the control of CaMV35S (constitutive) promoter and an egg-cell (inducible) promoter (ECS) was evaluated. The results demonstrated that constitutive expression via CaMV35S promoter revealed a low editing frequency compared to ECS promoter (Begemann et al., 2017). Likewise, the use of embryo-specific promoters (YAO) yielded better and more efficient GE in *Citrus sinesis* (Zhang et al., 2017). The use of different Cas proteins variants has a substantial impact on the reduction of off-target effects. A comparative study of Cas9 and Cas12a revealed that the Cas9 mechanism is more specific, efficient, and accurate (90–100%) than Cpf1 (0–60%) in maize plants (Lee et al., 2019). Similarly, modified variants like dead or deactivated Cas variants (dCas) have fewer off target activities (Brocken et al., 2017).

The delivery of the CRISPR-Cas system is one of the vital factors for achieving better on-target and the least off target activity. Several transformation methods such as PEG-, *Agrobacterium*-, biolistic-, protoplast, ribonuclease protein (RNP) complex-, lipid- and polymer-, and viral vectors-mediated methods are being practiced. These approaches, however, are not free from drawbacks and share limitations. Another ailment to CRISPR-Cas utility is the persistence of Cas activity in subsequent generations that can induce unanticipated mutations in stable lines and has been reported in *Arabidopsis*, cotton, and maize plants (Feng et al., 2014; Char et al., 2017; Wang et al., 2018). To circumvent this problem, different strategies have been employed, such as the use of transgene-free viral vectors for cassette delivery and *in vitro* expression of pre-assembled gRNA:Cas (RNP) complex (Woo et al., 2015; Zhang et al., 2016). RNP complex can yield Cas protein-free GE cells because it is degraded quickly after cleaving the target site, meaning that it can address persistent concerns about Cas activity. However, degraded Cas protein and degraded DNA fragments can still induce some undesired mutations at a lower frequency. The Transgene Killer CRISPR (TKC) technique is another promising tool to yield transgene-free GE plants (He et al., 2018). The TKC technique uses the temporal expression of the Cas protein and suicide genes; the former is expressed first at the transformation (at callus formation and organogenesis) stage, while the latter is expressed at the embryogenesis stage to kill all the pollens and embryos with the transgene. Eventually, transgene-free plants are yielded without labor-intensive screening and selection. These technologies will minimize the regulatory GM burden, mitigate ecological challenges, and foster public acceptance of GE plants and related byproducts.

Base editors (BEs) are another substantial addition to the CRISPR toolbox that have enabled site-specific modification (base substitution) without inducing DSBs. BEs and prime editors (Pes) are moving to the front lines of precision genetic engineering. In base editing, all four transition mutations (C→T, G→A, A→G, and T→C) can be achieved. The Adenine BE can substitute A→G, and the Cytosine BE can substitute C→T at the target site. These BEs can be employed to confer antiviral resistance by introducing stop codons in the coding regions of viral genomes via iSTOP (Billon et al., 2017) or CRISPR-stop (Kuscu et al., 2017) technologies. The resulting viral proteins will be non-functional and limit viral spread. Similarly, base editing could be used to develop plants with immunity against different single and multiple pathogens by targeting and modifying the host susceptibility factors (S-genes).

High copy numbers in polyploidy plants pose unique challenges, such as knock-out of all copy numbers with equal efficiency and of genes with high homology. To achieve a plant with all copy number mutations, a series of allelic variants are first executed, and then a subsequent selection is performed in the segregating population (Vats et al., 2019). CRISPR-Cas-based GE has enabled easy editing and the introgression of multiple traits in polyploid plants without any linkage drag, which otherwise is a tedious and laborious task through conventional breeding approaches.

In the plant virus-combating arsenal, the interaction and battle between host plants and viruses resemble a never-ending arms race. Viral genomes are dynamic entities so that CRISPR-Cas-mediated resistance can facilitate and speed up the evolution and generation of new viral variants. Some studies have already reported such notions in ACMV (Mehta et al., 2019), CLCuKoV, MeMV, TYLCV (Ali et al., 2016), and CaMV (Liu et al., 2018). However, the combinations of two single gRNAs and/or targeting non-coding IR regions (in the case of geminiviruses) resulted in a substantial delay in resistance breakdown (Ali et al., 2016).

Next-generation sequencing, as a go-to tool for plant virologists, has shaped plant virology by sequencing whole virus genomes, undertaking plant metagenomics studies, and characterizing viruses from archeological and quarantined plant samples. Since GE requires pre- and post-knowledge of the target site, the former is needed to precisely carry out the GE, while the latter is required to evaluate both on- and off-target efficiency. In this context, NGS has been quite instrumental, but the availability of the whole genome is just limited to a few plant species (mostly model plants). The successful execution of CRISPR-Cas in plants requires knowledge of genetic variations, chromosomal rearrangements, indels, SNPs, transposon occurrence, and copy number variations. NGS and whole-genome sequences are thus a prerequisite not just in model systems, but for all plant species for which antiviral systems are engineered.

Next-generation sequencing can demonstrate the characterization of latent viruses or viruses that are of less concern regarding agricultural production. Nevertheless, if such viruses evolve, adapt, and become an emerging threat in the future, then NGS revelation will undoubtedly help to develop rapid diagnostic assays and better management strategies. One challenge, however, lies in the cost and processing of NGS data, which requires sophisticated machines, tools, and expert personnel.

CONCLUDING REMARKS

Although NGS technologies have been evolving swiftly over the last decade, there are a variety of parallel options being practiced, particularly for virus characterization. Historically, ELISA in the 1980s and later PCR-based approaches in the 1990s contributed to detecting plant viruses and disease etiologies. However, NGS has enabled the detection and characterization of novel plant viruses that have remained undetectable by conventional diagnostic approaches. Recent versions of NGS technologies such as PacBio by Illumina, Oxford Nanopore, and ISS may significantly boost plant virology by providing faster, more reliable virus detection with reduced errors and direct RNA sequencing.

The CRISPR-Cas toolbox has a range of tools for GE and is still expanding so that almost all types of viral genomes can be targeted/engineered. In the field of plant virology, CRISPR-Cas provides a versatile platform that can be engineered for biosensing, detection of small molecules, site-directed mutagenesis, genotyping, SNPs detection, gene quantification, and substitution of a single nucleotide. After successfully executing CRISPR-Cas systems in plant species, extensive screening to yield transgene-free plants is required and NGS platforms have been used to detect on- and off-targets in several crops. Based on Illumina sequencing followed by an HRM analysis, a cost-effective and high-throughput mutant screening protocol can be developed for different crop plants, as a similar approach has been furnished in tetraploid tobacco plants where the PDS gene was targeted. NGS, coupled with CRISPR-Cas, has already contributed to the control of plant viral diseases. In the near future, basic biological issues for antiviral engineering will be addressed through CRISPR-Cas-based technologies and the current GMO-related concerns of the common people may be nullified.

AUTHOR CONTRIBUTIONS

MSh, AR, MSa, and ZI conceived the ideas and wrote and finalized the manuscript. MSh drew tables. MSa and ZI drew figures. AA-S prepared the final version. Finally, all authors proofread and approved the manuscript.

FUNDING

This project was financially supported by Sultan Qaboos University (RF/AGR/CROP/19/01, RF/AGR/CROP/19/03, and IG/AGR/CROP/20/2) and VALE, Oman (EG/AGR/CROP/12/03).

ACKNOWLEDGMENTS

The authors thank Dr. Joanna Gress (Emporia State University, United States) for critically reading and improving the English language of the manuscript.

REFERENCES

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA targeting with CRISPR–Cas13. *Nature* 550, 280–284.

Abudayyeh, O. O., Gootenberg, J. S., Kellner, M. J., and Zhang, F. (2019). Nucleic acid detection of plant genes using CRISPR-Cas13. *CRISPR J.* 2, 165–171. doi: 10.1089/crispr.2019.0011

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353:aaf5573. doi: 10.1126/science.aaf5573

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

Adams, I. P., Skelton, A., Macarthur, R., Hodges, T., Hinds, H., Flint, L., et al. (2014). Carrot yellow leaf virus is associated with carrot internal necrosis. *PLoS One* 9:e109125. doi: 10.1371/journal.pone.0109125

Ai, J.-W., Zhou, X., Xu, T., Yang, M., Chen, Y., He, G.-Q., et al. (2019). CRISPR-based rapid and ultra-sensitive diagnostic test for Mycobacterium tuberculosis. *Emerg. MicrobesInfect.* 8, 1361–1369. doi: 10.1080/22221751.2019.1664939

Al Rwahnih, M., Daubert, S., Golino, D., and Rowhani, A. (2009). Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* 387, 395–401. doi: 10.1016/j.virol.2009.02.028

Al Rwahnih, M., Rowhani, A., Westrick, N., Stevens, K., Diaz-Lara, A., Trouillas, F. P., et al. (2018). Discovery of viruses and virus-like pathogens in pistachio using high-throughput sequencing. *Plant Dis.* 102, 1419–1425. doi: 10.1094/pdis-12-17-1988-re

Ali, Z., Abulfaraj, A., Idris, A., Ali, S., Tashkandi, M., and Mahfouz, M. (2015a). CRISPR-Cas9-mediated viral interference in plants. *Gen. Biol.* 16:238.

Ali, Z., Abul-Faraj, A., Li, L., Ghosh, N., Piatek, M., Mahjoub, A., et al. (2015b). Efficient virus-mediated genome editing in plants using the CRISPR-Cas9 system. *Mol. Plant* 8, 1288–1291.

Ali, Z., Ali, S., Tashkandi, M., Zaidi, S. S. A., and Mahfouz, M. M. (2016). CRISPR-Cas9-mediated immunity to geminiviruses: differential interference and evasion. *Sci. Rep.* 6:26912.

Ali, Z., Eid, A., Ali, S., and Mahfouz, M. M. (2018). Pea early-browning virus-mediated genome editing via the CRISPR-Cas9 system in *Nicotiana benthamiana* and *Arabidopsis*. *Virus Res.* 244, 333–337. doi: 10.1016/j.virusres.2017.10.009

Ali, Z., Zaidi, S. S. A., Tashkandi, M., and Mahfouz, M. M. (2019). A simplified method to engineer CRISPR/Cas9-mediated geminivirus resistance in plants. Antiviral resistance in plants. *Methods Mol. Biol.* 2028, 167–183. doi: 10.1007/978-1-4939-9635-3_10

Alok, A., Sandhya, D., Jogam, P., Rodrigues, V., Bhati, K. K., Sharma, H., et al. (2020). The rise of the CRISPR/Cpf1 system for efficient genome editing in plants. *Front. Plant Sci.* 11:264. doi: 10.3389/fpls.2020.00264

Altpeter, F., Springer, N. M., Bartley, L. E., Blechl, A. E., Brutnell, T. P., Citovsky, V., et al. (2016). Advancing crop transformation in the era of genome editing. *Plant Cell* 28, 1510–1520.

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., et al. (2018a). RNA virus interference via CRISPR-Cas13a system in plants. *Genome Biol.* 19:1.

Aman, R., Mahas, A., Butt, H., Ali, Z., Aljedaani, F., and Mahfouz, M. (2018b). Engineering RNA virus interference via the CRISPR-Cas13 machinery in *Arabidopsis*. *Viruses* 10:732. doi: 10.3390/v10120732

Aman, R., Mahas, A., and Mahfouz, M. (2020). Nucleic acid detection using CRISPR-Cas biosensing technologies. *ACS Synthetic Biol.* 6, 1226–1233. doi: 10.1021/acssynbio.9b00507

Anderson, K. M., Poosala, P., Lindley, S. R., and Anderson, D. M. (2019). Targeted Cleavage and polyadenylation of RNA by CRISPR-Cas13. *BioRxiv* [Preprint]. 531111.

Aquino-Jarquin, G. (2019). CRISPR-Cas14 is now part of the artillery for gene editing and molecular diagnostic. *Nanomedicine* 18, 428–431. doi: 10.1016/j.nano.2019.03.006

Austin, R. S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N. J., Bonetta, D., et al. (2011). Next-generation mapping of *Arabidopsis* genes. *Plant J.* 67, 715–725. doi: 10.1111/j.1365-313X.2011.04619.x

Bag, S., Al Rwahnih, M., Li, A., Gonzalez, A., Rowhani, A., Uyemoto, J. K., et al. (2015). Detection of a new luteovirus in imported nectarine trees: a case study to propose adoption of metagenomics in post-entry quarantine. *Phytopathology* 105, 840–846. doi: 10.1094/phyto-09-14-0262-r

Balasubramanian, S. (2015). Solexa sequencing: decoding genomes on a population scale. *Clin. Chem.* 61, 21–24. doi: 10.1373/clinchem.2014.221747

Baltes, N. J., Hummel, A. W., Konecna, E., Cegan, R., Bruns, A. N., Bisaro, D. M., et al. (2015). Conferring resistance to geminiviruses with the CRISPR–Cas prokaryotic immune system. *Nat. Plants* 1:15145.

Barba, M., Czosnek, H., and Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106–136. doi: 10.3390/v6010106

Begemann, M. B., Gray, B. N., January, E., Gordon, G. C., He, Y., Liu, H., et al. (2017). Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. *Sci. Rep.* 7:11606. doi: 10.1038/s41598-017-11760-6

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J., and Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* 32, 76–84.

Billon, P., Bryant, E. E., Joseph, S. A., Nambiar, T. S., Hayward, S. B., Rothstein, R., et al. (2017). CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol. Cell* 67, 1068–1079. doi: 10.1016/j.molcel.2017.08.008

Bornancini, V. A., Irazoqui, J. M., Flores, C. R., Vaghi Medina, C. G., Amadio, A. F., and López Lambertini, P. M. (2020). Reconstruction and characterization of full-length begomovirus and alphasatellite genomes infecting pepper through metagenomics. *Viruses* 12:202. doi: 10.3390/v12020202

Bortesi, L., and Fischer, R. (2015). The CRISPR-Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi: 10.1016/j.biotechadv.2014.12.006

Boykin, L. M., Sseruwagi, P., Alicai, T., Ateka, E., Mohammed, I. U., Stanton, J.-A. L., et al. (2019). Tree lab: portable genomics for early detection of plant viruses and pests in Sub-Saharan Africa. *Genes* 10:632. doi: 10.3390/genes10090632

Briddon, R., Mansoor, S., Bedford, I., Pinner, M., Saunders, K., Stanley, J., et al. (2001). Identification of DNA components required for induction of cotton leaf curl disease. *Virology* 285, 234–243. doi: 10.1006/viro.2001.0949

Brocken, D. J., Tark-Dame, M., and Dame, R. T. (2017). dCas9: a versatile tool for epigenome editing. *Curr. Issues Mol. Biol.* 26, 15–32. doi: 10.21775/cimb.026.015

Bronzato Badial, A., Sherman, D., Stone, A., Gopakumar, A., Wilson, V., Schneider, W., et al. (2018). Nanopore sequencing as a surveillance tool for plant pathogens in plant and insect tissues. *Plant Dis.* 102, 1648–1652. doi: 10.1094/pdis-04-17-0488-re

Burstein, D., Harrington, L. B., Strutt, S. C., Probst, A. J., Anantharaman, K., Thomas, B. C., et al. (2017). New CRISPR-Cas systems from uncultivated microbes. *Nature* 542, 237–241. doi: 10.1038/nature21059

Candresse, T., Filloux, D., Muhire, B., Julian, C., Galzi, S., Fort, G., et al. (2014). Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLoS One* 9:e102945. doi: 10.1371/journal.pone.0102945

Candresse, T., Marais, A., Faure, C., and Gentit, P. (2013). Association of Little cherry virus 1 (LChV1) with the Shirofugen stunt disease and characterization of the genome of a divergent LChV1 isolate. *Phytopathology* 103, 293–298. doi: 10.1094/phyto-10-12-0275-r

Cao, Y., Zhou, H., Zhou, X., and Li, F. (2020). Control of plant viruses by CRISPR/Cas system-mediated adaptive immunity. *Front. Microbiol.* 11:593700. doi: 10.3389/fmicb.2020.593700

Carter, S., Fresco, L., Jones, P., and Fairbairn, J. (1997). *Introduction and Diffusion of Cassava in Africa: IITA Research Guide, No. 49*. Ibadan: International Institute of Tropical Agriculture.

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biol.* 16:232.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non−transgenic cucumber using CRISPR-Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi: 10.1111/mpp.12375

Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2015). Boosting plant immunity with CRISPR-Cas. *Genome Biol.* 16, 4–7.

Char, S. N., Neelakandan, A. K., Nahampun, H., Frame, B., Main, M., Spalding, M. H., et al. (2017). An *Agrobacterium*-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnol. J.* 15, 257–268. doi: 10.1111/pbi.12611

Chaudhary, K. (2018). CRISPR-Cas13a targeting of RNA virus in plants. *Plant Cell Rep.* 37, 1707–1712. doi: 10.1007/s00299-018-2297-2

Chen, L., Li, W., Katin-Grazzini, L., Ding, J., Gu, X., Li, Y., et al. (2018). A method for the production and expedient screening of CRISPR-Cas9-mediated non-transgenic mutant plants. *Hort. Res.* 5:13.

Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Method* 13:1050. doi: 10.1038/nmeth.4035

Cody, W. B., and Scholthof, H. B. (2019). Plant virus vectors 3.0: transitioning into synthetic genomics. *Ann. Rev. Phytopathol.* 57, 211–230. doi: 10.1146/annurev-phyto-082718-100301

Cody, W. B., Scholthof, H. B., and Mirkov, T. E. (2017). Multiplexed gene editing and protein overexpression using a tobacco mosaic virus viral vector. *Plant Physiol.* 175, 23–35. doi: 10.1104/pp.17.00411

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

Cox, D. B., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027. doi: 10.1126/science.aaq0180

Dahan−Meir, T., Filler−Hayut, S., Melamed−Bessudo, C., Bocobza, S., Czosnek, H., Aharoni, A., et al. (2018). Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. *Plant J.* 95, 5–16. doi: 10.1111/tpj.13932

Dayaram, A., Potter, K. A., Pailes, R., Marinov, M., Rosenstein, D. D., and Varsani, A. (2015). Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. *Infect. Genet. Evol.* 30, 278–287. doi: 10.1016/j.meegid.2014.12.037

Della Bartola, M., Byrne, S., and Mullins, E. (2020). Characterization of potato virus Y isolates and assessment of nanopore sequencing to detect and genotype potato viruses. *Viruses* 12:478. doi: 10.3390/v12040478

East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R., et al. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* 538, 270–273. doi: 10.1038/nature19802

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

El-Mounadi, K., Morales-Floriano, M. L., and Garcia-Ruiz, H. (2020). Principles, applications, and biosafety of plant genome editing using CRISPR-Cas9. *Front. Plant Sci.* 11:56. doi: 10.3389/fpls.2020.00056

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

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D. L., et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR-Cas-induced gene modifications in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4632–4637. doi: 10.1073/pnas.1400822111

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 532, 517–521. doi: 10.1038/nature17945

Fonseca, P. L., Badotti, F., de Oliveira, T. F., Fonseca, A., Vaz, A. B., Tomé, L. M., et al. (2018). Virome analyses of Hevea brasiliensis using small RNA deep sequencing and PCR techniques reveal the presence of a potential new virus. *Virol. J.* 15:184.

Fortes, P., and Morris, K. V. (2016). Long noncoding RNAs in viral infections. *Virus Res.* 212, 1–11. doi: 10.1016/j.virusres.2015.10.002

Fox, E. J., Reid-Bayliss, K. S., Emond, M. J., and Loeb, L. A. (2014). Accuracy of next generation sequencing platforms. *Next. Gener. Seq. Appl.* 1:1000106.

Freije, C. A., Myhrvold, C., Boehm, C. K., Lin, A. E., Welch, N. L., Carter, A., et al. (2019). Programmable inhibition and detection of RNA viruses using Cas13. *Mol. Cell* 76, 826–837. doi: 10.1016/j.molcel.2019.09.013

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., and Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* 32, 279–284. doi: 10.1038/nbt.2808

Gaj, T., Gersbach, C. A., and Barbas, C. F. III (2013). ZFN, TALEN, and CRISPR-Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi: 10.1016/j.tibtech.2013.04.004

Ghorbani, F. P., Farsi, M., Seifi, A., and Mirshamsi, K. A. (2020). Virus-induced CRISPR-Cas9 system improved resistance against tomato yellow leaf curl virus. *Mol. Biol. Rep.* 47, 3369–3376. doi: 10.1007/s11033-020-05409-3

Gibbs, A. J., Ohshima, K., Phillips, M. J., and Gibbs, M. J. (2008). The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PLoS One* 3:e2523. doi: 10.1371/journal.pone.0002523

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451. doi: 10.1016/j.cell.2013.06.044

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., et al. (2019). Simultaneous CRISPR-Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol. J.* 17, 421–434. doi: 10.1111/pbi.12987

Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., and Zhang, F. (2018). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* 360, 439–444. doi: 10.1126/science.aaq0179

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., et al. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356, 438–442. doi: 10.1126/science.aam9321

Greig, D. R., Jenkins, C., Gharbia, S., and Dallman, T. J. (2019). Comparison of single-nucleotide variants identified by Illumina and Oxford Nanopore technologies in the context of a potential outbreak of Shiga toxin–producing *Escherichia coli*. *GigaScience* 8:giz104.

Guy, P. L. (2013). Ancient RNA? RT-PCR of 50-year-old RNA identifies peach latent mosaic viroid. *Arch. Virol.* 158, 691–694. doi: 10.1007/s00705-012-1527-0

Hadidi, A., Flores, R., Candresse, T., and Barba, M. (2016). Next-generation sequencing and genome editing in plant virology. *Front. Microbiol.* 7:1325.

Hameed, A., Shan-e-Ali Zaidi, S., Sattar, M. N., Iqbal, Z., and Tahir, M. N. (2019). CRISPR technology to combat plant RNA viruses: a theoretical model for Potato virus Y (PVY) resistance. *Micr. Pathogen.* 133:103551. doi: 10.1016/j.micpath.2019.103551

Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D., and Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat. Rev. Microbiol.* 11, 777–788. doi: 10.1038/nrmicro3117

Hao, X., Zhang, W., Zhao, F., Liu, Y., Qian, W., Wang, Y., et al. (2018). Discovery of plant viruses from tea plant (*Camellia sinensis* (L.) o. kuntze) by metagenomic sequencing. *Front. Microbiol.* 9:2175.

Harper, S. J. (2013). Citrus tristeza virus: evolution of complex and varied genotypic groups. *Front. Microbiol.* 4:93.

Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., et al. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* 362, 839–842. doi: 10.1126/science.aav4294

Hashimoto, M., Neriya, Y., Keima, T., Iwabuchi, N., Koinuma, H., Hagiwara-Komoda, Y., et al. (2016). EXA1, a GYF domain protein, is responsible for loss-of-susceptibility to plantago asiatica mosaic virus in *Arabidopsis thaliana*. *Plant J.* 88, 120–131. doi: 10.1111/tpj.13265

He, X.-F., Fang, Y.-Y., Feng, L., and Guo, H.-S. (2008). Characterization of conserved and novel microRNAs and their targets, including a TuMV−induced TIR–NBS–LRR class R gene−derived novel miRNA in Brassica. *FEBS Lett.* 582, 2445–2452. doi: 10.1016/j.febslet.2008.06.011

He, Y., Yang, Z., Hong, N., Wang, G., Ning, G., and Xu, W. (2015). Deep sequencing reveals a novel closterovirus associated with wild rose leaf rosette disease. *Mol. Plant Pathol.* 16, 449–458. doi: 10.1111/mpp.12202

He, Y., Zhu, M., Wang, L., Wu, J., Wang, Q., Wang, R., et al. (2018). Programmed self-elimination of the CRISPR/Cas9 construct greatly accelerates the isolation of edited and transgene-free rice plants. *Mol. Plant* 11, 1210–1213. doi: 10.1016/j.molp.2018.05.005

Heu, C. C., McCullough, F. M., Luan, J., and Rasgon, J. L. (2020). CRISPR-Cas9-based genome editing in the silverleaf whitefly (*Bemisia tabaci*). *CRISPR J.* 3, 89–96. doi: 10.1089/crispr.2019.0067

Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell* 172, 1239–1259. doi: 10.1016/j.cell.2017.11.032

Ho, T., and Tzanetakis, I. E. (2014). Development of a virus detection and discovery pipeline using next generation sequencing. *Virology* 471–473, 54–60. doi: 10.1016/j.virol.2014.09.019

Idris, A., Al-Saleh, M., Piatek, M. J., Al-Shahwan, I., Ali, S., and Brown, J. K. (2014). Viral metagenomics: Analysis of begomoviruses by illumina high-throughput sequencing. *Viruses* 6, 1219–1236. doi: 10.3390/v6031219

Iqbal, Z., Sattar, M. N., and Shafiq, M. (2016). CRISPR/Cas9: a tool to circumscribe cotton leaf curl disease. *Front. Plant Sci.* 7:475.

Ji, X., Si, X., Zhang, Y., Zhang, H., Zhang, F., and Gao, C. (2018). Conferring DNA virus resistance with high specificity in plants using virus-inducible genome-editing system. *Genome Biol.* 19:197.

Ji, X., Zhang, H., Zhang, Y., Wang, Y., and Gao, C. (2015). Establishing a CRISPR–Cas-like immune system conferring DNA virus resistance in plants. *Nat. Plants* 1, 1–4.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife* 2:e00471. doi: 10.7554/eLife.00471

Jo, Y., Choi, H., Kim, S.-M., Kim, S.-L., Lee, B. C., and Cho, W. K. (2017). The pepper virome: natural co-infection of diverse viruses and their quasispecies. *BMC Genom.* 18:453.

Katsiani, A., Stainton, D., Lamour, K., and Tzanetakis, I. E. (2020). The population structure of Rose rosette virus in the USA. *J. Gen. Virol.* 101:jgv001418.

Kaya, H., Ishibashi, K., and Toki, S. (2017). A split Staphylococcus aureus Cas9 as a compact genome-editing tool in plants. *Plant Cell Physiol.* 58, 643–649. doi: 10.1093/pcp/pcx034

Kchouk, M., Gibrat, J.-F., and Elloumi, M. (2017). Generations of sequencing technologies: From first to next generation. *Biol. Med.* 9:3.

Ke, R., Mignardi, M., Hauling, T., and Nilsson, M. (2016). Fourth generation of next−generation sequencing technologies: promise and consequences. *Human Mut.* 37, 1363–1367. doi: 10.1002/humu.23051

Kehoe, M. A., Coutts, B. A., Buirchell, B. J., and Jones, R. A. C. (2014). Plant virology and next generation sequencing: experiences with a *Potyvirus*. *PLoS One* 9:e104580. doi: 10.1371/journal.pone.0104580

Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O., and Zhang, F. (2019). SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat. Prot.* 14, 2986–3012. doi: 10.1038/s41596-019-0210-2

Khan, M. Z., Haider, S., Mansoor, S., and Amin, I. (2019). Targeting plant ssDNA viruses with engineered miniature CRISPR-Cas14a. *Trends Biotechnol.* 37, 800–804. doi: 10.1016/j.tibtech.2019.03.015

Khan, S., Mahmood, M., Rahman, S., Rizvi, F., and Ahmad, A. (2020). Evaluation of the CRISPR-Cas9 system for the development of resistance against *Cotton leaf curl virus* in model plants. *Plant Protect. Sci.* 56, 154–162. doi: 10.17221/105/2019-pps

Kis, A., Hamar, É, Tholt, G., Bán, R., and Havelda, Z. (2019). Creating highly efficient resistance against wheat dwarf virus in barley by employing CRISPR-Cas9 system. *Plant Biotech. J.* 17:1004. doi: 10.1111/pbi.13077

Koonin, E. V., Makarova, K. S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37, 67–78. doi: 10.1016/j.mib.2017.05.008

Kreutzberg, V. (1940). A new virus disease of *Pistacia vera* L. *Dokl. Akad. Nauk SSSR* 27, 614–617.

Kundu, A., Paul, S., Dey, A., and Pal, A. (2017). High throughput sequencing reveals modulation of microRNAs in *Vigna mungo* upon Mungbean Yellow Mosaic India Virus inoculation highlighting stress regulation. *Plant Sci.* 257, 96–105. doi: 10.1016/j.plantsci.2017.01.016

Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., et al. (2017). CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* 14, 710–712. doi: 10.1038/nmeth.4327

Kutnjak, D., Elena, S. F., and Ravnikar, M. (2017). Time-sampled population sequencing reveals the interplay of selection and genetic drift in experimental evolution of *Potato Virus Y*. *J. Virol.* 91:e00690-17.

Laney, A. G., Keller, K. E., Martin, R. R., and Tzanetakis, I. E. (2011). A discovery 7x0 years in the making: characterization of the *Rose rosette* virus. *J. Gen. Virol.* 92, 1727–1732. doi: 10.1099/vir.0.031146-0

Lee, H., Chang, H. Y., Cho, S. W., and Ji, H. P. (2020). CRISPRpic: fast and precise analysis for CRISPR-induced mutations via p refixed i ndex c ounting. *NAR Gen.Bioinfo.* 2:lqaa012.

Lee, K., Zhang, Y., Kleinstiver, B. P., Guo, J. A., Aryee, M. J., Miller, J., et al. (2019). Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnol. J.* 17, 362–372. doi: 10.1111/pbi.12982

Legg, J., and Thresh, J. (2000). Cassava mosaic virus disease in East Africa: a dynamic disease in a changing environment. *Virus Res.* 71, 135–149. doi: 10.1016/s0168-1702(00)00194-5

Li, F., Yang, X., Bisaro, D. M., and Zhou, X. (2018). The βC1 protein of geminivirus–betasatellite complexes: A target and repressor of host defenses. *Mol. Plant* 11, 1424–1426. doi: 10.1016/j.molp.2018.10.007

Li, J., Chen, Z., Chen, F., Ling, Y., Peng, Y., Luo, N., et al. (2019a). Targeted mRNA demethylation using an engineered dCas13b-ALKBH5 fusion protein. *BioRxiv* [Preprint]. 614859.

Li, J., Manghwar, H., Sun, L., Wang, P., Wang, G., Sheng, H., et al. (2019b). Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR-Cas9-edited cotton plants. *Plant Biotechnol. J.* 17, 858–868. doi: 10.1111/pbi.13020

Li, L., Li, S., Wu, N., Wu, J., Wang, G., Zhao, G., et al. (2019). HOLMESv2: a CRISPR-Cas12b-assisted platform for nucleic acid detection and DNA methylation quantitation. *ACS Synthet. Biol.* 8, 2228–2237. doi: 10.1021/acssynbio.9b00209

Li, R., Gao, S., Berendsen, S., Fei, Z., and Ling, K.-S. (2015). Complete genome sequence of a novel genotype of squash mosaic virus infecting squash in Spain. *Genome Announc.* 3:e001583-14.

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

Liang, P., Navarro, B., Zhang, Z., Wang, H., Lu, M., Xiao, H., et al. (2015). Identification and characterization of a novel geminivirus with a monopartite genome infecting apple trees. *J. Gen. Virol.* 96, 2411–2420. doi: 10.1099/vir.0.000173

Liu, H., Soyars, C. L., Li, J., Fei, Q., He, G., Peterson, B. A., et al. (2018). CRISPR-Cas9-mediated resistance to cauliflower mosaic virus. *Plant Dis.* 2:e00047. doi: 10.1002/pld3.47

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of next-generation sequencing systems. *BioMed Res. Intl.* 2012:251364.

Loconsole, G., Önelge, N., Potere, O., Giampetruzzi, A., Bozan, O., Satar, S., et al. (2012a). Identification and characterization of citrus yellow vein clearing virus, a putative new member of the genus mandarivirus. *Phytopatholoy* 102, 1168–1175. doi: 10.1094/phyto-06-12-0140-r

Loconsole, G., Saldarelli, P., Doddapaneni, H., Savino, V., Martelli, G. P., and Saponari, M. (2012b). Identification of a single-stranded DNA virus associated with citrus chlorotic dwarf disease, a new member in the family *Geminiviridae*. *Virology* 432, 162–172. doi: 10.1016/j.virol.2012.06.005

Lõhmus, A., Hafrén, A., and Mäkinen, K. (2017). Coat protein regulation by CK2, CPIP, HSP70, and CHIP Is required for potato virus A replication and coat protein accumulation. *J. Virol.* 91, e01316–e01316. doi: 10.1128/jvi.01316-16

Loriato, V. A. P., Martins, L. G. C., Euclydes, N. C., Reis, P. A. B., Duarte, C. E. M., and Fontes, E. P. B. (2020). Engineering resistance against geminiviruses: a review of suppressed natural defenses and the use of RNAi and the CRISPR-Cas system. *Plant Sci.* 292:110410. doi: 10.1016/j.plantsci.2020.110410

Lozano-Durán, R., Rosas-Díaz, T., Luna, A. P., and Bejarano, E. R. (2011). Identification of host genes involved in geminivirus infection using a reverse genetics approach. *PLoS One* 6:e22383. doi: 10.1371/journal.pone.0022383

Ma, Y., Navarro, B., Zhang, Z., Lu, M., Zhou, X., Chi, S., et al. (2015). Identification and molecular characterization of a novel monopartite geminivirus associated with mulberry mosaic dwarf disease. *J. Gen. Virol.* 96, 2421–2434. doi: 10.1099/vir.0.000175

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Čermák, T., et al. (2018). Novel alleles of rice eIF4G generated by CRISPR-Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol. J.* 16, 1918–1927. doi: 10.1111/pbi.12927

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., and Joung, J. K. (2013). CRISPR RNA–guided activation of endogenous human genes. *Nat. Meth.* 10, 977–979. doi: 10.1038/nmeth.2598

Mahas, A., Ali, Z., Tashkandi, M., and Mahfouz, M. M. (2019a). "Virus-mediated genome editing in plants using the CRISPR/Cas9 system," in *Plant Genome Editing with CRISPR Systems. Methods in Molecular Biology*, Vol. 1917, ed. Y. Qi (New York, NY: Humana Press).

Mahas, A., Aman, R., and Mahfouz, M. (2019b). CRISPR-Cas13d mediates robust RNA virus interference in plants. *Gen. Biol.* 20, 1–16.

Makarova, K. S., and Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol.* 1311, 47–75. doi: 10.1007/978-1-4939-2687-9_4

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi: 10.1038/s41579-019-0299-x

Manghwar, H., Li, B., Ding, X., Hussain, A., Lindsey, K., Zhang, X., et al. (2020). CRISPR-Cas systems in genome editing: methodologies and tools for sgRNA Design, off-target evaluation, and strategies to mitigate off-target effects. *Adv. Sci.* 7:1902312. doi: 10.1002/advs.201902312

Manghwar, H., Lindsey, K., Zhang, X., and Jin, S. (2019). CRISPR/Cas system: recent advances and future prospects for genome editing. *Trends Plant Sci.* 24, 1102–1125. doi: 10.1016/j.tplants.2019.09.006

Mansoor, S., Khan, S. H., Bashir, A., Saeed, M., Zafar, Y., Malik, K. A., et al. (1999). Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology* 259, 190–199. doi: 10.1006/viro.1999.9766

Marais, A., Faure, C., Couture, C., Bergey, B., Gentit, P., and Candresse, T. (2014). Characterization by deep sequencing of divergent Plum bark necrosis stem pitting-associated virus (PBNSPaV) isolates and development of a broad-spectrum PBNSPaV detection assay. *Phytopathology* 104, 660–666. doi: 10.1094/phyto-08-13-0229-r

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature* 470, 198. doi: 10.1038/nature09796

McGavin, W. J., Mitchell, C., Cock, P. J., Wright, K. M., and MacFarlane, S. A. (2012). Raspberry leaf blotch virus, a putative new member of the genus Emaravirus, encodes a novel genomic RNA. *J. Gen. Virol.* 93, 430–437. doi: 10.1099/vir.0.037937-0

Mehta, D., Stürchler, A., Anjanappa, R. B., Zaidi, S. S., Hirsch-Hoffmann, M., Gruissem, W., et al. (2019). Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses. *Genome Biol.* 20:80.

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Gen. Res.* 15, 1767–1776. doi: 10.1101/gr.3770505

Mignardi, M., and Nilsson, M. (2014). Fourth-generation sequencing in the cell and the clinic. *Gen. Med.* 6:31. doi: 10.1186/gm548

Moreno, P., Ambros, 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

Myhrvold, C., Freije, C. A., Gootenberg, J. S., Abudayyeh, O. O., Metsky, H. C., Durbin, A. F., et al. (2018). Field-deployable viral diagnostics using CRISPR-Cas13. *Science* 360, 444–448. doi: 10.1126/science.aas8836

Naidu, R. A., Kimmins, F. M., Deom, C. M., Subrahmanyam, P., Chiyembekeza, A. J., and Van der Merwe, P. J. A. (1999). Groundnut rossette: a virus disease affecting groundnut production in sub-saharan Africa. *Plant Dis.* 83, 700–709. doi: 10.1094/pdis.1999.83.8.700

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., and Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* 7:482.

Ng, T. F. F., Duffy, S., Polston, J. E., Bixby, E., Vallad, G. E., and Breitbart, M. (2011). Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLoS One* 6:e19050. doi: 10.1371/journal.pone.0019050

Nicaise, V., Gallois, J.-L., Chafiai, F., Allen, L. M., Schurdi-Levraud, V., Browning, K. S., et al. (2007). Coordinated and selective recruitment of eIF4E and eIF4G factors for potyvirus infection in *Arabidopsis thaliana*. *FEBS Lett.* 581, 1041–1046. doi: 10.1016/j.febslet.2007.02.007

Paegel, B. M., Emrich, C. A., Wedemayer, G. J., Scherer, J. R., and Mathies, R. A. (2002). High throughput DNA sequencing with a microfabricated 96-lane capillary array electrophoresis bioprocessor. *Proc. Natl. Acad. Sci. U.S.A.* 99, 574–579. doi: 10.1073/pnas.012608699

Palukaitis, P. (2016). Satellite RNAs and satellite viruses. *Mol. Plant-Microbe Inter.* 29, 181–186. doi: 10.1094/mpmi-10-15-0232-fi

Pasin, F., Menzel, W., and Daròs, J. A. (2019). Harnessed viruses in the age of metagenomics and synthetic biology: an update on infectious clone assembly and biotechnologies of plant viruses. *Plant Biotechnol. J.* 17, 1010–1026. doi: 10.1111/pbi.13084

Patrick, R. M., Mayberry, L. K., Choy, G., Woodard, L. E., Liu, J. S., White, A., et al. (2014). Two *Arabidopsis* loci encode novel eukaryotic initiation factor 4E isoforms that are functionally distinct from the conserved plant eukaryotic initiation factor 4E. *Plant Physiol.* 164, 1820–1830. doi: 10.1104/pp.113.227785

Peyambari, M., Warner, S., Stoler, N., Rainer, D., and Roossinck, M. J. (2019). A 1,000-Year-Old RNA Virus. *J. Virol.* 93:e001188-18.

Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L., et al. (2018). Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* 8:10950.

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

Poojari, S., Alabi, O. J., Fofanov, V. Y., and Naidu, R. A. (2013). A leafhopper-transmissible DNA virus with novel evolutionary lineage in the family geminiviridae implicated in grapevine redleaf disease by next-generation sequencing. *PLoS One* 8:e64194. doi: 10.1371/journal.pone.0064194

Pyott, D. E., Sheehan, E., and Molnar, A. (2016). Engineering of CRISPR-Cas9-mediated potyvirus resistance in transgene-free *Arabidopsis* plants. *Mol. Plant Pathol.* 17, 1276–1288. doi: 10.1111/mpp.12417

Quan, J., Langelier, C., Kuchta, A., Batson, J., Teyssier, N., Lyden, A., et al. (2019). FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. *Nuclic Acids Res.* 47, e83–e83.

Rahman, M. U., Khan, A. Q., Rahmat, Z., Iqbal, M. A., and Zafar, Y. (2017). Genetics and genomics of cotton leaf curl disease, its viral causal agents and whitefly vector: a way forward to sustain cotton fiber security. *Front. Plant Sci.* 8:1157.

Rang, F. J., Kloosterman, W. P., and de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Gen. Biol.* 19:90.

Roossinck, M. J. (2017). Deep sequencing for discovery and evolutionary analysis of plant viruses. *Vir. Res.* 239, 82–86. doi: 10.1016/j.virusres.2016.11.019

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

Rosario, K., Capobianco, H., Ng, T. F. F., Breitbart, M., and Polston, J. E. (2014). RNA viral metagenome of whiteflies leads to the discovery and characterization of a whitefly-transmitted carlavirus in North America. *PLoS One* 9:e86748. doi: 10.1371/journal.pone.0086748

Rose, J. C., Popp, N. A., Richardson, C. D., Stephany, J. J., Mathieu, J., Wei, C. T., et al. (2020). Suppression of unwanted CRISPR-Cas9 editing by co-administration of catalytically inactivating truncated guide RNAs. *Nat. Commun.* 11:2697. doi: 10.1038/s41467-020-16542-9

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475:348.

Roy, A., Zhai, Y., Ortiz, J., Neff, M., Mandal, B., Mukherjee, S. K., et al. (2019). Multiplexed editing of a begomovirus genome restricts escape mutant formation and disease development. *PLoS One* 14:e0223765. doi: 10.1371/journal.pone.0223765

Ryan, D. E., Taussig, D., Steinfeld, I., Phadnis, S. M., Lunstad, B. D., Singh, M., et al. (2017). Improving CRISPR–Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res.* 46, 792–803. doi: 10.1093/nar/gkx1199

Sanfaçon, H. (2015). Plant translation factors and virus resistance. *Viruses* 7, 3392–3419. doi: 10.3390/v7072778

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Sastry, K. S., and Zitter, T. A., eds. (2014). "Management of virus and viroid diseases of crops in the tropics," in *Plant Virus and Viroid Diseases in the Tropics* (Dordrecht: Springer), 149–480. doi: 10.1007/978-94-007-7820-7_2

Savage, D. F. (2019). *Cas14: Big Advances from Small CRISPR Proteins*. Washington, DC: ACS Publications.

Scheben, A., Wolter, F., Batley, J., Puchta, H., and Edwards, D. (2017). Towards CRISPR/Cas crops–bringing together genomics and genome editing. *New Phytol.* 216, 682–698. doi: 10.1111/nph.14702

Schmidt, C., Pacher, M., and Puchta, H. (2019). "DNA break repair in plants and its application for genome engineering," in *Transgenic Plants*, S. Kumar, P. Barone and M. Smith (New York, NY: Humana Press), 237–266. doi: 10.1007/978-1-4939-8778-8_17

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., et al. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* 6, 550–551. doi: 10.1038/nmeth0809-550

Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P., and Pruett-Miller, S. M. (2018). A survey of validation strategies for CRISPR-Cas9 editing. *Sci. Rep.* 8:888.

Shahid, M., Shafiq, M., Ilyas, M., Raza, A., Al-Sadrani, M., Al-Sadi, A., et al. (2019). Frequent occurrence of Mungbean yellow mosaic India virus in tomato leaf curl disease affected tomato in Oman. *Sci. Rep.* 9, 1–14.

Shahid, M. S., Aboughanem-Sabanadzovic, N., Sabanadzovic, S., and Tzanetakis, I. E. (2017). Genomic characterization and population structure of a badnavirus infecting blackberry. *Plant Dis.* 101, 110–115. doi: 10.1094/pdis-04-16-0527-re

Shin, H. Y., Wang, C., Lee, H. K., Yoo, K. H., Zeng, X., Kuhns, T., et al. (2017). CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat. Commun.* 8:15464.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* 15, 169–182. doi: 10.1038/nrmicro.2016.184

Silva, T. F., Romanel, E. A., Andrade, R. R., Farinelli, L., Østerås, M., Deluen, C., et al. (2011). Profile of small interfering RNAs from cotton plants infected with the polerovirus Cotton leafroll dwarf virus. *BMC Mol. Biol.* 12:40. doi: 10.1186/1471-2199-12-40

Smith, O., Clapham, A., Rose, P., Liu, Y., Wang, J., and Allaby, R. G. (2014). A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Sci. Rep.* 4, 4003–4003.

Sugano, S. S., Nishihama, R., Shirakawa, M., Takagi, J., Matsuda, Y., Ishida, S., et al. (2018). Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in Marchantia polymorpha. *PLoS One* 13:e0205117. doi: 10.1371/journal.pone.0205117

Tang, X., Liu, G., Zhou, J., Ren, Q., You, Q., Tian, L., et al. (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol.* 19, 84.

Tashkandi, M., Ali, Z., Aljedaani, F., Shami, A., and Mahfouz, M. M. (2018). Engineering resistance against Tomato yellow leaf curl virus via the CRISPR-Cas9 system in tomato. *Plant Signal. Behav.* 13:e1525996. doi: 10.1080/15592324.2018.1525996

Thresh, J., and Cooter, R. (2005). Strategies for controlling cassava mosaic virus disease in Africa. *Plant pathol.* 54, 587–614. doi: 10.1111/j.1365-3059.2005.01282.x

Tripathi, A., Goswami, K., Tiwari, M., Mukherjee, S. K., and Sanan-Mishra, N. (2018). Identification and comparative analysis of microRNAs from tomato varieties showing contrasting response to ToLCV infections. *Physiol. Mol. Biol. Plants* 24, 185–202. doi: 10.1007/s12298-017-0482-3

Várallyay, É, Válóczi, A., Ágyi, Á, Burgyán, J., and Havelda, Z. (2010). Plant virus−mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *EMBO J.* 29, 3507–3519. doi: 10.1038/emboj.2010.215

Vats, S., Kumawat, S., Kumar, V., Patil, G. B., Joshi, T., Sonah, H., et al. (2019). Genome editing in plants: exploration of technological advancements and challenges. *Cells* 8:1386. doi: 10.3390/cells8111386

Villamor, D. E. V., Ho, T., Martin, R. R., and Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109, 716–725. doi: 10.1094/PHYTO-07-18-0257-RVW

Villamor, D. E. V., Mekuria, T. A., Pillai, S. S., and Eastwell, K. C. (2016). High-throughput sequencing identifies novel viruses in nectarine: insights to the etiology of stem-pitting disease. *Phytopathology* 106, 519–527. doi: 10.1094/phyto-07-15-0168-r

Vives, M. C., Velázquez, K., Pina, J. A., Moreno, P., Guerri, J., and Navarro, L. (2013). Identification of a new enamovirus associated with citrus vein enation disease by deep sequencing of small RNAs. *Phytopathology* 103, 1077–1086. doi: 10.1094/phyto-03-13-0068-r

Vu, G. T., Cao, H. X., Fauser, F., Reiss, B., Puchta, H., and Schubert, I. (2017). Endogenous sequence patterns predispose the repair modes of CRISPR-Cas9−induced DNA double−stranded breaks in *Arabidopsis thaliana*. *Plant J.* 92, 57–67. doi: 10.1111/tpj.13634

Wale, S., Platt, B., and Cattlin, N. D. (2008). *Diseases, Pests and Disorders of Potatoes: A Colour Handbook*. Amsterdam: Elsevier.

Wang, A. (2015). Dissecting the molecular network of virus-plant interactions: The complex roles of host factors. *Ann. Rev. Phytopathol.* 53, 45–66. doi: 10.1146/annurev-phyto-080614-120001

Wang, H., Nakamura, M., Abbott, T. R., Zhao, D., Luo, K., Yu, C., et al. (2019). CRISPR-mediated live imaging of genome editing and transcription. *Science* 365, 1301–1305. doi: 10.1126/science.aax7852

Wang, P., Zhang, J., Sun, L., Ma, Y., Xu, J., Liang, S., et al. (2018). High efficient multisites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system. *Plant Biotechnol. J.* 16, 137–150. doi: 10.1111/pbi.12755

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi: 10.1038/nbt.2969

Watts, D., and MacBeath, J. R. (2001). *Automated fluorescent DNA sequencing on the ABI PRISM 310 Genetic Analyzer: DNA Sequencing Protocols*. Berlin: Springer, 153–170.

Welliver, R. A. (2012). Plum pox virus case study: the eradication road is paved in gold. *Phytopathology* 102, 154–154.

Wessels, H.-H., Méndez-Mancilla, A., Guo, X., Legut, M., Daniloski, Z., and Sanjana, N. E. (2020). Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat. Biotechnol.* 38, 722–727. doi: 10.1038/s41587-020-0456-9

Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10092–10097. doi: 10.1073/pnas.1102716108

Woo, J. W., Kim, J., Kwon, S. I., Corvalan, C., Cho, S. W., Kim, H., et al. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164. doi: 10.1038/nbt.3389

Wu, Q., Ding, S.-W., Zhang, Y., and Zhu, S. (2015). Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. *Ann. Rev. Phytopathol.* 53, 425–444. doi: 10.1146/annurev-phyto-080614-120030

Xu, W., Fu, W., Zhu, P., Li, Z., Wang, C., Wang, C., et al. (2019). Comprehensive analysis of CRISPR-Cas9-mediated mutagenesis in *Arabidopsis thaliana* by genome-wide sequencing. *Int. J. Mol. Sci.* 20, 4125. doi: 10.3390/ijms20174125

Yamanaka, T., Ohta, T., Takahashi, M., Meshi, T., Schmidt, R., Dean, C., et al. (2000). TOM1, an Arabidopsis gene required for efficient multiplication of a tobamovirus, encodes a putative transmembrane protein. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10107–10112. doi: 10.1073/pnas.170295097

Yan, W. X., Chong, S., Zhang, H., Makarova, K. S., Koonin, E. V., Cheng, D. R., et al. (2018). Cas13d Is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Mol. Cell* 70, 327–339. doi: 10.1016/j.molcel.2018.02.028

Yin, K., Han, T., Liu, G., Chen, T., Wang, Y., Yu, A. Y. L., et al. (2015). A geminivirus-based guide RNA delivery system for CRISPR-Cas9 mediated plant genome editing. *Sci. Rep.* 5:14926.

Yin, K., Han, T., Xie, K., Zhao, J., Song, J., and Liu, Y. (2019). Engineer complete resistance to Cotton Leaf Curl Multan virus by the CRISPR-Cas9 system in *Nicotiana benthamiana*. *Phytopathol. Res.* 1:9.

Zaidi, S. S. A., and Mansoor, S. (2017). Viral vectors for plant genome engineering. *Front. Plant Sci.* 8:539.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi: 10.1016/j.cell.2015.09.038

Zhan, X., Zhang, F., Zhong, Z., Chen, R., Wang, Y., Chang, L., et al. (2019). Generation of virus-resistant potato plants by RNA genome targeting. *Plant Biotech. J.* 17, 1814–1822. doi: 10.1111/pbi.13102

Zhang, F., LeBlanc, C., Irish, V. F., and Jacob, Y. (2017). Rapid and efficient CRISPR/Cas9 gene editing in Citrus using the YAO promoter. *Plant Cell Rep.* 36, 1883–1887. doi: 10.1007/s00299-017-2202-4

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., et al. (2014). The CRISPR-Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotech. J.* 12, 797–807. doi: 10.1111/pbi.12200

Zhang, P., Du, H., Wang, J., Pu, Y., Yang, C., Yan, R., et al. (2020). Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus. *Plant Biotech. J.* 18, 1384–1395. doi: 10.1111/pbi.13302

Zhang, T., Zhao, Y., Ye, J., Cao, X., Xu, C., Chen, B., et al. (2019). Establishing CRISPR-Cas13a immune system conferring RNA virus resistance in both dicot and monocot plants. *Plant Biotech. J.* 17, 1185–1187. doi: 10.1111/pbi.13095

Zhang, T., Zheng, Q., Yi, X., An, H., Zhao, Y., Ma, S., et al. (2018). Establishing RNA virus resistance in plants by harnessing CRISPR immune system. *Plant Biotech. J.* 16, 1415–1423. doi: 10.1111/pbi.12881

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR-Cas9 DNA or RNA. *Nat. Commun.* 7, 1–8.

Zhang, Y., Zhang, F., Li, X., Baller, J. A., Qi, Y., Starker, C. G., et al. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol.* 161, 20–27. doi: 10.1104/pp.112.205179

Zou, L.-J., Deng, X.-G., Han, X.-Y., Tan, W.-R., Zhu, L.-J., Xi, D.-H., et al. (2016). Role of transcription factor HAT1 in modulating *Arabidopsis thaliana* response to cucumber mosaic virus. *Plant Cell Physiol.* 57, 1879–1889. doi: 10.1093/pcp/pcw109

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.

Copyright © 2021 Shahid, Sattar, Iqbal, Raza and Al-Sadi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

REVIEW
published: 27 January 2021
doi: 10.3389/fmicb.2020.609784

Application of CRISPR/Cas for Diagnosis and Management of Viral Diseases of Banana

Leena Tripathi[1], Valentine Otang Ntui[1], Jaindra Nath Tripathi[1] and P. Lava Kumar[2]*

[1] International Institute of Tropical Agriculture, Nairobi, Kenya, [2] International Institute of Tropical Agriculture, Ibadan, Nigeria

Viral diseases are significant biotic constraints for banana (*Musa* spp.) production as they affect the yield and limit the international movement of germplasm. Among all the viruses known to infect banana, the banana bunchy top virus and banana streak viruses are widespread and economically damaging. The use of virus-resistant bananas is the most cost-effective option to minimize the negative impacts of viral-diseases on banana production. CRISPR/Cas-based genome editing is emerging as the most powerful tool for developing virus-resistant crop varieties in several crops, including the banana. The availability of a vigorous genetic transformation and regeneration system and a well-annotated whole-genome sequence of banana makes it a compelling candidate for genome editing. A robust CRISPR/Cas9-based genome editing of the banana has recently been established, which can be applied in developing disease-resistant varieties. Recently, the CRISPR system was exploited to detect target gene sequences using Cas9, Cas12, Cas13, and Cas14 enzymes, thereby unveiling the use of this technology for virus diagnosis. This article presents a synopsis of recent advancements and perspectives on the application of CRISPR/Cas-based genome editing for diagnosing and developing resistance against banana viruses and challenges in genome-editing of banana.

Keywords: banana, viral diseases, BBTV, BSV, genome editing, CRISPR/Cas, diagnosis, disease-resistance

OPEN ACCESS

Edited by:
Ahmed Hadidi, Agricultural Research Service, United States Department of Agriculture, United States

Reviewed by:
Anupam Varma, Indian Agricultural Research Institute (ICAR), India
Herve Vanderschuren, KU Leuven, Belgium
Magdy Mahfouz, King Abdullah University of Science and Technology, Saudi Arabia

****Correspondence:***
Leena Tripathi
L.Tripathi@cgiar.org

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

Received: *24 September 2020*
Accepted: *29 December 2020*
Published: *27 January 2021*

Citation:
Tripathi L, Ntui VO, Tripathi JN and Kumar PL (2021) Application of CRISPR/Cas for Diagnosis and Management of Viral Diseases of Banana. Front. Microbiol. 11:609784. doi: 10.3389/fmicb.2020.609784

INTRODUCTION

Plant viruses are obligate intracellular pathogens, which utilize the host plant's molecular machinery to replicate. They cause many economically important plant diseases and are responsible for losses in crop yield and quality worldwide. Several viruses affect banana production worldwide because of their effects on yield, quality, and limitations to the international germplasm exchange (Tripathi et al., 2016). Among these, banana bunchy top virus (BBTV, genus *Babuvirus*) and banana streak virus (BSV, genus *Badnavirus*) are economically important viruses threatening banana production (Kumar et al., 2015). These viruses reduce crop yield and productivity, posing a severe threat to food and nutrition security in banana-growing regions.

Banana and plantain (*Musa* spp., hereafter referred to as banana), is one of the chief staple food crops in 136 countries in tropics and sub-tropics, with an annual production of 155 million tons on 11 million hectares of land, and feeding millions of people (FAOSTAT, 2018). One-third of its global production is from Africa, with East Africa being the largest banana-growing and consuming region. Numerous types of banana, such as dessert, cooking, roasting, and brewing

types are grown in different areas of the world and provide food for millions of people. Bananas are cultivated predominantly by smallholder farmers for home consumption and local or regional markets; only approximately 16% of production enters international markets (FAOSTAT, 2018). It is valuable food security and cash crop with huge potential to provide raw material to the budding agro-industry. It is cultivated in diverse environments and produces fruits throughout the year in favorable weather conditions.

The *Musa* spp. has four genomes corresponding to the genetic constitutions belonging to the four wild *Eumusa* species, *Musa acuminata* (AA genome), *Musa balbisiana* (BB genome), *Musa schizocarpa* (SS genome), and *Australimusa* species (TT genome; Davey et al., 2013). All cultivated bananas are generally seedless, parthenocarpic, and vegetatively propagated triploid (AAA, AAB, or ABB genome) hybrids between subspecies of *M. acuminata* (AA genome), or between *M. acuminata* and *M. balbisiana* (BB genome; McKey et al., 2010). Some cultivated bananas can have diploid or tetraploid genomes, including hybrids within or between the two *Musa* species. Hundreds of cultivars of bananas are grown and consumed worldwide. Still, large-scale farmers mainly grow the Cavendish subgroup (AAA genome) of dessert bananas for commercialization in local and international markets (Tripathi et al., 2020). Other dessert banana varieties such as Gros Michel (AAA genome), Sukali Ndiizi (commonly known as apple banana, AAB genome), Mysore (AAB genome), Silk (AAB genome), and Pome (AAB genome) are also grown at a small scale. Besides, cooking types such as the East African Highland Banana (EAHB, AAA genome) and bluggoe (ABB genome), the roasting type plantain (AAB genome), and the brewing type such as Pisang Awak (ABB genome) are also grown mainly in Africa. Plantain is mostly grown in Central and West Africa, and EAHB is cultivated in East Africa.

Banana is vegetatively propagated using suckers or *in vitro* plantlets and grown almost as perennial plantations (Kumar et al., 2015). As a vegetatively propagated crop, their production is affected due to the build−up of certain pests and pathogens, especially viruses, between successive plantings via infected planting material. Viruses of banana are challenging to control because of vegetative propagation, and many viruses are transmitted by insect vectors, contributing to further spread within the fields. Antiviral compounds are not available to cure banana plants infected with viruses. The control measures can contain the spread of viruses and prevent reinfection.

The diagnosis of a virus is the first step in the management of a viral disease. An efficient diagnostic and quarantine system is required to prevent the spread of viruses (Kumar et al., 2015). Viral infection of banana can be managed through phytosanitation, such as using virus-free planting material and strict regulation on the movement of infected planting materials. An alternative, cost-effective strategy for controlling banana viruses is to develop host plant resistance. Although conventional breeding has been used to create viral resistance in several crops, no success has been achieved in banana due to the unavailability of any known virus-resistant germplasm (Kumar et al., 2015). Developing virus-resistant varieties of banana using conventional breeding is challenging due to the low genetic variability in *Musa* germplasm, polyploidy, lengthy production cycle, and sterility of the majority of cultivars (Dale et al., 2017). Therefore, there is a critical need to delve into new breeding technologies such as transgenic and genome-editing to develop resistance against banana viruses. A few advances have been reported demonstrating enhanced resistance against BBTV using RNAi-mediated transgenic approaches (Shekhawat et al., 2012; Elayabalan et al., 2017). However, the commercialization of transgenic crops faces hurdles due to complicated regulatory approval processes. Genome-editing can fast-track breeding by making efficient and precise changes in the plant genome to develop new traits such as viral disease-resistance. A CRISPR/Cas9-based genome editing of banana has recently been established targeting the knockout of the *phytoene desaturase* (*PDS*) as a marker gene (Ntui et al., 2020). The highly efficient genome-editing tool developed using the different groups of banana has paved the path to develop disease-resistant varieties by knocking out single or multiple genes (Kaur et al., 2018; Naim et al., 2018; Ntui et al., 2020). Here, we present an overview of recent progress and perspectives to explore the application of CRISPR/Cas methods to diagnose and manage banana viruses.

MAJOR VIRAL DISEASES OF BANANA

There are about 20 viruses infecting banana globally, out of which four viruses, (BBTV, genus *Babuvirus*, family *Nanoviridae*), (BSV, genus *Badnavirus*, family *Caulimoviridae*), banana bract mosaic virus (BBrMV, genus *Potyvirus*, family *Potyviridae*) and cucumber mosaic virus (CMV, genus *Cucumovirus*, family *Bromoviridae*) are the most significant (Tripathi et al., 2016; **Figure 1**). Among them, BSV, BBrMV, and CMV are known to occur in all banana producing countries, whereas BBTV spread is limited to a few countries. Of all the viruses, BBTV and BSV are major threats to banana production.

Banana bunchy top disease (BBTD), caused by BBTV, is the most important viral disease of banana responsible for the significant adverse economic impact on banana production. BBTV is transmitted by the banana aphid (*Pentalonia nigronervosa*) and infected planting materials. BBTD was first noticed in Fiji in the 1880s and is currently present in more than 36 countries in Africa, Asia, Oceania, and South Pacific (Kumar et al., 2015; **Figure 2**). In Africa, BBTD is present in 17 countries, and adjacent banana-producing countries are at a high risk of being affected (Kumar et al., 2011; Adegbola et al., 2013; Jooste et al., 2016). In the last decade, BBTV has spread to at least six countries in Africa, including Benin, Cameroon, Mozambique, Nigeria, South Africa, and Zambia. In 2018, an incidence of BBTV was reported in Togo, but early detection and eradication prevented disease establishment (IITA News, 2019). This indicates of a continuous spread of the disease in banana-producing regions, causing decreased crop production. Fruit production in infected plants reduces by 70 to 100% within one season, and plantations cannot be recovered from infections. Most cultivars of *Musa* spp. are susceptible to BBTV and few

FIGURE 1 | Symptoms of viral diseases in infected banana plants. **(A)** Banana bunchy top virus (BBTV) infected plant showing stunting and bunchy leaves, **(B)** Banana streak virus (BSV) infected plant showing yellow streaks.

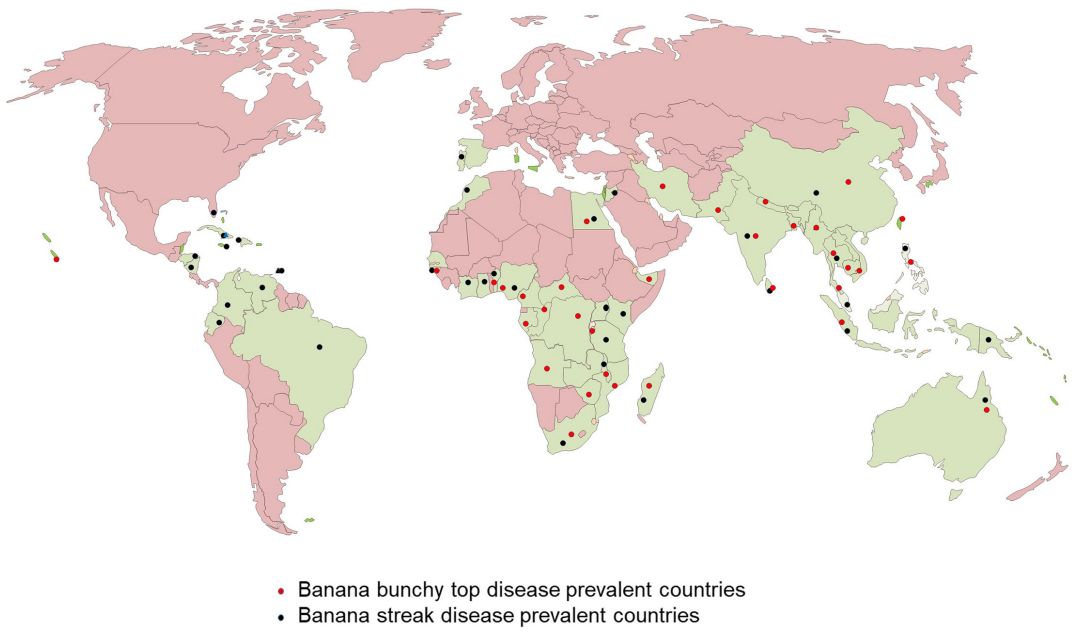


FIGURE 2 | World map showing the distribution of Banana bunchy top virus (BBTV) and Banana streak virus (BSV).

tolerant cultivars identified have limited potential for adoption in diverse production systems (Ngatat et al., 2017). The infected plants are stunted with bunchy and narrow leaves with brittle, and yellow edges (**Figure 1A**). Severely infected plants do not produce fruits; even if fruits are produced, they are deformed.

Banana bunchy top virus is a single-stranded DNA (ssDNA) virus with a multipartite genome comprising of six circular components (**Figure 3**) with an approximate size of 1.1 kb each (Harding et al., 1993; Burns et al., 1995). The six components, named DNA-R, -U3, -S, -M, -C, and -N (previously known as DNA 1-6), are encapsulated within separate virions, each about 18–20 nm in diameter (Harding et al., 1993). All the six components have a common genome organization comprising of a major common region (CR-M), stem-loop common region

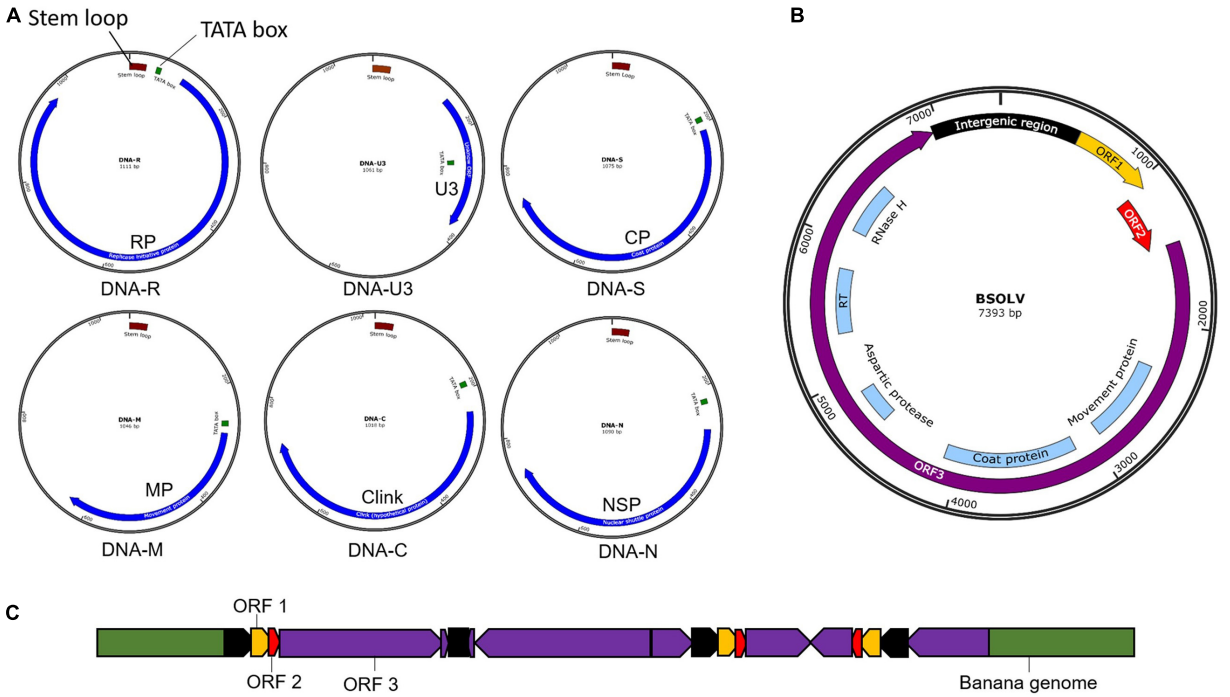


FIGURE 3 | Genomic maps of banana viruses. **(A)** Banana bunchy top virus (BBTV), **(B)** Banana streak virus (BSV), and **(C)** Endogenous Banana streak virus (eBSV) integrated into the plant genome.

(CR-SL), potential TATA box 3′ of the stem-loop, at least one open reading frame (ORF) for a major gene in the virion sense and polyadenylation signals associated with each gene (Burns et al., 1995; **Figure 3A**). DNA-R encodes a replication initiation protein (Rep) responsible for initiating viral DNA replication, DNA-S encodes the coat protein (CP), DNA-M encodes the movement protein (MP), DNA-C encodes the cell cycle link protein (Clink), DNA-N encodes the nuclear shuttle protein (NSP), while the function of DNA-U3 is unknown (Burns et al., 1995; Wanitchakorn et al., 1997, 2000). Two broad groups of BBTV isolates have been identified based on nucleotide sequence differences between their genome components, referred to as the "South Pacific" group having isolates from Africa, Australia, South Asia, South-Pacific, while the "Asian" group comprises isolates from East Asia (China, Indonesia, Japan, Philippines, Taiwan, Thailand, and Vietnam; Karan et al., 1994; Kumar et al., 2011).

Banana streak virus is a pararetrovirus infecting banana, infection of which will result in chlorotic and necrotic streaks on leaves and pseudostem (**Figure 1B**). The diseased plants may be stunted with distorted fruits and smaller bunches. BSV was first reported in 1958 in West Africa and subsequently detected in all the banana-growing countries (Fargette et al., 2006; **Figure 2**). BSV was differentiated based on the genome sequence diversity into several species, of which the four most significant species are: *Banana streak Obino l'Ewai virus*, *Banana streak Mysore virus*, *Banana streak Imové virus*, and *Banana streak Goldfinger virus*, all of which are known to present commonly in banana plantations (Geering et al., 2005).

Banana streak virus is a bacilliform double-stranded DNA (dsDNA) virus with a monopartite genome of 7–8 Kb long with three ORFs (**Figure 3B**). ORF 1 encodes a small protein associated with virions (Cheng et al., 1996). OFR 2 encodes a protein (∼14 KDa) involved with virion assembly (Cheng et al., 1996). ORF 3 encodes a 208 kDa polyprotein comprising of a coat protein, MP, reverse transcriptase, aspartic protease, and ribonuclease H functions (Harper and Hull, 1998). The polyprotein is cleaved post-translationally by the aspartic protease into functional proteins.

Banana streak virus occurs in episomal and endogenous forms. The BSV genome sequence integrated into the host's plant B genome is known as endogenous BSV (eBSV; Harper et al., 1999; Chabannes et al., 2013; **Figure 3C**), whereas the virus genome in the replicative form in the cell are known as episomal form. Multiple copies of eBSV sequences are integrated as direct and inverted tandem repeats at a single locus in the host B genome (Chabannes et al., 2013). The integrated form of eBSV remains dormant and develop no symptom. Under stress conditions, such as those experienced during environmental stress, micro-propagation, and interspecific crossing, eBSV becomes a functional episomal producing infectious viral particles, leading to disease symptoms in the banana plants with integrated eBSV (Coˆte et al., 2010; Tripathi et al., 2019). The natural field transmission of BSV is through mealybugs or the use of infected planting material. However, the epidemics happen due to the activation of eBSV, not through the natural transmission.

Banana streak virus was not considered a severe threat to banana production until early 2,000. Several outbreaks of the

disease reported in promising breeding lines and interspecific *Musa* hybrids as micropropagation and hybridization through conventional breeding triggered its activation (Dallot et al., 2001; Lheureux et al., 2003). Currently, BSV is considered a significant constraint in banana breeding programs, particularly for plantain (AAB genome), an important staple food in Africa. The use of the diploid progenitor *M. balbisiana* or its derivants carrying a B genome is restricted as parents for introgression of desirable agronomic traits (Duroy et al., 2016). It also limits the germplasm movement of genotypes with the B genome worldwide due to the potential activation of eBSV into the episomal infectious form of the virus. Control of BSV is difficult due to genomic integration and clonal propagation.

Cucumber mosaic virus is a positive-sense RNA virus with a tripartite genome infecting many plant species. The genome of CMV consists of three genomic RNAs (1, 2, and 3), which are necessary for systemic infections in plants (Palukaitis et al., 1992). RNA 1 and 2 are components of the CMV replicase and encode the 1a and 2a proteins, respectively. RNA 3 encodes two proteins, viral MP and the viral coat protein, expressed from subgenomic RNA 4. The MP protein facilitates the movement of CMV RNA from cell-to-cell. CMV has been found in banana-growing areas worldwide, causing chlorosis, mosaic, and heart rot (Niblette et al., 1994). CMV infection does not significantly impact banana production as BBTV or BSV; however, the infections may be severe under certain circumstances. It has been reported that CMV infection of banana has caused severe diseases in Morocco (Bouhida and Lockhart, 1990) and 100% yield losses of banana in China (Li, 1995).

CRISPR/CAS-BASED GENOME EDITING OF BANANA

Genome editing technologies facilitated by various sequence/site-directed nucleases (SDN), such as zinc-finger nucleases, meganucleases, transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats/CRISPR-associated protein (CRISPR/Cas), have emerged as powerful tools for crop improvement and functional genomics. CRISPR/Cas has rapidly become the most popular genome-editing approach because of its simplicity, efficiency, versatility, specificity, and multiplexing (Scheben et al., 2017; Ntui et al., 2020; Tripathi et al., 2020).

The CRISPR-Cas system is based on the adaptive immune system of *Streptococcus pyogenes* that eliminates invasion of foreign plasmid or viral DNA. The CRISPR/Cas editing system consists of two main components: gRNA (guide RNA) and the Cas nuclease. The Cas protein exhibits nuclease activity, recognizes target DNA by gRNA-DNA pairing between the 5′ leading sequence of gRNA. It also recognizes the PAM (Protospacer adjacent motif) sequence and starts editing upstream of the sequence (Schiml and Puchta, 2016). The PAM is a three-nucleotide sequence (NGG or NAG) serving as a recognition segment for Cas to start editing upstream. The gRNA contains a scaffold and a user-defined spacer sequence (approx. 20 nt) for genomic sequence targeting. It directs the Cas to induce precise double-stranded breaks at a target site.

The natural DNA repair mechanism of the host recognizes the breaks and repairs it using the homology-directed repair (HDR) or non-homologous end joining (NHEJ) to produce the desired mutation (Weinthal and Gürel, 2016). The genome-editing takes advantage of the targeted break and the host's natural repair mechanisms to introduce the precise, targeted changes or modifications. These modifications can be a small deletion, substitution, or the addition of nucleotides. Based on the type of repair, the editing can be SDN1, SDN2, or SDN3 (Podevin et al., 2013). SDN1 relies on the spontaneous repair of the double-stranded break by NHEJ. As NHEJ is an error-prone repair, it can lead to random mutations in the host genome, causing gene silencing, gene knock-out, or alteration in the gene function. SDN2 repairs the cleavage through HDR using a repair template complementary to the break site and containing one or few nucleotide changes and copied into the host's genome during the repair mechanism resulting in a mutation of the target gene. SDN3 also repairs the double-stranded break via HDR using the repair template; however, in this case, the repair template is more prolonged, which might be of allelic, additional, or foreign gene, leading to the targeted insertion of the genetic material.

The types of genome-editing need to be distinguished due to potential discrepancies in the regulatory approaches. SDN1 and SDN2 are similar to mutations obtained through chemical mutagenesis, irradiation, or spontaneous natural mutations and do not lead to the insertion of foreign DNA (Schmidt et al., 2020). Any foreign DNA integrated into the plants during the genome-editing process is segregated out by crossing, especially in the seed crops. The final SDN1 and SDN2 type genome-edited products contain the desired mutations, but no foreign DNA is integrated and not considered genetically modified organisms (GMO). Segregating out the foreign gene integrated into the vegetatively propagated crop such as banana is challenging. SDN1 and SDN2 types of products with no foreign gene insertion in such crops can be obtained using the DNA-free genome-editing method, such as direct delivery of preassembled complexes of purified Cas9 protein-gRNA ribonucleoproteins (RNP) or by transient expression of the CRISPR/Cas construct. The SDN1 and SDN2 type of genome-edited products with no foreign DNA integration in the plant genome is not regulated as GMO in several countries such as Argentina, Australia, Brazil, Canada, Chile, Colombia, Japan, Israel, and the United States (Schmidt et al., 2020; Tripathi et al., 2020). However, the SDN3 type of product is subjected to regulatory controls as GMO, if the insert constitutes a foreign gene (Schmidt et al., 2020).

The CRISPR/Cas system has been extensively used for the genetic improvement of many crops (Scheben et al., 2017; Tripathi et al., 2019). Genome-edited products with improved traits can enhance yield potential by reducing the losses due to biotic and abiotic stresses. The availability of a well-annotated whole-genome reference sequence of the banana genome[1], advancement in bioinformatics tools, and robust genetic transformation protocols (Tripathi et al., 2015)

[1]http://banana-genome-hub.southgreen.fr

make CRISPR/Cas a suitable technology to develop disease-resistant banana.

Genome-editing in banana was first demonstrated in the cultivar "Rasthali" (AAB genome), targeting the PDS as a marker gene (Kaur et al., 2018). In this report, the authors have used a single gRNA and created mutations in the PDS gene leading to an albino phenotype. But with a relatively low mutation efficiency of 59%. In the same year, Naim et al. (2018) reported the mutation of the PDS gene in "Cavendish Williams" (AAA genome) with higher editing efficiency of 100% using polycistronic gRNAs. Similarly, Ntui et al. (2020) demonstrated high mutation efficiency of 100% using multiple gRNAs targeting the PDS gene in banana cultivar "Sukali Ndiizi" (AAB genome) and plantain cultivar "Gonja Manjaya" (AAB genome). The PDS gene is generally used as a visual-marker gene for establishing genome-editing systems in plants. The PDS gene encodes a key enzyme in the carotenoid biosynthesis pathway and converts phytoene into carotenoid precursors phytofluene and ζ-carotene. The disruption of its function leads to albino phenotype, which is a visible indicator. However, the knockout of PDS negatively affects plant growth. Alternatively, Zorrilla-Fontanesi et al. (2020) used *RP43/CHAOS39*, a gene encoding the chloroplast signal recognition particle (cpSRP) machinery, as a visual marker to optimize the genome-editing protocol for banana. The *CHAOS39* edited banana plants showed pale-green phenotypes with normal *in vitro* growth. Nevertheless, the researchers need to be careful in using *cpSRP43/CHAOS39* as a visual marker as the pale green phenotype can be achieved with other factors such as nutrient deficiency and improper light.

Our laboratory has established a robust genome editing platform for banana and plantain by using the multiplexed CRISPR/Cas9 system. This platform is now routinely used for the generation of genome-edited banana and plantain for disease resistance. Application of CRISPR/Cas9-based genome-editing system for banana was demonstrated by inactivating the endogenous eBSV sequence integrated into the B genome of plantain using multiple gRNAs (Tripathi et al., 2019). Recently, CRISPR/Cas9 technology was applied to create mutations in the *M. acuminata gibberellin* 20ox2 (MaGA20ox2) gene to develop semi-dwarf plants of the banana cultivar "Gros Michel" by disrupting gibberellin production (Shao et al., 2020).

APPLICATION OF CRISPR/CAS FOR DIAGNOSIS OF BANANA VIRAL DISEASES

Sensitive and reliable diagnostic tools for detecting BBTV and BSV are crucial for surveillance programs, clean planting material production, and phenotyping. Many serological and nucleic acid-based methods have been established to detect these two viruses (Kumar et al., 2015). Polymerase chain reaction (PCR)-based methods are widely used for the detection of BBTV and BSV. However, immunocapture PCR or rolling circle amplification methods are used for distinguishing episomal and endogenous forms of BSV. The PCR-based methods offer reliable detection of the two banana infecting viruses, but they require sophisticated equipment and laboratory facilities, limiting the adoption of these methods. The ability of CRISPR/Cas precise targeting of nucleotide sequences has been harnessed to develop highly sensitive and rapid isothermal diagnostic tools to detect viruses, bacteria, and cancer diagnosis (Chertow, 2018; Khambhati et al., 2019). Similar to genome editing, a gRNA aids in the specific recognition of a target nucleic acid sequence and activates enzymatic cleavage of the target nucleic acid by the Cas enzyme, which is then detected using a chromogenic or fluorometry detection system.

The CRISPR/Cas9-based tools were first used to detect the Zika virus after isothermal amplification of the target RNA (Pardee et al., 2016). In this method, the viral RNA was amplified by reverse transcription (RT)-PCR or RT isothermal amplification methods. The amplified product was detected using the CRISPR/Cas9 cleaving of the amplified DNA and the results were visualized by colorimetric toehold using RNA switch sensors. An improved method for the same virus using CRISPR/Cas9 triggered isothermal exponential amplification reaction (CAS-EXPAR) offered detection of viral genome at attomolar (aM) sensitivity and single-base specificity capable of differentiating African and American strains by colorimetric detection using SYBR Green florescence signal (Huang et al., 2018). The discovery of RNA-guided, RNA-targeting CRISPR effector Cas13a, and subsequently identified Cas12a, Cas13b, and Cas14a facilitated CRISPR-Cas12a, Cas13a, Cas13b, and Cas14-based nucleic methods for the detection of several human-infecting viruses and bacterial pathogens (Gootenberg et al., 2017, 2018). Of these, Cas13 types are suitable for direct detection of homologous RNA targets using RNA guide, and the Cas12 and Cas14 types are appropriate for the detection of single-stranded and dsDNA-targets, respectively, as these enzymes direct RNA guides to homologous DNA targets. A diagnostic procedure named the "Specific High Sensitivity Enzymatic Reporter UnLOCKing (SHERLOCK)" system was developed to detect the target sequence by isothermal amplification of target molecule with Recombinase Polymerase Amplification (RPA)/Reverse Transcriptase (RT)-RPA or Loop-mediated Isothermal Amplification (LAMP)/RT-LAMP (Gootenberg et al., 2017). The target amplicons are then subjected to *in vitro* T7 transcription, followed by the detection of RNA molecules generated by Cas13-guided reporter assay. The final products are detected by chromogenic detection on a later-flow device or with a fluorometer when fluorescent-labeled probes are used (Gootenberg et al., 2018). As a further improvement, the HUDSON (heating unextracted diagnostic samples to obliterate nucleases) method was standardized for the direct detection of viruses in the bodily fluid samples without nucleic acid extraction step for sensitive, rapid, and instrument-free detection of target viral nucleic acid molecule by pairing with the SHERLOCK (Myhrvold et al., 2018). This method was shown to be sensitive for detection of virus species, strains, and clinically relevant mutations directly, without nucleic acid extraction to speed up the virus detection Myhrvold et al., 2018). The Cas12a and Cas12b enzymes were also used for developing CRISPR/Cas based diagnostics. The Cas12a recognizes T rich PAM sequence for targeted cleavage of dsDNA targets using a

method termed DETECTR (DNA endonuclease targeted CRISPR *trans*-reporter; Chen et al., 2018). In this, target RNA or DNA molecule is amplified by PCR/RT-PCR or isothermal amplification methods (RPA/RT-RPA, LAMP/RT-LAMP). The amplified dsDNA products are detected using the sg-RNA-Cas12a complex, which triggers the degradation of the ssDNA fluorophore-quencher reporter probe. The results are detected by colorimetric or fluorometry detection (Chen et al., 2018). The Cas14a enzyme function similar to that of the Cas12 system, except that it detects ssDNA (Harrington et al., 2018). Most of the CRISPR-Cas diagnostics involve pre-amplification of the target molecule by PCR or RT-PCR, or isothermal amplification methods such as RPA or RT-RPA, LAMP, or RT-LAMP depending on the type of target virus genome. The target molecules are detected using the signals generated from fluorophore-quencher-based reporter RNA molecule or by separating reactions on lateral flow devices or SYBR Green fluorescence detection system (Wang et al., 2019). For BBTV, an Exo-RPA isothermal detection system targeting the BBTV-R DNA segment has already been developed and standardized for the virus detection directly in the plant tissue without the need for DNA extraction (IITA News, 2019). The amplified products are detected using the FAM-labeled Exo-RPA probes in a fluorometer (Khambhati et al., 2019). Efforts are ongoing to convert the BBTV Exo-RPA assay into a HUDSON-SHERLOCK detection system for low-cost, rapid, and sensitive detection of the virus both under laboratory and field conditions. Simultaneously, existing PCR-based methods for BSV detection are being converted to CRISPR/Cas diagnostics to detect BSV and its variants in banana. Because of high specificity and sensitivity for detecting near single copy of the target molecules, CRISPR-based diagnostics tools have the potential to offer highly robust, rapid, and sensitive detection of banana viruses for seed health certification, surveillance, and other applications.

Recently, an efficient and rapid RT-RPA-CRISPR/Cas12a system has been developed as a one-step detection assay to diagnose plant RNA viruses (Aman et al., 2020). This diagnostic assay uses a fluorescence visualizer to detect the plant RNA viruses. It can be performed in less than 30 min at a single temperature in the field. This assay could be used quickly and efficiently to detect banana RNA viruses fast-tracking their containment strategies.

CONTROL OF VIRAL DISEASES USING CRISPR/CAS

The CRISPR/Cas system is becoming the method of choice to control plant viruses, either by targeting the viral factors for viral genome editing in viruses or by targeting the host plant factors responsible for the viral cycle. CRISPR/Cas-based genome editing for controlling plant viruses is reported for ssDNA viruses, dsDNA viruses, and ssRNA viruses (Baltes et al., 2015; Ali et al., 2016; Hadidi et al., 2016; Zaidi et al., 2016; Zhang et al., 2018; Gomez et al., 2019; Tripathi et al., 2019; **Table 1**). Although most of the application of the CRISPR/Cas to develop virus-resistance is reported for ssDNA and dsDNA viruses in plants, its application against RNA viruses is also reported (Zhao et al., 2020). The CRISPR/Cas-based resistance to plant RNA viruses is based on the editing of host plant factors influencing viral infection rather than the viral genes. The plant host factors like eukaryotic translation initiation factor (eIF) are required to maintain replication of RNA viruses. Several eIF such as eIF4E and eIF(iso)4E, have been identified as recessive resistance alleles to confer resistance against several potyviruses (Khatodia et al., 2017). The genome-edited plants with mutations in the eIF(iso)4E gene demonstrated enhanced resistance against cucumber vein yellowing virus, cassava brown streak virus, Ugandan cassava brown streak virus, papaya ringspot virus-type W, zucchini yellow mosaic virus, and turnip mosaic virus (Chandrasekaran et al., 2016; Pyott et al., 2016; Gomez et al., 2019).

Banana bunchy top virus is a multipartite ssDNA virus that replicates either by the host or virally encoded DNA polymerases through a dsDNA intermediate form during replication. In contrast, BSV is a monopartite dsDNA virus, with a DNA-RNA intermediate during replication similar to that of pararetrovirus replication. The dsDNA structure of both viruses makes them a good target for CRISPR/Cas9 mediated genome editing. Although no progress on CRISPR/Cas mediated resistance against BBTV has been reported, several investigations have documented the use of CRISPR/Cas9 to induce durable virus resistance to ssDNA viruses through gene knockout in many plant species (Yin and Qiu, 2019; Kalinina et al., 2020; **Table 1**). These mechanisms could be harnessed to develop resistance to BBTV in banana.

CRISPR/Cas-mediated resistance to DNA viruses was first demonstrated for geminiviruses (Ali et al., 2015; Baltes et al., 2015; Ji et al., 2015). The CRISPR/Cas9 technology was applied for developing resistance against beet severe curly top virus (BSCTV, genus *Geminivirus*) in Arabidopsis and *Nicotiana benthamiana,* targeting the replication-associated protein (Rep), coat protein (CP), and intergenic region (IR; Ji et al., 2015). The genome-edited plants demonstrated high resistance against BSCTV with up to 87% reduction in viral load. Similarly, a high level of resistance against the bean yellow dwarf virus (BeYDV, genus *Mastrevirus*) in *N. benthamiana* using CRISPR/Cas9 plasmid construct targeting the *Rep* gene delivered through *Agrobacterium*-mediated transformation was recorded (Baltes et al., 2015). The edited plants expressing CRISPR-Cas reagents showed targeted mutations within the viral genome and demonstrated reduced virus load and symptoms upon challenge with BeYDV. These studies reported novel strategies for controlling geminiviruses.

In another approach, *N. benthamiana* was engineered with resistance against the tomato yellow leaf curl virus (TYLCV, genus *Begomovirus*), a monopartite begomovirus, by transiently delivering a CRISPR/Cas9 construct targeting the viral *Rep*, *CP*, and the conserved region of IR of TYLCV using the tobacco rattle virus (TRV) vector (Ali et al., 2015). The *N. benthamiana* plants exhibited resistance to TYLCV. Further, stable transgenic edited plants were generated through *Agrobacterium*-mediated

TABLE 1 | Summary of developing virus resistance in plant species using CRISPR/Cas-based genome-editing.

Plant	Editing system	Target gene	Target trait	Outcome	References
Arabidopsis	CRISPR-Cas9	*eIF4E*	Turnip mosaic virus (TuMV)	Complete resistance to TuMV	Pyott et al., 2016
Arabidopsis	CRISPR-Cas9	*CP*	Cauliflower mosaic virus (CaMV)	Resistance to CaMV	Liu et al., 2018
Arabidopsis	CRISPR-Cas9	Viral genome of CMV	Cucumber mosaic virus (CMV), Tobacco mosaic virus (TMV)	Resistance to CMV and TMV	Zhang et al., 2018
Banana	CRISPR-Cas9	ORF1, ORF2, and ORF3 of BSV	Banana streak virus (BSV)	Inactivation of integrated endogenous BSV (eBSV)	Tripathi et al., 2019
Barley	CRISPR-Cas9	*CP, MP, Rep, RepA*	Wheat dwarf virus (WDV)	Resistance/tolerance to WDV	Kis et al., 2019
Cassava	CRISPR-Cas9	*nCBP-1 and nCBP-2*	Cassava brown streak virus *(CBSV),* Ugandan cassava brown streak virus (UCBSV)	Reduction in disease symptom severity and incidence	Gomez et al., 2019
Cassava	CRISPR-Cas9	*AC2 and AC3*	African Cassava mosaic virus (ACMV)	No resistance against ACMV	Mehta et al., 2019
Cucumber	CRISPR-Cas9	*eIF4E*	Cucumber vein yellowing virus (CVYV), Zucchini yellow mosaic virus (ZYMV), Papaya ringspot virus-type W *(*PRSV-W*)*	Resistance to multiple viruses (CVYN, ZYMV, PRSV-W)	Chandrasekaran et al., 2016
Potato	CRISPR-Cas13a	*P3*, *CI*, *Ni*, *CP*	Potato virus Y (PVY)	Suppressed PVY accumulation and disease symptoms	Zhan et al., 2019
Rice	CRISPR-Cas9	*eIF4G*	Rice tungro spherical virus (RTSV)	Resistance to RTSV	Macovei et al., 2018
Rice	CRISPR-Cas13a	Three regions in viral genome	Southern rice black-streaked dwarf virus (SRBSDV)	Reduction in disease symptoms	Zhang et al., 2019
Rice	CRISPR-Cas13a	Three regions in viral genome	Rice stripe mosaic virus (RSMV)	Reduction in disease symptoms	Zhang et al., 2019
Tobacco	CRISPR-Cas9	Coding and non-coding region of TYMV	Tomato yellow leaf curl virus (TYLCV)	Significant reduction in disease symptoms	Ali et al., 2015
Tobacco	CRISPR-Cas9	*CP, Rep*	Cotton leaf curl kokhran virus (CLCuKoV)	Reduced viral activities	Ali et al., 2016
Tobacco	CRISPR-Cas9	*CP, RCRII*	Merremia mosaic virus (MeMV)	Reduced viral activities	Ali et al., 2016
Tobacco	CRISPR-Cas9	*6* regions in the viral genome	Bean yellow dwarf virus (BeYDV)	Reduction in viral accumulation	Baltes et al., 2015
Tobacco	CRISPR-Cas9	Viral genome	Beet severe curly top virus (BSCTV)	Reduction in viral accumulation	Ji et al., 2015
Tobacco	CRISPR-Cas9	*IR, CP, Rep*	Tomato yellow leaf curl virus (TYLCV)	Resistance to TYLCV	Tashkandi et al., 2018
Tobacco	CRISPR-Cas9	Viral genome of CMV	Cucumber mosaic virus (CMV), Tobacco mosaic virus (TMV)	Resistance to CMV and TMV	Zhang et al., 2018
Tobacco	CRISPR-Cas13a	*4* regions of TuMV	Turnip mosaic virus (TuMV)	Virus interference	Aman et al., 2018
Tobacco	CRISPR-Cas9	*IR, V2/V1, C1/C4*	Chili leaf curl virus (ChiLCV)	Reduction in viral accumulation	Roy et al., 2019
Tobacco	CRISPR-Cas9	*IR and CI* coding regions	Cotton leaf curl multan virus (CLCuMV)	Complete resistance to CLCuMV	Yin et al., 2019
Tobacco	CRISPR-Cas9	*Rep*,*βC1*	Cotton leaf curl virus (CLCuV)	Delayed symptom development and reduction in viral accumulation	Khan et al., 2020
Tomato	CRISPR-Cas9	*IR, CP, Rep*	Tomato yellow leaf curl virus (TYLCV)	Resistance	Tashkandi et al., 2018

transformation using the same CRISPR/Cas9 construct. These transgenic *N. benthamiana* plants exhibited broad-spectrum resistance against the monopartite beet curly top virus (genus *Curtovirus*), the bipartite Merremia mosaic virus (MeMV, genus *Begomovirus*), and TYLCV. Later, *N. benthamiana* was engineered using a CRISPR/Cas9 system to interfere with the coding sequences of TYLCV, MeMV, and cotton leaf curl Kokhran virus (genus *Begomovirus*; Ali et al., 2016). However, this led to the emergence of a new mutated virus variant, which evaded the CRISPR/Cas9 activity, and viruses continued to replicate and spread systemically. Further, when the IR sequences were targeted, the new mutated virus variants were not detected, suggesting that targeting non-coding viral sequences may be better for controlling multiple geminiviruses simultaneously using CRISPR/Cas. Similarly, tomato plants exhibiting resistance against TYLCV were produced using the CRISPR/Cas9 system targeting the IR region (Tashkandi et al., 2018).

Recently, complete resistance against the cotton leaf curl virus (CLCuV, genus *Begomovirus*) was demonstrated in *N. benthamiana* using a CRISPR/Cas9 system by multiplexing gRNAs targeting *Rep* and IR sequences (Yin et al., 2019). Later, *N. benthamiana* plants agroinfiltrated with CRISPR/Cas9

construct having multiplex gRNAs targeting the *Rep*, and β*C1* gene of the betasatellites showed delayed disease symptoms and lower titer of CLCuV (Khan et al., 2020). Interestingly, the inactivation of the *AC2* gene encoding the transcription activator protein and the *AC3* gene encoding the replication enhancer protein of African cassava mosaic virus (ACMV, genus *Begomovirus*) failed to produce resistance against the virus in transgenic cassava (Mehta et al., 2019). The authors reported that CRISPR editing led to the formation and escape of new CRISPR-resistant ACMV variants, which were probably generated due to the post cleavage NHEJ repair.

Genome-edited barley plants with resistance to the wheat dwarf virus (genus *Mastrevirus*) were generated by multiplexing four gRNAs targeting the overlapping region of *MP* and *CP* coding sequence, *Rep/RepA* coding region at the N-terminus of the proteins, long intergenic region (LIR) region, and the genomic region encoding the C-terminus of Rep (Kis et al., 2019). Similar approaches using CRISPR/Cas systems targeting *MP, CP*, and *Rep* can be applied in banana to enhance resistance against BBTV. Improved resistance to BBTV by silencing the *CP, MP*, and *Rep* using the RNAi approach was demonstrated (Shekhawat et al., 2012; Krishna et al., 2013; Elayabalan et al., 2017). However, the transgenic banana plants only showed partial resistance to BBTV. It might probably be because RNAi does not always result in a complete knockout; therefore, genome-editing could potentially be used to simultaneously knocked out several genes. The banana plants resistant to BBTV can be developed by delivering the CRISPR/Cas9 reagents targeting either the viral genes such as *CP, MP*, and *Rep*, or the host plant factors involved in viral infection like *eIF* gene.

The CRISPR/Cas system has also been used to successfully provide resistance against other DNA viruses, besides ssDNA viruses. For example, a CRISPR/Cas9 system was used to enhance resistance against cauliflower mosaic virus (genus *Caulimovirus*), a dsDNA pararetrovirus, in *Arabidopsis* by creating targeted mutations in the coat protein (Liu et al., 2018). However, the authors also reported some mutated forms of the virus, which could escape and spread in systemically infected leaves.

CRISPR/Cas technology can also be applied to pararetroviruses such as BSV or retroviruses with dsDNA as part of their replication. Currently, the application of CRISPR/Cas9 to disrupt both episomal and integrated dsDNA viruses is reported for only one plant virus (Tripathi et al., 2019). However, it has been demonstrated for control of several human viruses, including papillomaviruses (HPV16 and HPV18), hepatitis B virus, Epstein-Barr virus (EBV), HIV-1, polyomavirus 2 (John Cunningham Virus), Herpes simplex virus-1, and other herpesviruses (White et al., 2016). CRISPR/Cas system has been used to treat HIV-1 infection by targeting both viral and host factors (Chen et al., 2018). HIV is a retrovirus integrating the viral DNA into the host DNA and reactivates, causing HIV-AIDS. CRISPR/Cas9 technology was used to inactivate the HIV-1 by knocking out proviral DNA integrated into the host cells by targeting long terminal repeat (LTR) flanking sequences or overlapping ORFs or multiple regulatory genes within the HIV-1 provirus (Wang et al., 2016; Wang et al., 2016; Ophinni et al., 2018). CRISPR/Cas9 targeting the LTR significantly suppresses the activation of HIV-1; however, viral escape was also reported (Wang et al., 2016). The probability of the virus escape could be overcome by multiplexing gRNAs targeting different ORFs (White et al., 2016). Similarly, CRISPR/Cas9 targeting multiple ORF targets (E6 and E7) was used to inactivate HPV, a dsDNA virus that gets integrated into the chromosomes of host cells (Kennedy et al., 2014; Zhen et al., 2014). Similarly, a CRISPR-SpCas9 tool was applied, targeting six different regions in the EBV genome (Wang and Quake, 2014). Subsequently, Yuen et al. (2015) multiplexed a CRISPR/Cas9 system targeting a 558 bp fragment in the promoter region of BART (*Bam*HI A rightward transcript) and a primary viral transcript encoding the viral microRNAs (miRNAs).

Similarly, a multiplex CRISPR/Cas9 system targeting all three ORFs of BSV was used to inactivate the integrated dsDNA of eBSV from the banana genome (Tripathi et al., 2019). The CRISPR/Cas9 construct with multiple gRNAs targeting the ORF1, ORF2, and aspartic protease gene of ORF3 was delivered into the host cells of plantain "Gonja Manjaya" through *Agrobacterium*-mediated genetic transformation to inactivate the virus. The regenerated genome-edited plantain with targeted mutations in the viral genome prevented proper transcription or/and translation into infectious viral proteins. Phenotyping of the potted genome-edited plants under water stress conditions in the greenhouse confirmed the inactivation of the virus as 75% of the tested plants remained asymptomatic under stress conditions; in contrast, the wildtype control plants showed BSV disease symptoms. All the asymptomatic plants have mutations in all the three ORFs. The application of CRISPR/Cas9-mediated genome editing of banana for controlling viruses is only reported for BSV.

There are reports of the emergence of new mutated virus variants in plants and human using the CRISPR-Cas9 with a single sgRNA targeting viral genes (Ali et al., 2016, Wang et al., 2016, Mehta et al., 2019). The emergence of mutated new variants of the viruses, which may be hypervirulent, can be delayed by targeting multiple viral genes for editing, using the more efficient versions of Cas9, targeting the non-coding region of virus genome, or deleting the larger portions of the viral genome (White et al., 2016; Mehta et al., 2019). The variability in the virus sequences also plays a role in circumventing resistance in the plants expressing the CRISPR machinery targeting viral genes. The CRISPR-Cas9 system should be carefully designed to engineer virus resistance to minimize chances of viral escape.

CHALLENGES IN GENOME-EDITING OF BANANA

Banana is a polyploid heterozygous crop containing a high number of multigene families with paralogs (Cenci et al., 2014). One of the significant challenges of genome editing in a banana is to target multiple alleles and gene copies simultaneously.

Sometimes knockdown or knockout of a particular gene does not result in any phenotypic change, maybe due to the dose-effect of other paralogous copies of genes. Therefore, the gRNA needs to be designed to target all the copies and alleles of the gene and screen many mutants to recover an edited line with multiallelic mutations. Multiplexed genome-editing using multiple gRNAs targeting several genes and their paralogs in a gene family can be an efficient tool for improving polyploid crops (Ansari et al., 2020).

Another major challenge is that genome editing in a banana crop is currently achieved by plasmid-based delivery through *Agrobacterium* transformation (Ntui et al., 2020). In other crops, the CRISPR reagents are delivered through a range of transformation methods, such as protoplast transfection, agroinfiltration, and stable transformation through *Agrobacterium* and microprojectile bombardment (Nadakuduti et al., 2018). However, transient delivery systems like agroinfiltration or protoplast fusion are not successful in banana. The mutants generated through stable transformation are considered GMOs due to transgene integration in the plant genome and had to go through time-consuming regulatory approvals, and this could reduce their acceptability. Since banana is a vegetatively propagated crop, segregating out of the Cas9 gene, marker gene, and *Agrobacterium*–derived DNA sequences through backcrossing is not feasible due to the sterility of the majority of the farmer-preferred cultivars, unlike seeded crops (Nadakuduti et al., 2018; Tripathi et al., 2020). To overcome these biosafety concerns, there is a need to develop transgene-free genome-edited banana plants. Recently, several efforts have been put to produce transgene-free genome-edited plants by directly delivering preassembled Cas9 protein-gRNA RNP, otherwise known as RNPs, into plant cells (Woo et al., 2015; Malnoy et al., 2016; Svitashev et al., 2016; Liang et al., 2018; Tripathi et al., 2020). Upon delivery, the RGENs-RNPs edit the target sites immediately and are rapidly degraded by endogenous proteases in cells, leaving no traces of foreign DNA elements (Kanchiswamy et al., 2015; Woo et al., 2015).

In banana, preassembled RGENs-RNPs targeting viral genes or plant host factors could be coated to gold particles and delivered to banana cell suspension cultures by particle bombardment. Alternatively, the CRISPR/Cas9 constructs targeting the viral genes or plant host factors could transiently be delivered into banana cells through microprojectile bombardment. The complete plants can be regenerated from the bombarded banana cells. The edited plants with the targeted mutations and absence of foreign gene integration should be selected based on the molecular characterization. The virus-resistant genome-edited plants generated using preassembled RGENs-RNPs or transient delivery of CRISPR/Cas9 reagents will not have any foreign gene integration and might not require GMO regulatory approval (Tripathi et al., 2020).

The major biosafety concerns with genome-edited crops are unwanted genetic changes in plants due to off-target mutations and transgene integration. The off-target effects can be minimized by improving strategies for designing the gRNA very specific to the target and RNP-based delivery as they are active for a short duration in the host cell and use of inducible CRISPR/Cas system to avoid the strong doses of CRISPR-Cas9 beyond the target cells/tissues and throughout the life span of the plant.

Another pressing issue is the lack of high throughput screening methods to identify genome-edited events. Currently, the edited plants of banana are screened using PCR and target sequencing to detect the mutation. However, these methods are expensive and time-consuming. Screening the edited plants using the high-throughput phenotyping for the desired trait, followed by target sequencing of the selected events, will be more efficient and cost-efficient.

The challenges with the differences among the countries regarding the regulation of genome-edited products cannot be ignored (Tripathi et al., 2020). The genome-editing products with no foreign gene integration are not regulated in many countries (Schmidt et al., 2020). Only the EU and New Zealand consider genome-edited products under the existing GMO biosafety (Schmidt et al., 2020).

CONCLUSION

Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein-based genome editing is fast revolutionizing its applications in crop improvement for desired traits such as disease resistance. BBTV and BSV are economically important viruses threatening banana production. The most sustainable way to reduce losses due to these viral diseases is the use of virus-resistant banana varieties. In the absence of any known resistant germplasm, it requires the development of durable virus-resistant varieties using modern biotechnological tools, complementing conventional breeding. Genome editing tools provide a new weapon in the arsenal against plant viruses. The availability of a robust genome-editing system and reference genome make the banana a potential candidate for developing virus-resistant varieties using CRISPR/Cas-based genome editing. So far, genome-editing is applied to banana only for control of BSV by inactivating the eBSV integrated into the host genome. CRISPR/Cas system targeting the essential genes of the virus or host plant genes involved in the susceptibility can be applied in banana to develop resistance against BBTV. Genome-edited virus-resistant banana varieties can be generated with no foreign gene integration, making them more acceptable for commercialization.

AUTHOR CONTRIBUTIONS

LT conceived the idea. All authors contributed to writing, reviewing, and editing the manuscript.

ACKNOWLEDGMENTS

The authors wish to thank CGIAR Research Program on Roots, Tubers and Banana (CRP-RTB) for financial support.

REFERENCES

Adegbola, R., Ayodeji, O., Awosusi, O., Atiri, G., and Kumar, P. L. (2013). First report of *Banana bunchy top virus* in banana and plantain (*Musa* spp .) in Nigeria. *Disease Note* 97, 10–13. doi: 10.1094/PDIS-08-12-0745-PDN

Ali, Z., Abulfaraj, A., Idris, A., Ali, S., Tashkandi, M., and Mahfouz, M. M. (2015). CRISPR/ Cas9-mediated viral interference in plants. *Genome Biol.* 16:238. doi: 10.1186/s13059-015-0799-6

Ali, Z., Ali, S., Tashkandi, M., Zaidi, S. S., and Mahfouz, M. M. (2016). CRISPR/Cas9-mediated immunity to geminiviruses: differential interference and evasion. *Sci. Rep.* 6:26912. doi: 10.1038/srep26912

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., et al. (2018). RNA virus interference via CRISPR/Cas13a system in plants. *Genome Biol.* 19:1. doi: 10.1186/s13059-017-1381-1

Aman, R., Mahas, A., Marsic, T., Hassan, N., and Mahfouz, M. M. (2020). Efficient, rapid and sensitive detection of plant RNA viruses with one-pot RT-RPA– 2 CRISPR/Cas12a assay. *Front. Microbiol.* 11:610872. doi: 10.3389/fmicb.2020.610872

Ansari, W. A., Chandanshive, S. U., Bhatt, V., Nadaf, A. B., Vats, S., Katara, J. L., et al. (2020). Genome editing in cereals: approaches, applications and challenges. *Int. J. Mol. Sci.* 21:4040. doi: 10.3390/ijms21114040

Baltes, N. J., Hummel, A. W., Konecna, E., Cegan, R., Bruns, A. N., Bisaro, D. M., et al. (2015). Conferring resistance to geminiviruses with the CRISPR – Cas prokaryotic immune system. *Nat. Plants* 1:15145. doi: 10.1038/nplants.2015.145

Bouhida, M., and Lockhart, B. E. (1990). Increase in importance of cucumber mosaic virus infection in greenhouse-grown bananas in Morocco. *Phytopathology* 80:981.

Burns, T. M., Harding, R. M., and Dale, J. L. (1995). The genome organization of *Banana bunchy top virus*: analysis of six ssDNA components. *J. Gen. Virol.* 76, 1471–1482. doi: 10.1099/0022-1317-76-6-1471

Cenci, A., Guignon, V., Roux, N., and Rouard, M. (2014). Genomic analysis of NAC transcription factors in banana (*Musa acuminata*) and definition of NAC orthologous groups for monocots and dicots. *Plant Mol. Biol.* 85, 63–80. doi: 10.1007/s11103-013-0169-2

Chabannes, M., Baurens, F.-C., Duroy, P.-O., Bocs, S., Vernerey, S., Rodier-Goud, M., et al. (2013). Three infectious viral species lying in wait in the banana genome. *J. Virol.* 87, 8624–8637. doi: 10.1128/JVI.00899-13

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non- transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi: 10.1111/mpp.12375

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, A., Tian, X., Palesfsky, J. M., et al. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded dnase activity. *Science* 360, 436–439. doi: 10.1126/science.aar6245

Cheng, C., Lockhart, B. E. L., and Olszewski, N. E. (1996). The ORF I and II proteins of Commelina yellow mottle virus are virion-associated. *Virology* 223, 263–271. doi: 10.1006/viro.1996.0478

Chertow, D. S. (2018). Next-generation diagnostics with CRISPR. *Science* 360, 381–382. doi: 10.1126/science.aat4982

Coîe, F. X., Galzi, S., Folliot, M., Lamagne're, Y., Teycheney, P. Y., and Iskra-Caruana, M. L. (2010). Micropropagation by tissue culture triggers differential expression of infectious endogenous Banana streak virus sequences (eBSV) present in the B genome of natural and synthetic interspecific banana plantains. *Mol. Plant Pathol.* 11, 137–144. doi: 10.1111/j.1364-3703.2009.00583.x

Dale, J., Paul, J. Y., Dugdale, B., and Harding, R. (2017). Modifying bananas: from transgenics to organics. *Sustainability* 9:333. doi: 10.3390/su9030333

Dallot, S., Acuña, P., Rivera, C., Ramirez, P., Cote, F., Lockhart, B. E., et al. (2001). Evidence that the proliferation stage of micropropagation procedure is determinant in the expression of Banana streak virus integrated into the genome of the FHIA 21 hybrid (*Musa*AAAB). *Arch. Virol.* 146, 2179–2190. doi: 10.1007/s007050170028

Davey, M., Gudimella, R., Harikrishna, J. A., Sin, L. W., Khalid, N., and Keulemans, J. (2013). A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter-and intra-specific *Musa* hybrids. *BMC Genomics* 14:683. doi: 10.1186/1471-2164-14-683

Duroy, P. O., Perrier, X., Laboureau, N., Jacquemoud-Collet, J. P., and Iskra-Caruana, M. L. (2016). How endogenous plant pararetroviruses shed light on *Musa* evolution. *Ann. Bot.* 117, 625–641. doi: 10.1093/aob/mcw011

Elayabalan, S., Subramaniam, S., and Selvarajan, R. (2017). Construction of BBTV rep gene RNAi vector and evaluate the silencing mechanism through injection of *Agrobacterium tumefaciens* transient expression system in BBTV infected hill banana plants cv. *Virupakshi* (AAB). *Indian J. Nat. Sci.* 7, 12395–12403.

FAOSTAT (2018). Available online at: http://www.fao.org/faostat/en/#data/QC (accessed August 31, 2018).

Fargette, D., Konate, G., Fauquet, C., Muller, E., Peterschmitt, M., and Thresh, J. M. (2006). Molecular ecology and emergence of tropical plant viruses. *Annu. Rev. Phytopathol.* 44, 235–260. doi: 10.1146/annurev.phyto.44.120705.104644

Geering, A. D. W., Olszewski, N. E., Harper, G., Lockhart, B. E. L., Hull, R., and Thomas, J. E. (2005). Banana contains a diverse array of endogenous badnaviruses. *J. Gen. Virol.* 86, 511–520. doi: 10.1099/vir.0.80261-0

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., et al. (2019). Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol. J.* 17, 421–434. doi: 10.1111/pbi.12987

Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., and Zhang, F. (2018). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* 360, 439–444. doi: 10.1126/science.aaq0179

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., et al. (2017). Nucleic acid detection with CRISPR-cas13a/c2c2. *Science* 356, 438–442. doi: 10.1126/science.aam9321

Hadidi, A., Flores, R., Candresse, T., and Barba, M. (2016). Next-generation sequencing and genome editing in plant virology. *Front. Microbiol.* 7:1325. doi: 10.3389/fmicb.2016.01325

Harding, R. M., Burns, T. M., Hafner, G., Dietzgen, R. G., and Dale, J. L. (1993). Nucleotide sequence of one component of the *Banana bunchy top virus* genome contains a putative replicase gene. *J. Gen. Virol.* 74, 323–328. doi: 10.1099/0022-1317-74-3-323

Harper, G., and Hull, R. (1998). Cloning and sequence analysis of Banana streak virus DNA. *Virus Genes* 17, 271–278. doi: 10.1023/A:1008021921849

Harper, G., Osuji, J. O., Heslop-Harrison, J. S., and Hull, R. (1999). Integration of Banana streak virus infection into the *Musa* genome: molecular and cytogenetic evidence. *Virology* 255, 207–213. doi: 10.1006/viro.1998.9581

Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., et al. (2018). Programmed DNA destruction by miniature CRISPR-cas14 enzymes. *Science* 362, 839–842. doi: 10.1126/science.aav4294

Huang, M., Zhou, X., Wang, H., and Xing, D. (2018). Clustered regularly interspaced short palindromic repeats/cas9 triggered isothermal amplification for site-specific nucleic acid detection. *Anal. Chem.* 90, 2193–2200. doi: 10.1021/acs.analchem.7b04542

IITA News (2019). *Building Regional Capacity to Contain Banana Bunchy Top Disease (BBTD) Spread in West Africa. IITA Bulletin (2469)*, Vol. 1. Available online at: https://www.iita.org/wp-content/uploads/2019/01/Bulletin_2469.pdf (accessed January 18, 2019).

Ji, X., Zhang, H., Zhang, Y., Wang, Y., and Gao, C. (2015). Establishing a CRISPR-Cas-like immune system conferring DNA virus resistance in plants. *Nat. Plants* 1:15144. doi: 10.1038/nplants.2015.144

Jooste, A. E. C., Wessels, N., and van der Merwe, M. (2016). First report of *Banana bunchy top virus* in banana (*Musa* spp.) from South Africa. *Plant Dis.* 100, 1251–1252. doi: 10.1094/PDIS-12-15-1422-PDN

Kalinina, N. O., Khromov, A., Love, A. J., and Taliansky, M. E. (2020). CRISPR application in plant virology: virus resistance and beyond. *Phytopathology* 110, 18–28. doi: 10.1094/PHYTO-07-19-0267-IA

Kanchiswamy, C. N., Malnoy, M., Velasco, R., Kim, J.-S., and Viola, R. (2015). Non-GMO genetically edited crop plants. *Trends Biotechnol.* 33, 489–491. doi: 10.1016/j.tibtech.2015.04.002

Karan, M., Harding, R. M., and Dale, J. L. (1994). Evidence for two groups of banana bunchy top virus isolates. *J. Gen. Virol.* 75, 3541–3546. doi: 10.1099/0022-1317-75-12-3541

Kaur, N., Alok, A., Shivani, Kaur, N., Pandey, P., Awasthi, P., et al. (2018). CRISPR/Cas9-mediated efficient editing in *phytoene desaturase* (PDS) demonstrates precise manipulation in banana cv. Rasthali genome. *Funct. Integr. Genomics* 18, 89–99. doi: 10.1007/s10142-017-0577-5

Kennedy, E. M., Kornepati, A. V., Goldstein, M., Bogerd, H. P., Poling, B. C., Whisnant, A. W., et al. (2014). Inactivation of the human papillomavirus E6 or E7 gene in cervical carcinoma cells by using a bacterial CRISPR/Cas RNA-guided endonuclease. *J. Virol.* 88, 11965–11972. doi: 10.1128/JVI.01879-1

Khambhati, K., Bhattacharjee, G., and Singh, V. (2019). Current progress in CRISPR-based diagnostic platforms. *J. Cell Biochem.* 120, 2721–2725. doi: 10.1002/jcb.27690

Khan, S., Mahmood, M. S., Rahman, S. U., Rizvi, F., and Ahmad, A. (2020). Evaluation of the CRISPR/Cas9 system for the development of resistance against *Cotton leaf curl virus* in model plants. *Plant Protect. Sci.* 56, 154–162. doi: 10.17221/105/2019-PPS

Khatodia, S., Bhatotia, K., and Tuteja, N. (2017). Development of CRISPR/Cas9 mediated virus resistance in agriculturally important crops. *Bioengineered* 8, 274–279. doi: 10.1080/21655979.2017.1297347

Kis, A., Hamar, É, Tholt, G., Bán, R., and Havelda, Z. (2019). Creating highly efficient resistance against *Wheat dwarf virus* in barley by employing CRISPR/Cas9 system. *Plant Biotechnol. J.* 17, 1004–1006. doi: 10.1111/pbi.13077

Krishna, B., Kadu, A. A., Vyavhare, S. N., Chaudhary, R. S., Joshi, S. S., Patil, A. B., et al. (2013). RNAi-mediated resistance against *Banana bunchy top virus* (BBTV) in "Grand Nain" banana. *Acta Hortic.* 974, 157–164. doi: 10.17660/ActaHortic.2013.974.18

Kumar, P. L., Hanna, R., Alabi, O. J., Soko, M. M., Oben, T. T., Vangu, G. H. P., et al. (2011). *Banana bunchy top virus* in sub-Saharan Africa: investigations on virus distribution and diversity. *Virus Res.* 159, 171–182. doi: 10.1016/j.virusres.2011.04.021

Kumar, P. L., Selvarajan, R., Iskra-Caruana, M., Chabannes, M., and Hanna, R. (2015). Biology, etiology, and control of virus diseases of banana and plantain. *Adv. Virus Res.* 91, 229–269. doi: 10.1016/bs.aivir.2014.10.006

Lheureux, F., Carreel, F., Jenny, C., Lockhart, B. E. L., and Iskra-Caruana, M. L. (2003). Identification of genetic markers linked to banana streak disease expression in inter-specific *Musa* hybrids. *Theor. Appl. Genet.* 106, 594–598. doi: 10.1007/s00122-002-1077-z

Li, H. P. (1995). *Characterization of Cucumber Mosaic Virus Strains in Banana in China*. Ph.D. Dissertation, South China Agricultural University, Guangzhou.

Liang, Z., Chen, K., Zhang, Y., Liu, J., Yin, K., Qiu, J.-L., et al. (2018). Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins. *Nat. Protoc.* 13:413. doi: 10.1038/nprot.2017.145

Liu, H., Soyars, C. L., Li, J., Fei, Q., He, G., Peterson, B. A., et al. (2018). CRISPR/Cas9-mediated resistance to cauliflower mosaic virus. *Plant Direct* 2:e00047. doi: 10.1002/pld3.47

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Ċermák, T., et al. (2018). Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol. J.* 16, 1918–1927. doi: 10.1111/pbi.12927

Malnoy, M., Viola, R., Jung, M.-H., Koo, O.-J., Kim, S., Kim, J. S., et al. (2016). DNA- free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. *Front. Plant Sci.* 7:1904. doi: 10.3389/fpls.2016.01904

McKey, D., Elias, M., Pujol, B., and Duputié, A. (2010). The evolutionary ecology of clonally propagated domesticated plants. *New Phytol.* 186, 318–332. doi: 10.1111/j.1469-8137.2010.03210.x

Mehta, D., Stürchler, A., Anjanappa, R. B., Zaidi, S. S.-E.-A., Hirsch-Hoffmann, M., Gruissem, W., et al. (2019). Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses. *Genome Biol.* 20:80. doi: 10.1186/s13059-019-1678-3

Myhrvold, C., Freije, C. A., Gootenberg, J. S., Abudayyeh, O. O., Metsky, H. C., Durbin, A. F., et al. (2018). Field-deployable viral diagnostics using CRISPR-Cas13. *Science* 360, 444–448. doi: 10.1126/science.aas8836

Nadakuduti, S. S., Buell, C. R., Voytas, D. F., Starker, C. G., and Douches, D. S. (2018). Genome editing for crop improvement – applications in clonally propagated polyploids with a focus on potato (*Solanum tuberosum* L.). *Front. Plant Sci.* 9:1607. doi: 10.3389/fpls.2018.01607

Naim, F., Dugdale, B., Kleidon, J., Brinin, A., Shand, K., Waterhouse, P., et al. (2018). Gene editing the phytoene desaturase alleles of Cavendish banana using CRISPR/Cas9. *Transgenic Res.* 27, 451–460. doi: 10.1007/s11248-018-0083-0

Ngatat, S., Hanna, R., Kumar, P. L., Gray, S. M., Cilia, M., Ghogomu, R. T., et al. (2017). Relative susceptibility of *Musa* genotypes to banana bunchy top disease in Cameroon and implication for disease management. *Crop Protection* 101, 116–122. doi: 10.1016/j.cropro.2017.07.018

Niblette, C. L., Pappu, S. S., Bird, J., and Lastra, R. (1994). "Infectious chlorosis, mosaic, and heart rot," in *Compendium of Tropical Fruit Diseases*, eds R. C. Ploetz, G. M. Zentmyer, W. T. Nishijima, K. G. Rohrbach, and H. D. Ohr (St. Paul, MN: The American Phytopathological Society), 18–19.

Ntui, V. O., Tripathi, J. N., and Tripathi, L. (2020). Robust CRISPR/Cas9 mediated genome editing tool for banana and plantain (Musa spp.). *Cur. Plant Biol.* 21:100128. doi: 10.1016/j.cpb.2019.100128

Ophinni, Y., Inoue, M., Kotaki, T., and Kameoka, M. (2018). CRISPR/Cas9 system targeting regulatory genes of HIV-1 inhibits viral replication in infected T-cell cultures. *Sci. Rep.* 8:7784. doi: 10.1038/s41598-018-26190-1

Palukaitis, P., Roossinck, M. J., Dietzgen, R. G., and Francki, R. I. B. (1992). Cucumber mosaic virus. *Adv. Virus Res.* 41, 281–348. doi: 10.1016/s0065-3527(08)60039-1

Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., et al. (2016). Rapid, low-cost detection of Zika virus using programmable biomolecular components. *Cell* 165, 1255–1266. doi: 10.1016/j.cell.2016.04.059

Podevin, N., Davies, H. V., Hartung, F., Nogué, F., and Casacuberta, J. M. (2013). Site-directed nucleases: a paradigm shift in predictable, knowledge-based plant breeding. *Trends Biotechnol.* 31, 375–383. doi: 10.1016/j.tibtech.2013.03.004

Pyott, D. E., Sheehan, E., and Molnar, A. (2016). Engineering of CRISPR/ Cas9-mediated potyvirus resistance in transgene- free *Arabidopsis* plants. *Mol. Plant Pathol.* 4, 1–13. doi: 10.1111/mpp.12417

Roy, A., Zhai, Y., Ortiz, J., Neff, M., Mandal, B., Mukherjee, S. K., et al. (2019). Multiplexed editing of a begomovirus genome restricts escape mutant formation and disease development. *PLoS One* 14:e0223765. doi: 10.1371/journal.pone.0223765

Scheben, A., Wolter, F., Batley, J., Puchta, H., and Edwards, D. (2017). Towards CRISPR/Cas crops bringing together genomics and genome editing. *New Phytol.* 216, 682–698. doi: 10.1111/nph.14702

Schiml, S., and Puchta, H. (2016). Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. *Plant Methods* 12:8. doi: 10.1186/s13007-016-0103-0

Schmidt, S. M., Belisle, M., and Frommer, W. B. (2020). The evolving landscape around genome editing in agriculture. *EMBO Rep.* 21:e50680. doi: 10.15252/embr.202050680

Shao, X., Wu, S., Dou, T., Zhu, C., Huo, H., He, W., et al. (2020). Using CRISPR/Cas9 genome editing system to create *MaGA20ox2* gene-modified semi-dwarf banana. *Plant Biotechnol. J.* 18:17019. doi: 10.1111/pbi.13216

Shekhawat, U. K. S., Ganapathi, T. R., and Hadapad, A. B. (2012). Transgenic banana plants expressing small interfering RNAs targeted against viral replication initiation gene display high-level resistance to Banana bunchy top virus infection. *J. Gen. Virol.* 93, 1804–1813. doi: 10.1099/vir.0.041871-0

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., and Cigan, M. A. (2016). Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nat. Commun.* 7:13274. doi: 10.1038/ncomms13274

Tashkandi, M., Ali, M., Aljedaani, F., Shami, A., and Mahfouz, M. M. (2018). Engineering resistance against Tomato yellow leaf curl virus via the CRISPR/Cas9 system in tomato. *Plant Signal Behav.* 13:e1525996. doi: 10.1080/15592324.2018.1525996

Tripathi, J. N., Ntui, V. O., Ron, M., Muiruri, S. K., Britt, A., and Tripathi, L. (2019). CRISPR/Cas9 editing of endogenous Banana streak virus in the B genome of *Musa* spp. overcomes a major challenge in banana breeding. *Commun. Biol.* 2:46. doi: 10.1038/s42003-019-0288-7

Tripathi, J. N., Oduor, R. O., and Tripathi, L. (2015). A high-throughput regeneration and transformation platform for production of genetically modified banana. *Front. Plant Sci.* 6:1025. doi: 10.3389/fpls.2015.01025

Tripathi, L., Ntui, V. O., and Tripathi, J. N. (2019). Application of genetic engineering and genome editing for developing climate smart bananas. *Food Energy Secur.* 8:e00168. doi: 10.1002/fes3.168

Tripathi, L., Ntui, V. O., and Tripathi, J. N. (2020). CRISPR/Cas9 based genome editing of banana for disease resistance. *Curr. Opin. Plant Biol.* 56, 118–126. doi: 10.1016/j.pbi.2020.05.003

Tripathi, S., Patil, B. L., and Verma, R. (2016). "Viral diseases of banana and their management," in *Plant Viruses: Evolution and Management*, eds R. Gaur, N. Petrov, B. Patil, and M. Stoyanova (Singapore: Springer). doi: 10.1007/978-981-10-1406-2_17

Wang, B., Wang, R., Wang, D., Wu, J., Li, J., Wang, J., et al. (2019). Cas12avdet: a CRISPR/Cas12a-based platform for rapid and visual nucleic acid detection. *Anal. Chem.* 91, 12156–12161. doi: 10.1021/acs.analchem.9b01526

Wang, G., Zhao, N., Berkhout, B., and Das, A. T. (2016). CRISPR-Cas9 can inhibit HIV-1 replication but NHEJ repair facilitates virus escape. *Mol. Ther.* 24, 522–526. doi: 10.1038/mt.2016.24

Wang, J., and Quake, S. R. (2014). RNA-guided endonuclease provides a therapeutic strategy to cure latent herpesviridae infection. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13157–13162. doi: 10.1073/pnas.1410785111

Wang, Z., Pan, Q., Gendron, P., Zhu, W., Guo, F., Cen, S., et al. (2016). CRISPR/Cas9-derived mutations both inhibit HIV-1 replication and accelerate viral escape. *Cell Rep.* 15, 481–489. doi: 10.1016/j.celrep.2016.03.042

Wanitchakorn, R., Hafner, G. J., Harding, R. M., and Dale, J. L. (2000). Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *J. Gen. Virol.* 81, 299–306. doi: 10.1099/0022-1317-81-1-299

Wanitchakorn, R., Harding, R. M., and Dale, J. L. (1997). Banana bunchy topvirus DNA-3 encodes the viral coat protein. *Arch. Virol.* 142, 1673–1680. doi: 10.1007/s007050050188

Weinthal, D. M., and Gürel, F. (2016). "Plant genome editing and its applications in cereals," in *Genetic Engineering*, ed. F. Jamal (London: Intech Open).

White, M. K., Hu, W., and Khalili, K. (2016). Gene editing approaches against viral infections and strategy to prevent occurrence of viral escape. *PLoS Pathol.* 12:e1005953. doi: 10.1371/journal.ppat.1005953

Woo, J. W., Kim, J., Kwon, S. I., Corvalan, C., Cho, S. W., Kim, H., et al. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* 33, 1163–1165. doi: 10.1038/nbt.3389

Yin, K., Han, T., Xie, K., Zhao, J., Song, J., and Liu, Y. (2019). Engineer complete resistance to Cotton Leaf Curl Multan virus by the CRISPR/Cas9 system in *Nicotiana benthamiana*. *Phytopathol. Res.* 1, 1–9. doi: 10.1186/s42483-019-0017-7

Yin, K., and Qiu, J. -L. (2019). Genome editing for plant disease resistance: applications and perspectives. *Philos. Trans. R. Soc. B Biol. Sci.* 374:20180322. doi: 10.1098/rstb.2018.0322

Yuen, K. S., Chan, C. P., Wong, N. H., Ho, C. H., Ho, T. H., Lei, T., et al. (2015). CRISPR/Cas9-mediated genome editing of Epstein-Barr virus in human cells. *J. Gen. Virol.* 96, 626–636. doi: 10.1099/jgv.0.000012

Zaidi, S. S.-e.-A., Tashkandi, M., Mansoor, S., and Mahfouz, M. M. (2016). Engineering plant immunity: using CRISPR/Cas9 to generate virus resistance. *Front. Plant Sci.* 7:1673. doi: 10.3389/fpls.2016.01673

Zhan, X., Zhang, F., Zhong, Z., Chen, R., Wang, Y., Chang, L., et al. (2019). Generation of virus-resistant potato plants by RNA genome targeting. *Plant Biotechnol. J.* 17, 1814–1822. doi: 10.1111/pbi.13102

Zhang, T., Zhao, Y., Ye, J., Cao, X., Xu, C., Chen, B., et al. (2019). Establishing CRISPR/Cas13a immune system conferring RNA virus resistance in both dicot and monocot plants. *Plant Biotechnol. J.* 17, 1185–1187. doi: 10.1111/pbi.13095

Zhang, T., Zheng, Q., Yi, X., An, H., Zhao, Y., Ma, S., et al. (2018). Establishing RNA virus resistance in plants by harnessing CRISPR immune system. *Plant Biotechnol. J.* 16, 1415–1423. doi: 10.1111/pbi.12881

Zhao, Y., Yang, X., Zhou, G., and Zhang, T. (2020). Engineering plant virus resistance: from RNA silencing to genome editing strategies. *Plant Biotechnol. J.* 18, 328–336. doi: 10.1111/pbi.13278

Zhen, S., Hua, L., Takahashi, Y., Narita, S., Liu, Y. H., and Li, Y. (2014). In vitro and in vivo growth suppression of human papillomavirus 16-positive cervical cancer cells by CRISPR/Cas9. *Biochem. Biophys. Res. Commun.* 450, 1422–1426. doi: 10.1016/j.bbrc.2014.07.014

Zorrilla-Fontanesi, Y., Pauwels, L., Panis, B., Signorelli, S., Vanderschuren, H., and Swennen, R. (2020). Strategies to revise agrosystems and breeding to control fusarium wilt of banana. *Nature Food* 1, 599–604. doi: 10.1038/s43016-020-00155-y

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.

Copyright © 2021 Tripathi, Ntui, Tripathi and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 April 2021
doi: 10.3389/fmicb.2021.646273

Next-Generation Sequencing Identification and Characterization of MicroRNAs in Dwarfed Citrus Trees Infected With Citrus Dwarfing Viroid in High-Density Plantings

*Tyler Dang†, Irene Lavagi-Craddock†, Sohrab Bodaghi and Georgios Vidalakis**

Department of Microbiology and Plant Pathology, University of California, Riverside, Riverside, CA, United States

OPEN ACCESS

Edited by:
Ahmed Hadidi, Agricultural Research Service, United States Department of Agriculture, United States

Reviewed by:
Teruo Sano, Hirosaki University, Japan
Xuefeng Wang, Citrus Research Institute, Chinese Academy of Agricultural Sciences, China

***Correspondence:**
Georgios Vidalakis
georgios.vidalakis@ucr.edu; vidalg@ucr.edu

† 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: *05 January 2021*
Accepted: *06 April 2021*
Published: *30 April 2021*

Citation:
Dang T, Lavagi-Craddock I, Bodaghi S and Vidalakis G (2021) Next-Generation Sequencing Identification and Characterization of MicroRNAs in Dwarfed Citrus Trees Infected With Citrus Dwarfing Viroid in High-Density Plantings. Front. Microbiol. 12:646273. doi: 10.3389/fmicb.2021.646273

Citrus dwarfing viroid (CDVd) induces stunting on sweet orange trees [*Citrus sinensis* (L.) Osbeck], propagated on trifoliate orange rootstock [*Citrus trifoliata* (L.), syn. *Poncirus trifoliata* (L.) Raf.]. MicroRNAs (miRNAs) are a class of non-coding small RNAs (sRNAs) that play important roles in the regulation of tree gene expression. To identify miRNAs in dwarfed citrus trees, grown in high-density plantings, and their response to CDVd infection, sRNA next-generation sequencing was performed on CDVd-infected and non-infected controls. A total of 1,290 and 628 miRNAs were identified in stem and root tissues, respectively, and among those, 60 were conserved in each of these two tissue types. Three conserved miRNAs (csi-miR479, csi-miR171b, and csi-miR156) were significantly downregulated (adjusted p-value < 0.05) in the stems of CDVd-infected trees compared to the non-infected controls. The three stem downregulated miRNAs are known to be involved in various physiological and developmental processes some of which may be related to the characteristic dwarfed phenotype displayed by CDVd-infected *C. sinensis* on *C. trifoliata* rootstock field trees. Only one miRNA (csi-miR535) was significantly downregulated in CDVd-infected roots and it was predicted to target genes controlling a wide range of cellular functions. Reverse transcription quantitative polymerase chain reaction analysis performed on selected miRNA targets validated the negative correlation between the expression levels of these targets and their corresponding miRNAs in CDVd-infected trees. Our results indicate that CDVd-responsive plant miRNAs play a role in regulating important citrus growth and developmental processes that may participate in the cellular changes leading to the observed citrus dwarf phenotype.

Keywords: gene regulation, miRNA, sRNA, RNAi, vdsRNA, siRNA, gene silencing, plant antiviral response

INTRODUCTION

Small RNAs (sRNAs) can be divided into several categories, which include small-interfering (si)RNAs, *trans*-acting (ta)-siRNAs, microRNAs (miRNAs), natural-antisense siRNAs (nat-siRNAs), and Piwi-interacting RNAs (piwi-RNAs) (Borges and Martienssen, 2015; Czech et al., 2018; Zhu et al., 2018; Treiber et al., 2019). One of the major components of endogenous plant sRNAs are miRNAs. miRNAs are encoded by plant *MIR* genes and have independent transcriptional units with their own regulatory promoters. They form double stranded stem loop structures that are processed to produce single stranded transcripts, typically 21–24 nucleotides (nt) in length (Wang et al., 2019).

MicroRNAs have essential functions in plant development and are involved in regulating a myriad of plant processes such as leaf, root, stem, and floral organ morphogenesis and development, biosynthesis, metabolism, homeostasis, vegetative to reproductive growth transition, senescence, signal transduction, and response to biotic and abiotic stress. Upon expression and processing, plant miRNAs are incorporated into the activated RNA-induced silencing complex (RISC) to target RNAs which are complementary to the miRNA guide strand. Once the activated miRNA–RISC complex finds the complementary plant mRNA, it silences the target via RNA degradation or translational repression (Wang et al., 2019) (**Figure 1A**).

Viroid derived sRNAs (vdsRNA) are products of the RNA interference (RNAi) basal plant antiviral defense response (Navarro et al., 2012; Dadami et al., 2013, 2017; Eamens et al., 2014; Adkar-Purushothama et al., 2015, 2017; Reis et al., 2015; Taliansky et al., 2021). Viroids (246–401 nt), highly structured, autonomously replicating RNA plant pathogenic agents, and trigger RNAi during their replication due to the formation of double stranded intermediate RNAs (Flores et al., 2009; Dadami et al., 2017). Similarly to plant endogenous sRNAs, vdsRNAs are 21–22 and 24 nt in length and have been detected in plants infected by several different viroids (Navarro et al., 2009; Bolduc et al., 2010; Diermann et al., 2010; Tsushima et al., 2011). vdsRNAs play an important role in viroid-mediated biological and pathogenic activities by guiding the RISC-mediated cleavage of host RNAs (**Figure 1B**) (Wassenegger et al., 1994; Itaya et al., 2007; Navarro et al., 2012; Dadami et al., 2013, 2017; Eamens et al., 2014; Adkar-Purushothama et al., 2015, 2017; Reis et al., 2015; Flores et al., 2017; Ramesh et al., 2020). Viroid infection might cause symptoms through the action of vdsRNAs which alter the expression levels of plant miRNAs, which in turn affects the expression levels of the plant mRNA targets of those plant miRNAs. It was reported that potato spindle tuber viroid (PSTVd) infection of tomato affects host miRNA production (Diermann et al., 2010) and host mRNA production (Wang et al., 2011; Owens et al., 2012). It was also reported, citrus bark cracking viroid infection was shown to affect plant miRNA regulation of plant transcription factors regulating leaf, cone and root growth and development of hop plants (Mishra et al., 2016).

The *Citrus* genus (family *Rutaceae*), includes several cultivars of high economic value including oranges, mandarins, grapefruits and lemons (2018–2019 US citrus crop packinghouse-door equivalent $3.35 billion) (USDA-NASS, 2019). Citrus flavors and aromas are among the most recognizable and preferred worldwide. In addition, citrus fruits are a rich source of vitamins, antioxidants, minerals, and dietary fiber essential for overall nutritional wellbeing (Van Duyn and Pivonka, 2000; Yao et al., 2004). Citrus trees are produced by grafting a desired scion variety onto a suitable rootstock species that then are planted in commercial citrus orchards. Tree spacing in citrus orchards has varied depending upon the cultivated species and a variety of factors such as soil type, climatic conditions and available farming equipment (Platt, 1973; Tucker and Wheaton, 1978). The historical global trend of citrus orchard spacing has been toward higher tree densities to maintain yield on the reduced available agricultural land and to increase economic returns.

Citrus dwarfing viroid (CDVd) infection of navel orange trees [*Citrus sinensis* (L.) Osb.] propagated on trifoliate orange [*Citrus trifoliata* (L.), syn. *Poncirus trifoliata* (L.) Raf.] rootstock has been previously reported to reduce canopy volume by approximately 50% (**Figure 1B**) (Vidalakis et al., 2011) and we recently demonstrated that the observed reduction in tree size results from $a > 20\%$ reduction in the apical growth of individual shoots within the tree canopy (Lavagi-Craddock et al., 2020). Understanding the molecular mechanism of the CDVd-induced citrus tree size reduction, will be most valuable as it could provide information on how to systematically produce dwarf trees for high density plantings.

To date, very few studies exist of miRNAs in citrus and even fewer in navel orange trees (Lu et al., 2015; Ma et al., 2016; Liang et al., 2017; Xie et al., 2017; Huang et al., 2019) and to our knowledge, there are no published studies on citrus miRNAs in response to viroid infection of citrus field trees. To explore the effect of CDVd-infection on citrus miRNAs and gain insight into the symptom development mechanism leading to the dwarfed phenotype observed in field plantings, we analyzed the effect of CDVd infection using next-generation sequencing (NGS) approach. The increasing number of miRNAs deposited in the miRBase database (Kozomara and Griffiths-Jones, 2014; Kozomara et al., 2019) from a wide range of species (<200), including *C. sinensis,* enables the discovery of novel miRNAs and their responses to pathogen infection, which may account for the observed species specific reactions and symptom development. Many plant miRNAs are conserved (Axtell and Bartel, 2005) but some are species specific (Moxon et al., 2008) and expressed at lower levels, thus making NGS the ideal approach to discover them and study their expression profiles (Jagadeeswaran et al., 2010; Motameny et al., 2010). Indeed, miRNAs from different plant species such as maize (Zhang et al., 2009), potato (Zhang et al., 2013), peanut (Zhang et al., 2017), barley (Ferdous et al., 2017), soybean (Zhang et al., 2008), and hop (Mishra et al., 2016), have been identified using NGS approaches.

In this study, we analyzed sRNA libraries prepared from field grown CDVd-infected navel orange and non-infected control trees to characterize miRNAs in the *C. sinensis* (stems) and *C. trifoliata* (roots) genomes and their expression profile in response to CDVd infection. This work provides valuable information at the molecular level and establishes

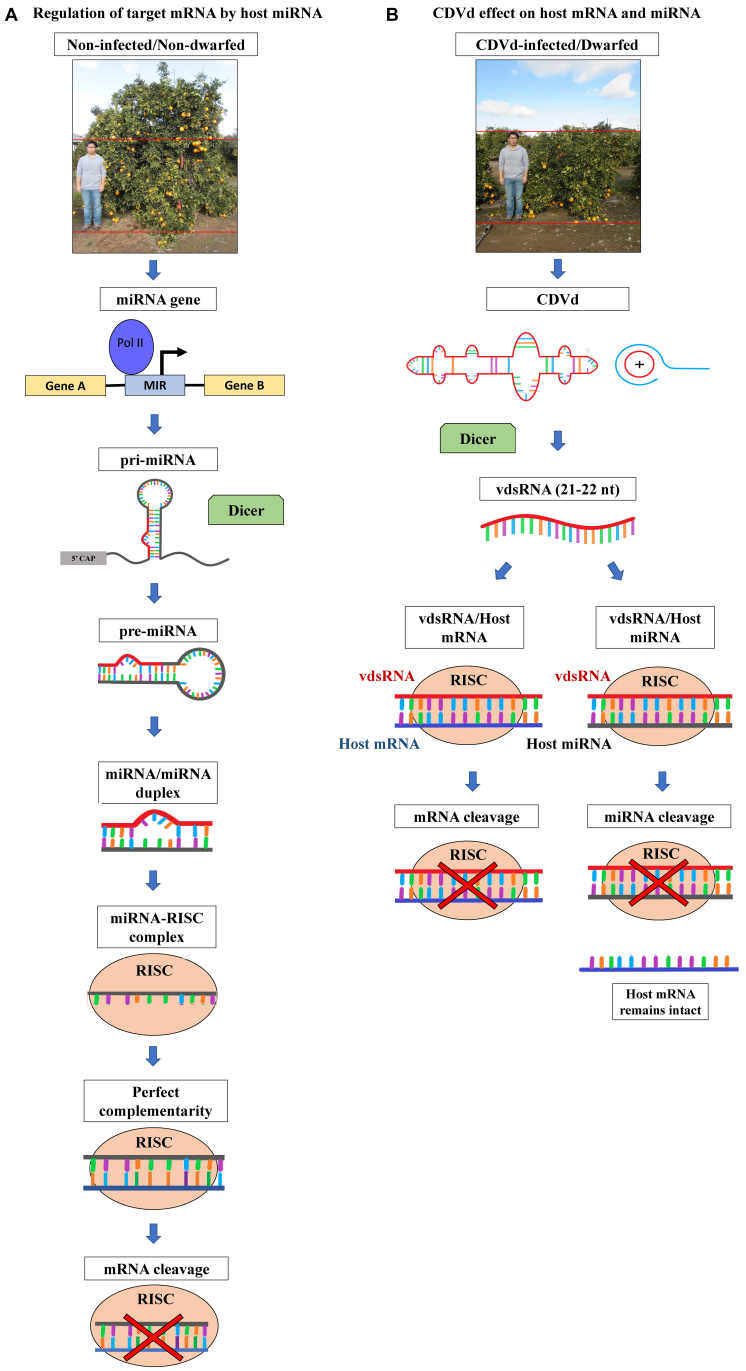


FIGURE 1 | Citrus phenotypes and schematic representation of host microRNA (miRNA)-based gene expression regulatory pathway and the hypothesized effects of citrus dwarfing viroid (CDVd) on the expression of host target genes. **(A)** Non-viroid-infected, non-dwarf phenotype (~3 m height) of navel orange [*Citrus sinensis* (L.) Osbeck] on trifoliate orange [*Citrus trifoliata* (L.) Raf.] rootstock, citrus tree and regulation of target messenger RNA (mRNA) by host miRNA. miRNA genes are transcribed by RNA polymerase II into primary miRNA (pri-miRNA) that folds into self-complementary stem-loop structures. The pri-RNA is then processed by the dicer-like (DCL) RNAse III endonucleases to form precursor miRNA (pre-miRNA), and the pre-miRNA is cleaved again to generate a miRNA/miRNA duplex. One strand of the duplex is degraded and the guide strand is incorporated into the RNA-inducing silencing effector complex (RISC). RISC, with RNAse H activity, targets RNA complementary to the miRNA guide strand. Once the complex finds the complementary plant mRNA, cleaves it, thus regulating its gene expression levels (for a review see Wang et al., 2019). **(B)** Viroid-infected, dwarf phenotype (~2 m height) of navel orange trees on trifoliate orange rootstock and the production of viroid derived small RNAs (vdsRNAs). The highly structured CDVd RNA molecule and double stranded intermediate RNAs produced during the viroid rolling-circle replication, are processed by DCL into 21–22 nt vdsRNAs. vdsRNAs are incorporated into RISC and can be guided to form either vdsRNA/host mRNA or vdsRNA/host miRNA complexes. As a result, CDVd vdsRNAs could alter the expression levels of plant mRNA (directly or indirectly via cleavage of plant miRNAs), therefore resulting in the observed citrus tree dwarf phenotype in the field (for a review see Dadami et al., 2017).

the foundational framework that is necessary to dissect the subcellular mechanisms responsible for the observed citrus dwarf phenotype in the field.

MATERIALS AND METHODS

Plant Material and RNA Isolation

Plant material (stems and roots) was collected in January 2016 from six 18-years-old "Parent Washington" navel [*C. sinensis* (L.) Osbeck] on "Rich 16-6" trifoliate orange [*C. trifoliata* (L.), syn. *Poncirus trifoliata* (L.) Raf.] rootstock infected (n = 3) and non-infected (n = 3) with CDVd, respectively. Trees were planted in an East–West running orchard located at the University of California (UC), Agriculture and Natural Resources, Lindcove Research and Extension Center (Exeter, CA, United States). CDVd-infected trees were planted at high density (3 × 6.7 m), whereas non-infected control trees were spaced at standard density (6.1 × 6.7 m).

Stem and root samples were processed in the field and immediately frozen in liquid nitrogen. For each tree, eight stem samples from around the canopy were collected. Leaves and petioles were removed, the stems were roughly chopped into approximately 0.5–1 cm pieces, placed into 50 ml conical tubes, and flash frozen. Root samples were collected from around the tree, at approximately 1 m away from the trunk and 20 cm deep, near the irrigation emitters, using a corer. The roots from eight core soil samples were washed thoroughly with water, gently blotted dry with paper towels, chopped into 0.5–1 cm pieces, placed into 50 ml conical tubes, and flash frozen. In between each sample collection and processing, cutting tools, and working surfaces were sanitized with 10% bleach solution (0.5–1% sodium hypochlorite) and rinsed with water and new sterile disposable plasticware and razor blades were used. Samples were transported into the Citrus Clonal Protection Program (CCPP), Citrus Diagnostic Therapy and Research Laboratory at the UC Riverside (Riverside, CA, United States) on dry ice and stored at −80°C until analysis.

Total RNA was isolated using the Invitrogen™ TRIzol™ (Thermo Fisher Scientific, Waltham, MA, United States) reagent. For each sample, 300 mg of frozen tissue were ground in liquid nitrogen with mortar and pestle. The ground material was transferred to a 5 ml Eppendorf tube and 3 ml of TRIzol™ reagent was added immediately. RNA extraction was performed according to the manufacturer's instructions. The eluted RNA was aliquoted into four 1.5 ml microcentrifuge tubes to prevent freezing-thawing cycles during downstream analysis. The RNA concentration and quality was assessed with a spectrophotometer and the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, United States) using the Plant RNA Nano assay (RIN values were between 7.9 and 8.6).

The presence or absence of CDVd in each sample was confirmed by reverse transcription quantitative polymerase chain reaction (RT-qPCR) using a CCPP developed and validated assay [F: 5′-AACTTACCTGTCGTCGTC-3′; R: 5′-CGTGTTTTACCCTGGAGG-3′; Probe (FAM): 5′-CTCCGCTAGTCGGAAAGACTCCGC-3′]. The assay was performed using the iTaq Universal Probes One-Step Kit (Bio-Rad, Hercules, CA, United States) in 20 μL reactions with 10 μL of iTaq universal probe reaction mix, 0.5 μL of reverse transcriptase, 0.6 μL of forward primer (300 nM final concentration), 1.2 μL reverse primer (600 nM final concentration), 0.4 μL of probe (200 nM final concentration), 1 μL of RNA template, and 6.3 μL of water. The RT-qPCR was performed in the Bio-Rad CFX-96 and the reaction conditions were as follows: 30 min at 50°C, 5 min at 95°C, followed by 45 cycles of 10 s 95°C, 30 s at 59°C.

Next-Generation Sequencing, sRNA Library Preparation and Sequencing Analysis to Identify Conserved and Novel miRNAs and Their Predicted Targets

The sRNA libraries were prepared using the Illumina TruSeq Small RNA Kit (San Diego, CA, United States) following the manufacturer's recommended protocol. The libraries were sequenced using an Illumina HiSeq™ 2500 instrument with single-end 50 bp reads (SeqMatic, Fremont, CA, United States). Raw reads were trimmed to remove low quality bases and adapters using cutadapt v. 1.15 (Martin, 2011) to generate clean sRNAs reads ranging from 18 to 28 nt in length.

The clean reads were then filtered for rRNA, tRNA, snRNA, snoRNA, repeat sequences, and other ncRNAs, using Rfam v.13.0 (Kalvari et al., 2018) with default parameters. The remaining reads were mapped to known miRNAs from the miRBase database (release 21, June 2014) to identify conserved miRNAs (Kozomara and Griffiths-Jones, 2011, 2014). The reads were further analyzed to predict potential novel miRNAs using miR-PREFer v. 0.24 using default parameters (Lei and Sun, 2014).

The conserved and novel miRNA sequences were analyzed against *C. sinensis* mRNA transcripts and *C. trifoliata* coding sequences (CDS) using psRNATarget v. 2.0 (Dai et al., 2018) to predict potential miRNA-mRNA interactions. DESeq2 v. 1.18 (Love et al., 2014) was used for the differential expression analyses of the miRNAs. The annotated mRNA targets were identified using the Blast2Go (Götz et al., 2008) tool within the OmicsBox software suite v. 1.4.11 (Cambridge, MA, United States). Figures were created using GraphPad Prism v. 9.0 (San Diego, CA, United States).

Expression Analysis of Citrus miRNAs and miRNA Target Genes Using RT-qPCR

To validate the expression levels of conserved and novel miRNAs, custom stem-loop RT-qPCR assays (catalog number: 4398987) were designed by Thermo Fisher Scientific based on the sequences provided in **Supplementary Table 1**. For the relative expression quantification, U6 spliceosomal RNA was used as an internal control gene to normalize the efficiency between the target and internal control using the comparative Cq method (Schmittgen and Livak, 2008; Kou et al., 2012). The assay was carried out based on the manufacturer's recommended protocol and all samples were standardized to the same concentration to ensure equal representation. The reverse transcription reactions were performed in a total volume of 15 μL with the TaqMan™

MicroRNA Reverse Transcription Kit (Thermo Fisher) which contained 0.15 μL of 100 mM dNTP, 1 μL of MultiScribe Reverse transcriptase, 1.5 μL of 10x RT Buffer, 0.19 μL of RNase Inhibitor, 4.16 μL of nuclease-free water, 5 μL of total RNA, and 3 μL of 5x RT primer. The reverse transcription reactions were performed with the ProFlex PCR System (Thermo Fisher) as follows: 16°C for 30 min, 42°C for 30 min, 85°C for 5 min, and 4°C hold. The endpoint qPCR was performed in triplicates, according to the MIQE guidelines (Bustin et al., 2009), on a QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher) with the TaqMan™ Fast Advanced Master Mix (Thermo Fisher) in a total of 20 μL reactions which included: 10 μL of master mix, 7.67 μL of nuclease-free water, 1 μL of TaqMan Small RNA Assay, 1.33 μL of the cDNA template. The endpoint PCR conditions were as follows: 50°C for 2 min, 95°C for 20 s, followed by 40 cycles of 95°C for 1 s, and 60°C for 20 s.

To verify the relative expression levels of the miRNA target genes, primers for the predicted target genes of miRNAs, were designed for RT-qPCR (**Supplementary Table 2**). Actin2 was used as an internal control gene to determine the relative abundance of the target mRNA expression levels by the comparative Cq method (Schmittgen and Livak, 2008; Mafra et al., 2012). Reverse transcription was performed using the Invitrogen™ SuperScript™ II Reverse Transcriptase (RT) (Carlsbad, CA, United States). The reaction was performed using the manufacturer's recommended protocol as follows: 1μL of olig (dT) (500 μg/mL), 1 μL of dNTP (10 mM), 2 μL of total RNA and 8 μL of nuclease-free water. The mixture was incubated for 5 min at 65°C and subsequently chilled on ice. The reaction was prepared with 4 μL of 5x First-Strand Buffer, 2 μL of 0.1M DTT, and 1 μL of RNaseOUT (40 units/μL) and then incubated for 2 min at 42°C. Finally, 1 μL of SuperScript™ II RT (200 units) was added and the reaction was incubated at 42°C for 50 min followed by 70°C for 15 min. Downstream qPCR was also performed in triplicates, according to the MIQE guidelines, using the iTaq Universal SYBR Supermix (Bio-Rad): 10 μL of iTaq Universal SYBR Supermix, 1 μL of cDNA, 0.6 μL of each forward and reverse primers and 7.8 μL of nuclease-free water. The qPCR was performed on the Bio-Rad CFX-96 with the following conditions: 95°C for 1 min, followed by 40 cycles of 95°C for 10 s, and 60°C for 15 s.

RESULTS

Next-Generation Sequencing and Characterization of Potential Citrus miRNAs

To characterize citrus miRNAs and their expression profile in response to CDVd infection, we prepared and analyzed two sRNA libraries from stems and root samples of CDVd-infected and non-infected controls of navel orange citrus trees on trifoliate orange rootstock (**Table 1**).

From the non-infected trees, 8.1% of the stem and 6.1% of the root were classified as miRNAs. Similarly, for the CDVd-infected stems and roots, 7.2 and 5.8% of the reads, respectively, were classified as miRNAs. The unique unannotated sequences in both the non-infected and CDVd-infected stems represented at least 90% of the total reads while for the roots they represented over 82% (**Table 1**). The total unique miRNA reads for both non-infected and CDVd-infected stems represented 0.05% of the reads, while both non-infected and CDVd-infected roots represented 0.0006% of the reads (**Table 1**).

The most common size among the total mapped miRNAs sequences ranged between 20 and 24 nt in length, with 21-nt being the predominant miRNA class across different treatments and tissue types. This is consistent with plant antiviral RNAi responses and DCL-mediated processing of dsRNA producing 21 nt siRNAs.

Identification of Conserved miRNAs and Their Expression Profiles

The miRNA sequencing from non-infected control and CDVd-infected stems and roots identified 60 unique conserved miRNAs that ranged from 20 to 24 nt (**Supplementary Tables 3**, **4**). Based on differential expression analysis, four conserved miRNAs (three in the stems and one in the roots) were found to be significantly altered in response to CDVd-infection (*P*-value and adjusted *P*-value < 0.05) (**Table 2**). Our results indicated that different members of the three miRNA families of interest had different expression levels between the non-infected and the CDVd-infected trees. The conserved miRNA families in the stems included csi-miR156, csi-miR171b, and csi-miR479, while csi-miR535 was the only conserved miRNA found in the roots. The conserved stem miRNAs were moderately more abundant compared to the conserved root miRNAs (**Table 2**). All four conserved miRNAs had higher expression levels in the non-infected control than the CDVd-infected trees (**Table 2**).

Five miRNA families present in both stem and root tissues were identified: miR166, miR171b, miR399, miR477, and miR482. In stems, the highest represented miRNA families were miR166, and miR399, with five members each, followed by miR171b with four members, miR396, miR477, and miR482 with three members and the remaining 40 miRNAs were represented by a single member (**Supplementary Table 3**). In the roots, two miRNA families (miR166 and miR399) were represented by five members, five miRNA families (miR167, miR172, miR396, miR477, and miR482) were represented by three members, two miRNA families (miR530 and miR171b) contained two members, and the remaining 40 root miRNA families were represented by a single member (**Supplementary Table 4**).

Stem-loop RT-qPCR analysis was performed on the four conserved miRNAs in root and stem tissues (csi-miR479, csi-miR156, csi-miR171b, and csi-miR535) from non-infected and CDVd-infected trees to determine their relative abundance. The expression levels of the four conserved miRNAs were significantly altered as a result of CDVd infection. In the stems, csi-miR479's expression decreased 3.55 fold and csi-miR171b had a fold decrease of 2.24, while csi-miR156 had the smallest negative fold change (0.11) (**Figure 2**). In the roots, csi-miR535′s expression decreased 1.12 fold (**Figure 2**). The results obtained from the RT-qPCR analysis (**Figure 2**) were consistent with the NGS read

TABLE 1 | Statistical summary of small RNA (sRNA) sequences from non-infected and citrus dwarfing viroid (CDVd)-infected libraries from stem (*Citrus sinensis*) and root (*Citrus trifoliata*) tissues.

	Non-infected stem		CDVd-infected stem	
	Reads	**Unique sequences**	**Reads**	**Unique sequences**
Raw reads	16,008,944	N/A	13,764,218	N/A
Clean reads (18–28 nt sRNA)	6,742,931 (100%)	1,453,586 (100%)	5,733,421 (100%)	1,214,014 (100%)
miRNA	545,243 (8.1%)	683 (0.05%)	412,526 (7.2%)	607 (0.05%)
rRNA, tRNA, snRNA, and snoRNA	3,220,307 (47.8%)	115,180 (7.9%)	2,880,237 (50.2%)	109,817 (9.05%)
Unannotated	2,977,381 (44.2%)	1,336,138 (91.9%)	2,440,659 (42.6%)	1,102,588 (90.8%)
	Non-infected root		**CDVd-infected root**	
	Reads	**Unique sequences**	**Reads**	**Unique sequences**
Raw reads	6,524,898	NA	5,864,614	NA
Clean reads (18–28 nt sRNA)	2,030,419 (100%)	515,978 (100%)	1,873,309 (100%)	468,357 (100%)
miRNA	125,224 (6.1%)	313 (0.00061%)	108,262 (5.8%)	315 (0.0007%)
rRNA, tRNA, snRNA, and snoRNA	1,027,218 (50.5%)	87,435 (16.9%)	984,550 (52.5%)	81,988 (17.5%)
Unannotated	877,976 (43.2%)	427,990 (82.9%)	780,496 (41.6%)	385,868 (82.4%)

frequencies (**Table 2**) indicating a strong correlation between the RT-qPCR analysis and read frequencies obtained through small sRNA sequencing.

Identification of Novel miRNAs and Their Expression Profiles

The lengths of the predicted novel miRNAs from stems and roots ranged between 19 and 24 nt. A total of 646 stem and 108 root novel miRNAs were identified (**Supplementary Tables 5**, **6**). No novel root miRNAs had significant differential expression levels. On the other hand, three novel stem miRNAs (csi-miRNA-75, csi-miRNA-114, and csi-miRNA-435) had significantly different expression levels in response to CDVd infection (*P*-value and adjusted *P*-value < 0.05). All three novel stem miRNAs had higher expression levels in the non-infected trees than in the CDVd-infected trees (**Table 3**).

Stem-loop RT-qPCR was performed to confirm the NGS read frequency results of the identified novel miRNAs in response to CDVd-infection. csi-miRNA-114 showed the largest expression fold change (−9.23), while csi-miRNA-75 and csi-miRNA-435 showed similar fold changes (−1.44 and −1.22, respectively) (**Figure 3**). These results also support the reliability of RT-qPCR (**Figure 3**) and NGS read frequencies (**Table 3**) as measures of the expression levels of miRNAs.

Citrus miRNA-Target Prediction and Functional Analysis

To understand the function of the identified citrus miRNAs, host target genes were analyzed using the psRNATarget program by cross referencing the results against the *C. sinensis* genome for the stems (ref: GCF_000317415.1) and the *C. trifoliata* CDS for the roots (Kawahara et al., 2020). For both conserved and novel stem miRNAs, 83.1% of the miRNA targets were predicted to be regulated by cleavage and 16.9% by translational inhibition (**Supplementary Tables 7**, **8**). Similarly, in the roots, 86.4% of the miRNA targets were predicted to be regulated by cleavage and 13.6% by translational inhibition (**Supplementary Tables 9**, **10**).

Based on the extent of sequence complementarity between miRNAs and their targets, a total of 5,542 potential targets were predicted for the conserved and novel stem and root miRNAs (conserved: stem 63 and root 64; novel: stem 647 and root 109). Of the 5,542 potential miRNA target genes, 494 and 3,926 were targets of the conserved and novel stem miRNAs while 495 and 627 were targets of the conserved and novel root miRNAs, respectively.

The conserved stem and root miRNAs (miR479, mi171b, miR156, and miR535) analyzed in this study were associated with 10 different groups of target genes while the novel stem miRNAs (csi-miRNA-75, csi-miRNA-114, and csi-miR435) associated with three different target genes (see section "Discussion") (**Supplementary Table 11**).

Clusters of orthologous groups (COG) functional classification of the targets of conserved and novel miRNAs revealed that the highest proportion of the genes were associated with (i) the nucleus (21% conserved and 11% novel); (ii) the integral component of membrane (14% conserved and 22% novel); and (iii) ATP binding (11% conserved and 10% novel) (**Figures 4A,B**).

Other miRNA targets shared by the conserved and novel miRNAs include (i) ADP binding (9% conserved and 2% novel); (ii) cytoplasm (4% conserved and 6% novel); (iii) DNA-binding transcription factor activity (3% conserved and 1% novel); (vi) oxidation-reduction processes (7% conserved and 4% novel); (v) plasma membrane (6% conserved and 1% novel); (vi) protein phosphorylation (2% conserved and 4% novel); and (vii) regulation of transcription (14% conserved and 3% novel) (**Figures 4A,B**).

Our data were further annotated based on ontological definitions of the gene ontology (GO) terms, which categorized the predicted targets of the conserved miRNAs differentially expressed in response to CDVd infection into various biological,

TABLE 2 | Subsets of the conserved microRNAs (miRNAs) and their recovery profile in response to citrus dwarfing viroid (CDVd)-infection in stem (*Citrus sinensis*) and root (*C. trifoliata*) tissues.

Family	miRNA name	Sequence (5′–3′)	Length (nt)	Tissue type	Normalized value		Log 2 fold change	*P*-value	Adjusted *P*-value	Significance label
					Non-infected	CDVd-infected				
MIR156	csi-miR156	UGACAGAAGAGAGUGAGCAC	20	Stem	944.28	532.04	−0.828	0.001084	0.0412	**
MIR171	csi-miR171b	CGAGCCGAAUCAAUAUCACUC	21	Stem	433.35	264.95	−0.711	0.000740	0.0338	**
MIR171	csi-miR479	UGUGAUAUUGGUUCGGCUCAUC	22	Stem	433.35	264.95	−0.711	0.000740	0.0338	**
MIR166	csi-miR166a	UCGGACCAGGCUUCAUUCCCCC	22	Stem	225,444.12	190,251.69	−0.24	0.29	0.614	
MIR166	csi-miR166b	UCGGACCAGGCUUCAUUCCCGU	22	Stem	190,840.39	211,388.61	0.15	0.25	0.610	
MIR166	csi-miR166c	UCGGACCAGGCUUCAUUCCC	20	Stem	248,503.49	210,279.64	−0.24	0.27	0.613	
MIR166	csi-miR166d	UCGGACCAGGCUUCAUUCCCU	21	Stem	10,291.47	9,164.90	−0.17	0.35	0.645	
MIR166	csi-miR166e	UCGGACCAGGCUUCAUUCCCC	21	Stem	225,437.68	190,245.62	−0.24	0.29	0.614	
MIR396	csi-miR396a	UUCCACAGCUUUCUUGAACUG	21	Stem	4,439.016609	6,389.073731	0.53	0.01	0.10	
MIR396	csi-miR396b	UUCCACAGCUUUCUUGAACUG	21	Stem	4,625.81381	6,577.736377	0.51	0.01	0.10	
MIR396	csi-miR396c	UUCAAGAAAUCUGUGGGAAG	20	Stem	3,340.320865	2,152.155423	−0.63	0.02	0.16	
MIR399	csi-miR399a	UGCCAAAGGAGAUUUGCCCGG	21	Stem	0.82	2.04	1.47	0.22	NA	
MIR399	csi-miR399b	UGCCAAAGGAGAGUUGCCCUA	21	Stem	24.93	41.24	0.74	0.08	0.379	
MIR399	csi-miR399c	UGCCAAAGGAGAAUUGCCCUG	21	Stem	2.28	4.29	0.99	0.25	NA	
MIR399	csi-miR399d	UGCCAAAGGAGAGUUGCCCUG	21	Stem	90.63	113.30	0.33	0.32	0.644	
MIR399	csi-miR399e	UGCCAAAGGAGAAUUGCCCUG	21	Stem	2.28	3.93	0.87	0.32	NA	
MIR477	csi-miR477a	ACCUCCCUCGAAGGCUUCCAA	21	Stem	63.15	53.52	−0.23	0.57	0.750	
MIR477	csi-miR477b	CUCUCCCUCAAGGGCUUCUCU	21	Stem	1,367.46	1,020.24	−0.42	0.05	0.289	
MIR477	csi-miR477c	UCCCUCGAAGGCUUCCAAUAUA	22	Stem	63.15	53.52	−0.23	0.57	0.750	
MIR482	csi-miR482a	UCUUCCCUAUGCCUCCCAUUCC	22	Stem	1,029.05	915.31	−0.17	0.35	0.645	
MIR482	csi-miR482b	UCUUGCCCACCCCUCCCAUUCC	22	Stem	632.93	519.99	−0.28	0.06	0.316	
MIR482	csi-miR482c	UUCCCUAGUCCCCCUAUUCCUA	22	Stem	207.25	288.14	0.48	0.06	0.316	
MIR535	csi-miR535	UGACAAUGAGAGAGAGCACAC	21	Root	63.97	24.38	−1.42	0.00	0.01	**
MIR159	csi-miR159	UUUGGAUUGAAGGGAGCUCUA	21	Root	1,334.66	1,369.76	0.04	0.88	1.00	
MIR479	csi-miR479	UGUGAUAUUGGUUCGGCUCAUC	22	Root	57.62	30.32	−0.96	0.01	0.48	
MIR319	csi-miR319	UUUGGACUGAAGGGAGCUCCU	21	Root	59.43	74.55	0.30	0.40	1.00	
MIR166	csi-miR166a	UCGGACCAGGCUUCAUUCCCCC	22	Root	40,255.18	37,986.77	−0.08	0.76	0.997	
MIR166	csi-miR166b	UCGGACCAGGCUUCAUUCCCGU	22	Root	56,293.03	48,592.72	−0.21	0.37	0.997	
MIR166	csi-miR166c	UCGGACCAGGCUUCAUUCCC	20	Root	41,752.00	39,585.50	−0.08	0.78	0.997	
MIR166	csi-miR166d	UCGGACCAGGCUUCAUUCCCU	21	Root	1,451.02	1,455.03	0.01	0.98	0.997	
MIR166	csi-miR166e	UCGGACCAGGCUUCAUUCCCC	21	Root	40,249.81	37,980.07	−0.08	0.76	0.997	
MIR167	csi-miR167a	UGAAGCUGCCAGCAUGAUCUG	21	Root	196.25	146.19	−0.41	0.19	1.00	
MIR167	csi-miR167b	UGAAGCUGCCAGCAUGAUCUU	21	Root	247.06	243.45	−0.02	0.96	1.00	
MIR167	csi-miR167c	UGAAGCUGCCAGCAUGAUCUG	21	Root	35.56	32.61	−0.07	0.82	1.00	

(Continued)

TABLE 2 | Continued

Family	miRNA name	Sequence (5′–3′)	Length (nt)	Tissue type	Normalized value		Log 2 fold change	*P*-value	Adjusted *P*-value	Significance label
					Non-infected	CDVd-infected				
MIR171	csi-miR171a	UUGAGCCGCGCCAAUAUCAC	20	Root	0.27	0.28	0.29	0.90	1.00	
MIR171	csi-miR171b	CGAGCCGAAUCAAUAUCACUC	21	Root	57.62	30.32	−0.96	0.01	0.48	
MIR172	csi-miR172a	AGAAUCUUGAUGAUGCUGCA	20	Root	72.88	57.79	−0.32	0.40	1.00	
MIR172	csi-miR172b	AGAAUCUUGAUGAUGCGGCAA	21	Root	0.34	0.38	0.30	0.89	1.00	
MIR172	csi-miR172c	UGGAAUCUUGAUGAUGCUGCAG	22	Root	50.84	45.38	−0.15	0.73	1.00	
MIR399	csi-miR399a	UGCCAAAGGAGAUUUGCCCGG	21	Root	0.14	0.38	−0.01	1.00	0.997	
MIR399	csi-miR399b	UGCCAAAGGAGAGUUGCCCUA	21	Root	20.99	28.18	0.37	0.46	0.997	
MIR399	csi-miR399c	UGCCAAAGGAGAAUUGCCCUG	21	Root	0.20	1.60	1.93	0.17	0.997	
MIR399	csi-miR399d	UGCCAAAGGAGAGUUGCCCUG	21	Root	37.82	42.06	0.14	0.87	0.997	
MIR399	csi-miR399e	UGCCAAAGGAGAAUUGCCCUG	21	Root	0.20	1.60	1.93	0.17	0.997	
MIR477	csi-miR477a	ACCUCCCUCGAAGGCUUCCAA	21	Root	42.54	43.27	0.01	0.99	0.997	
MIR477	csi-miR477b	CUCUCCCUCAAGGGCUUCUCU	21	Root	38.37	27.60	−0.52	0.14	0.997	
MIR477	csi-miR477c	UCCCUCGAAGGCUUCCAAUAUA	22	Root	42.54	43.27	0.01	0.99	0.997	
MIR482	csi-miR482a	UCUUCCCUAUGCCUCCCAUUCC	22	Root	71.80	65.94	−0.10	0.74	0.997	
MIR482	csi-miR482b	UCUUGCCCACCCCUCCCAUUCC	22	Root	89.33	89.17	−0.03	0.92	0.997	
MIR482	csi-miR482c	UUCCCUAGUCCCCCUAUUCCUA	22	Root	27.56	29.59	0.14	0.70	0.997	
MIR530	csi-miR530a	UGCAUUUGCACCUGCACCUUG	21	Root	0.810	0.555	−0.893	0.570	0.997	
MIR530	csi-miR530b	UGCAUUUGCACCUGCAUCUUG	21	Root	0.677	1.047	0.508	0.703	0.997	

*miRNAs with statistically significant values are noted with **. The complete miRNA dataset can be found in* ***Supplementary Tables 3–6****.*

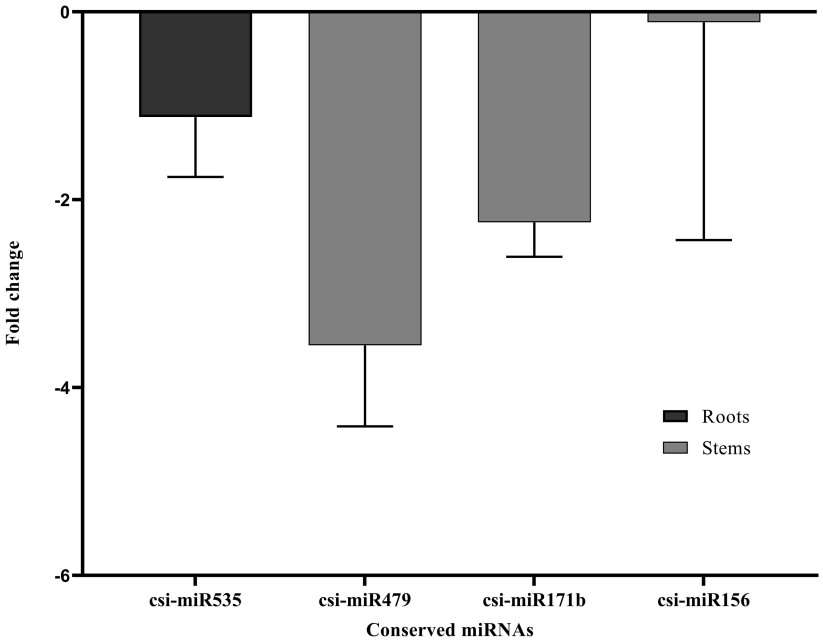


FIGURE 2 | Differential expression analysis of four identified conserved miRNAs in response to CDVd infection. The relative abundance of each analyzed miRNA in CDVd-infected and non-infected trees was determined using the comparative Cq method by normalization to the U6 spliceosomal RNA. Conserved miRNAs with a significant change (*P*-value and adjusted *P*-value < 0.05) were considered to be differentially expressed. The bar graph shows the $\log_2$ fold change of expression levels of the miRNAs in CDVd-infected samples relative to non-infected samples in stem and root tissues.

molecular and cellular processes (**Figure 5A**). Under the biological process, the predicted targets of conserved miRNA responsive to CDVd infection were subcategorized to (i) metabolic process; (ii) cellular process; (iii) biological regulation; and (iv) regulation of biological processes. The number of sequences associated with these four biological process subcategories had similar values in the stems and roots with the exception of the metabolic process that was higher in the roots (**Figure 5A**). For the molecular process, the majority of the predicted target genes of conserved miRNAs responsive to CDVd infection in the roots were subcategorized to catalytic activity while the targets in the stems were mostly subcategorized to binding. Targets belonging to the cellular components category were subcategorized to (i) cellular anatomical entity; and (ii) intracellular subcategories (**Figure 5A**) (**Supplementary Table 11**).

The targets of the predicted novel miRNAs displaying significant differential expression in response to CDVd infection (csi-miRNA-75, csi-miRNA-114, and csi-miRNA-435) were categorized to (i) biological process; (ii) molecular process; and (iii) cellular components. The cellular components subcategories (i) cellular anatomical entity; and (ii) intracellular contained most of the sequences (**Figure 5B**) (**Supplementary Table 11**).

Expression Profiles and Experimental Validation of miRNA Target Transcripts

The expression levels of eight predicted mRNA targets of the conserved CDVd-responsive miRNAs was determined via RT-qPCR. The results indicate that the expression of the miRNA target genes correlates negatively with the expression of their corresponding miRNA (**Figure 6**), thus confirming the relationship between CDVd-infection and altered expression levels of specific miRNA targets. Targets of miRNAs belonging to the same miRNA family showed variable results. For example, orange1.1g011651m (**Figure 6**, bar #5), orange1.1g032310m (**Figure 6**, bar #6), and orange1.1g008776m (**Figure 6**, bar #7), which are all targets of different members of the csi-miR156 family, were not uniform and showed fold-change differences.

DISCUSSION

Citrus production exceeded 157.9 million tons in over 9.8 million hectares worldwide for 2019[1], while in California alone the citrus industry is valued at $3.4 billion dollars with an estimated total economic impact of $7.1 billion (Babcock, 2018). Global decrease in farmland availability, increasing land, water and labor costs, and the continued spread of the deadly Huanglongbing (HLB) disease of citrus, make it imperative to develop tools that allow for high-density citrus plantings for maximization of yields and economic returns per land surface unit. In addition, these factors have forced the citrus industry toward the implementation of novel cultivation practices that would allow for mechanized citrus production under protective structures (Gottwald, 2010; Vidalakis et al., 2011; Lambin, 2012; Verburg et al., 2013; da Graça et al., 2016).

[1]http://www.fao.org/faostat/en/#data/QC

TABLE 3 | Subsets of the novel microRNAs (miRNAs) and their recovery profile in response to citrus dwarfing viroid (CDVd)-infection in stem (*Citrus sinensis*) and root (*C. trifoliata*) tissues.

miRNA name	Sequence (5′–3′)	Length (nt)	Tissue type	Normalized value		Log 2 fold change	*P*-value	Adjusted *P*-value	Significance label
				Non-infected	CDVd-infected				
csi-miRNA-75	GUGACAGAAGAGAGUGAGCAC	21	Stem	82.35	35.66	−1.219	0.000074	0.0084	**
csi-miRNA-114	UUGGGCUCUCUUCCUCUCAUG	21	Stem	30.71	6.73	−2.247	0.000023	0.0052	**
csi-miRNA-435	GUCCCUCUCACAGCUACAGUACCC	24	Stem	45.62	23.46	−0.984	0.000380	0.0289	**
csi-miRNA-02	AAAAGGAGGACUAAGUUAAAAGCA	24	Stem	77.68	59.82	−0.39	0.08	0.38	
csi-miRNA-23	AUAUUGGAGUGUUUGACCAGU	21	Stem	60.26	46.28	−0.38	0.14	0.53	
csi-miRNA-87	ACAAGAGUUUGUGACUGUAUCAUU	24	Stem	8.55	10.60	0.20	0.67	1.00	
csi-miRNA-20	GUGACAGAAGAGAGUGAGCAC	21	Root	48.73	39.25	−0.29	0.44	1.00	
csi-miRNA-57	AUUCCUCAUUGUUUGGUCAACAGC	24	Root	13.49	9.75	−0.42	0.41	1.00	
csi-miRNA-63	UUCCAAAGGGAUCGCAUUGAUC	22	Root	59.08	53.41	−0.17	0.59	1.00	
csi-miRNA-77	AGGCAGUCUCCUUGGCUAAG	20	Root	9.54	5.30	−0.90	0.08	1.00	
csi-miRNA-91	AAGCACGAGAGAAAGACGAGAGAA	24	Root	5.21	5.17	−0.05	0.94	1.00	
csi-miRNA-100	AUUUCGGUAACUAAUAGGAUAUAC	24	Root	3.48	3.24	−0.07	0.93	1.00	

*miRNAs with statistically significant values are noted with **. The complete miRNA dataset can be found in* ***Supplementary Tables 3–6****.*

The idea of using "graft-transmissible dwarfing agents" for tree size reduction has been continuously investigated in citrus since originally proposed by Cohen (1968) and Mendel (1968) and countries such as Australia and Israel have explored the application of such technology in commercial settings (Broadbent et al., 1992; Bar-Joseph, 1993; Hutton et al., 2000; Semancik, 2003). The observation that CDVd significantly reduced *C. sinensis* canopy volume on *C. trifoliata* rootstock (Semancik et al., 1997; Vidalakis et al., 2011); by reducing vegetative growth (Lavagi-Craddock et al., 2020) indicated that CDVd may be used as a possible tool for high-density plantings of citrus, and provided key information on the possible biological mechanism through which CDVd affects specific rootstock-scion combinations to reduce tree canopy volume (Vidalakis et al., 2011). Furthermore, understanding the detailed molecular regulatory mechanisms that lead to a reduction in tree canopy volume in response to CDVd infection would provide the necessary knowledge to produce reduced-size citrus trees without the need of a -graft-transmissible viroid agent.

Small RNAs play an essential regulatory role in cellular and plant development functions including antiviral host responses and potentially viroid pathogenesis (Borges and Martienssen, 2015; Dadami et al., 2017; Flores et al., 2017; Czech et al., 2018; Zhu et al., 2018; Treiber et al., 2019; Wang et al., 2019; Ramesh et al., 2020). Two models have been proposed to explain the involvement of RNAi in the pathogenic process induced by viroid infections and both involve vdsRNAs. In the first model, vdsRNAs might act as miRNAs, downregulating the expression of physiologically important host genes, thus inducing disease associated symptoms. vdsRNAs are expected to contain significant identity to a region of the host genome for this model to work and resistance of viroids to RNAi is a feature of the viroid genome (Wang et al., 2004). According to the second model, disease symptoms caused by the nucleus replicating pospiviroids might result from the incorporation of viroid replication intermediates into the trans-acting small interfering RNA (ta-siRNA) biogenesis pathway. The nucleolus is a ta-siRNA free zone, and mature viroid forms produced in the nucleolus are resistant to degradation. In contrast, (vd)ta-siRNA produced in the nucleus from replication intermediates can then translocate to the cytoplasm where they guide the cleavage of target host mRNA leading to observed symptoms (Gómez et al., 2009). Both models involve viroid secondary structures as a key element that can therefore be interpreted as the evolutionary compromise between the need to interact with host factors and the necessity to survive RNAi. Regardless of whether vdsRNAs are produced according to the first or second model, it is also important to point out that rather than acting directly on host mRNA, vdsRNAs may affect host mRNA targets genes via host miRNAs as previously described (Mishra et al., 2016; Dadami et al., 2017).

Plant disease resistance gene families are typically very large with thousands of members and are commonly considered the putative targets of sRNAs (Chen et al., 2010), thus making the study of sRNAs in response to viroid infection a valid approach to investigate the biological mechanisms associated with symptoms.

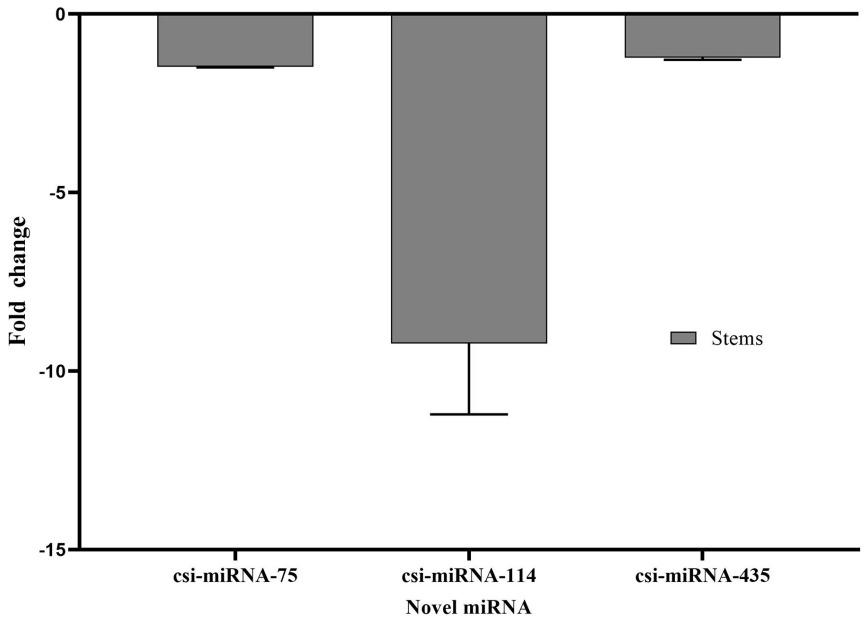


FIGURE 3 | Differential expression analysis of three predicted novel miRNAs in response to CDVd infection. The relative abundance of each analyzed miRNA in CDVd-infected compared to non-infected trees was determined using the comparative Cq method by normalization to the U6 spliceosomal RNA (U6). Novel miRNAs with significant fold changes (*P*-value and adjusted *P*-value < 0.05) were considered to be differentially expressed. The bar graph shows $\log_2$ fold changes in expression levels of miRNAs in CDVd-infected samples relative to non-infected samples from stems.

The systematic profiling of sRNAs in CDVd-infected trees, using NGS technologies, was the next logical step to gain insight into the function and regulatory mechanisms of miRNAs through which CDVd may reduce tree canopy size. In this study, we identified conserved and novel miRNAs in citrus and their response to CDVd infection. Consistently with the distribution patterns of sRNAs in other plant species, most sRNAs from both the CDVd-infected and non-infected libraries were found in the 21 and 24 nt classes (Jia et al., 2014; Gao et al., 2015; Mishra et al., 2016; Farooq et al., 2017; Zhang et al., 2018). CDVd-infected stems produced higher frequencies of the 21-nt class than their non-infected counterparts, indicating a CDVd induction of the 21-nt class since the other sRNA classes remained at comparable levels with the non-infected libraries. The increased abundance of the 21-nt class of sRNAs in response to viroid infection observed here is in agreement with previous reports for viral (viroid and virus) infections (Minoia et al., 2014; Zavallo et al., 2015).

All identified differentially expressed conserved and novel miRNAs, in this study, displayed overall reduced expression levels in response to CDVd-infection (**Figures 2**, **3**). Several evolutionary deeply-conserved miRNAs have been shown to retain homologous targets across plant phyla (Axtell et al., 2007) and these include miR156 (stem), miR535 (roots), and miR171b (stem), which represent three out of the four conserved miRNA with differential expression levels in response to CDVd infection identified in this study. In agreement with previous studies, miR156 was shown here to direct the cleavage of squamosa-promoter binding-like protein (SBP) box genes (orange1.1g011651m; orange1.1g032310m; and orange1.1g008776m) (Cardon et al., 1999; Rhoades et al., 2002; Wu and Poethig, 2006; Xie et al., 2006; Gandikota et al., 2007; Riese et al., 2007) (**Supplementary Table 7** and **Figure 6**). Members of this transcription factor family are known to play important roles in flower and fruit development, plant architecture, and in the transitions from juvenile to adult stages and to flowering (Chen et al., 2010; Yu et al., 2012). Even though miR535 is also known to target squamosa promoter-binding-like protein 3 (Shi et al., 2017; Zhou et al., 2020), our study identified rhomboid-like protein 1 as the target of miR535 in the roots (P_trifoliata_00066_mRNA_51.1). In *Arabidopsis*, a rhomboid-like protein was identified, providing evidence for the existence of regulated intramembrane proteolysis (RIP), a fundamental mechanism for controlling a wide range of cellular functions, in plants (Kanaoka et al., 2005). miR171b directs the cleavage of GRAS domain transcription factor genes (Ma et al., 2014). However, in our study, we found that miR171b's target, a probable glutathione-S-transferase, was altered in response to CDVd infection (orange1.1g033674m). Glutathione-S-transferases are ubiquitous and multifunctional enzymes encoded by large gene families that can be highly induced by biotic stress including bacterial, fungal, and viral infection (Gullner et al., 2018). We found that the less conserved miR479 cleaves the UDP-glucose flavonoid glucosyl-transferase (orange1.1g033614m) (Vogt and Jones, 2000; Offen et al., 2006) which is involved in the process of conjugating hormones, stabilizing secondary metabolites, solubility, transport, and regulating bioavailability of compounds for other metabolic process in *Arabidopsis thaliana*and DEAD/DEAH box helicases which play an important in regulatory events such as organ maturation and cellular growth and differentiation (orange1.1g028826m and

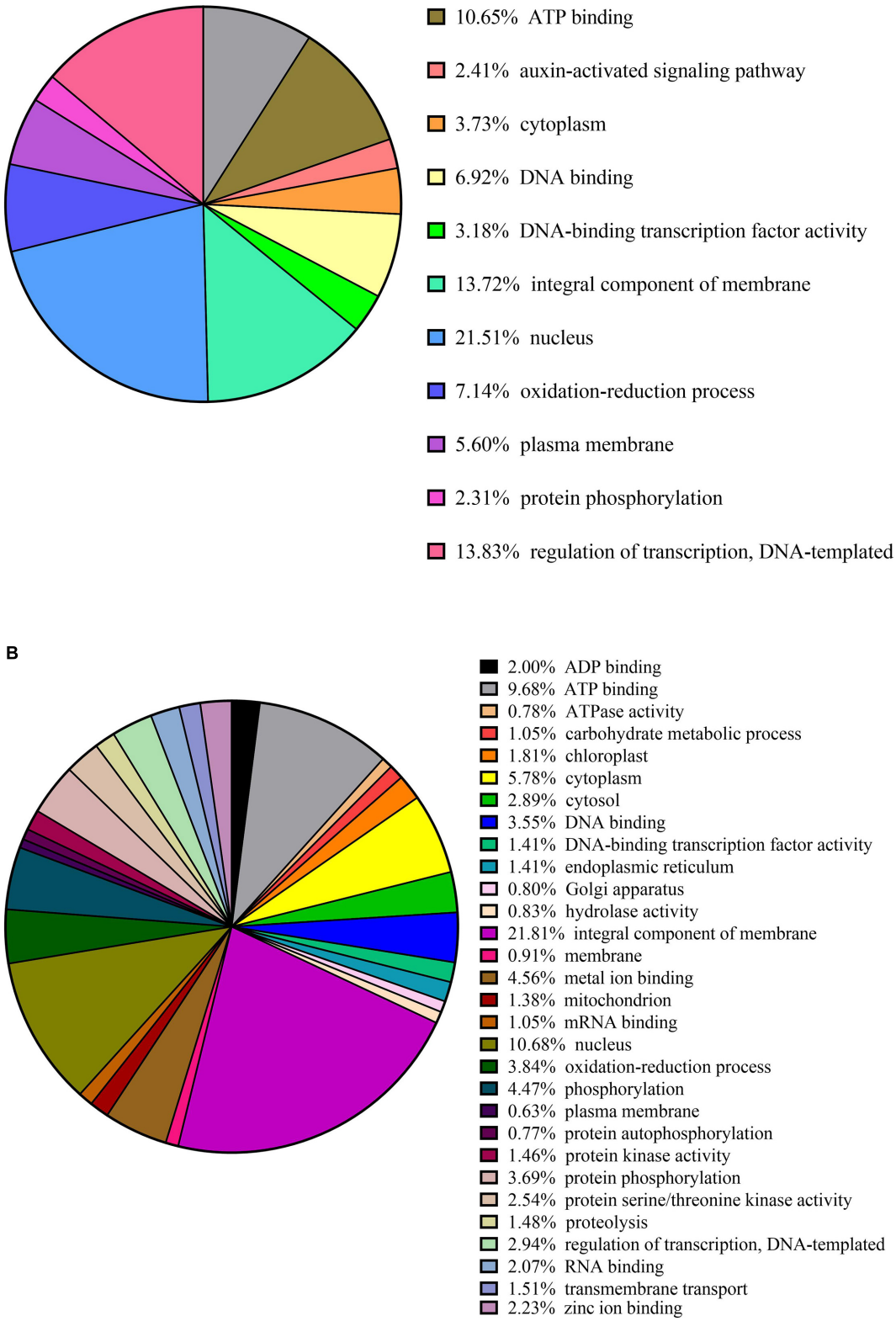


FIGURE 4 | Cluster orthologous groups (COG) function calcification of predicted citrus target genes of conserved in panel **(A)** and novel in panel **(B)** miRNAs.

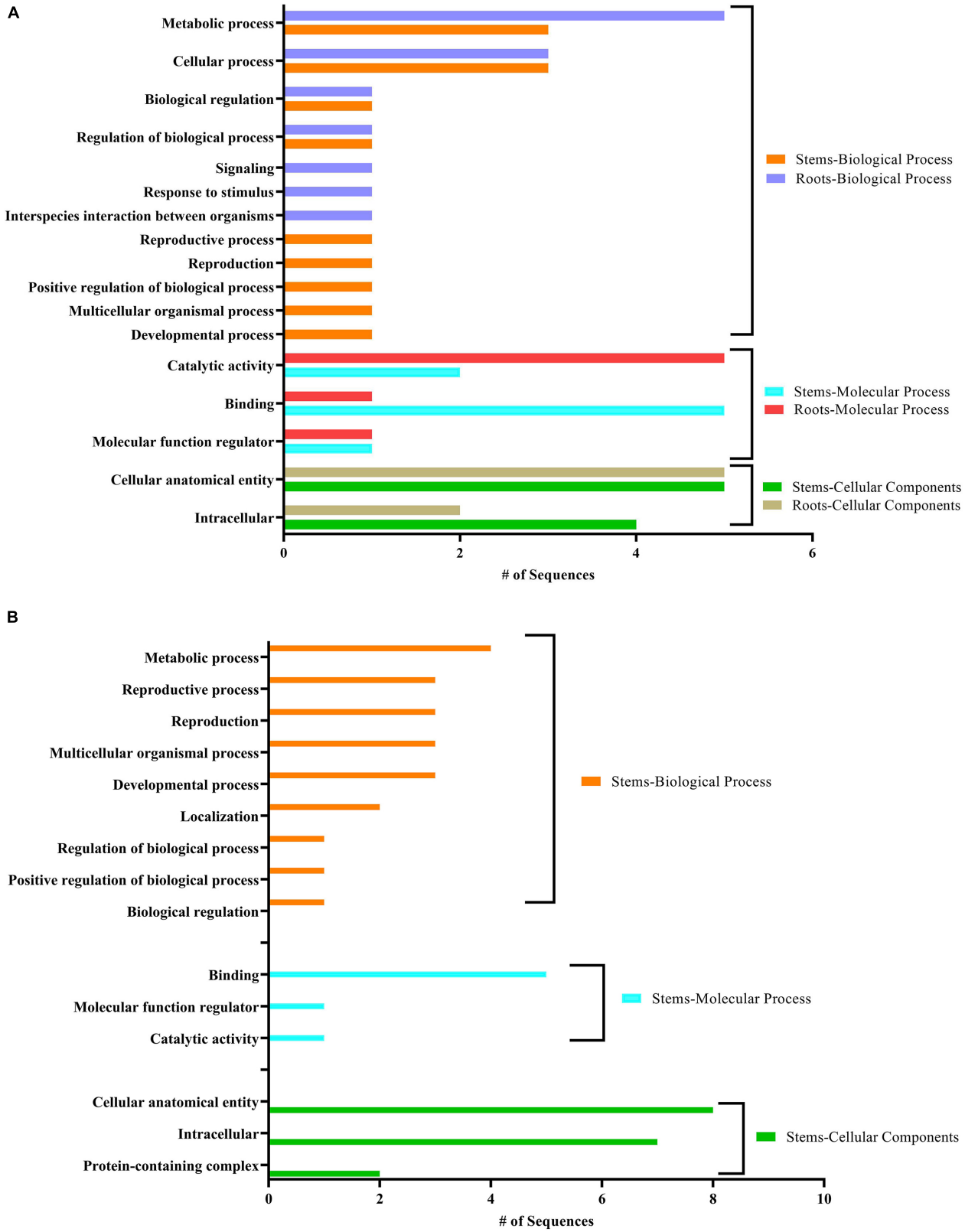


FIGURE 5 | Clusters of orthologous groups functional classification of predicted target genes of conserved in panel **(A)** and novel in panel **(B)** CDVd-responsive miRNAs. The bar graph shows the number of sequences and distribution in different functional categories of the predicted miRNA targets at gene ontology (GO) level 2.

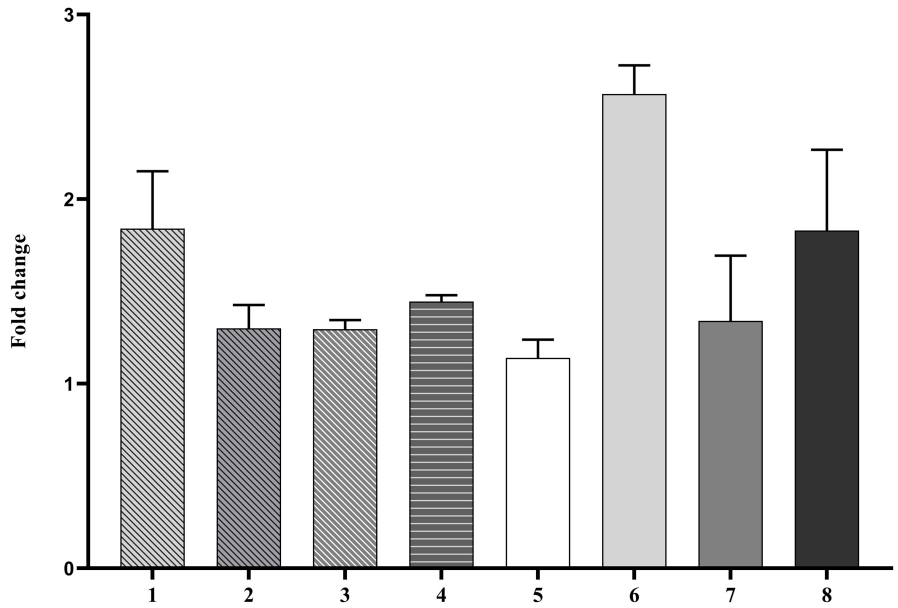


FIGURE 6 | . Differential expression profile of selected miRNA target genes. The relative gene expression was evaluated by the comparative Cq method using actin2 as a reference gene. The bar graph shows log_2 fold changes of expression levels of target genes in CDVd-infected stems and roots relative to non-infected tissues. The predicted target genes used in the analysis were (1) UDP-glucose flavonoid glucosyl-transferase (orange1.1g033614m, target of csi-miR479-1-stem); (2) DEAD/DEAH box helicase (orange1.1g028826m, target of csi-miR479-2-stem); (3) DEAD/DEAH box helicase (orange1.1g026925m, target of csi-miR479-3-stem); (4) glutathione S-transferase (orange1.1g033674m, target of csi-miR171b-stem); (5) squamosa promoter binding protein-like 2 (orange1.1g011651m, target of csi-miR156-1-stem); (6) squamosa promoter binding protein-like 3 (orange1.1g032310m, target of csi-miR156-2-stem); (7) squamosa promoter-binding protein-like transcription factor family protein (orange1.1g008776m, target of csi-miR156-3-stem); and (8) RHOMBOID-like protein, P_trifoliata_00066_mRNA_51.1, target of csi-miR535-root).

orange1.1g026925m) (Macovei et al., 2012). Both target genes may be related to the observed dwarf phenotype (**Supplementary Table 11**).

The predicted target genes of the novel miRNAs identified in this study (csi-miRNA-75, csi-miRNA-114, and csi-miRNA-435) include proteins were associated with target genes including SBP proteins (orange1.1g030599m, orange1.1g029650m, orange1.1g021420m, orange1.1g017256m, orange1.1g016971m, orange1.1g032310m orange1.1g046416m, orange1.1g011662m, orange1.1g010865m, and orange1.1g010591m; target of csi-miRNA-75); plastid-lipid associated protein (orange1.1g030218m, orange1.1g025746m, orange1.1g030180m, and orange1.1g020639m; target of csi-miRNA-114) which are structures that contain lipids and proteins that sequester the overaccumulation of carotenoids during flower development and fruit ripening (Moriguchi et al., 1998; Leitner-Dagan et al., 2006); and vacuolar protein sorting-associated proteins (orange1.1g021304m and orange1.1g017530m; target of csi-miR435), which direct protein cargo from the Golgi apparatus to the vacuoles (Xiang et al., 2013) and have been shown to be important in plant development (Cai et al., 2014) (**Supplementary Table 11**). The results suggest that CDVd-infection affects a wide range of biological functions via different miRNAs. In addition, the GO distribution analysis performed in this study, identified targets of conserved and novel CDVd-responsive miRNAs involved in various processes (**Figure 5**). Taken together, our findings might suggest that CDVd-infection could lead to developmental reprogramming and growth alterations of citrus trees, leading to the observed symptoms of reduced vegetative growth and overall smaller tree size. Future transcriptome studies could provide additional evidence to elucidate the molecular details in support of this hypothesis.

In this study, the overall miRNA profile of roots (trifoliate orange, *C. trifoliata*) was not altered in response to CDVd infection to the same extent demonstrated by the stem (navel orange, *C. sinensis*) miRNA profile. Although CDVd-derived sRNAs were detected in the roots, the trifoliate orange rootstock does not display major symptoms in response to CDVd infection (Vidalakis et al., 2004; Vernière et al., 2006; Murcia et al., 2009). This observation is consistent with the findings that the striking dwarfed citrus tree phenotype caused by CDVd infection results from the reduced vegetative growth of the stems, supporting the hypothesis that the molecular mechanisms responsible for this reprograming must be primarily active in the stems. Future greenhouse studies could provide additional evidence to support this hypothesis.

Finally, the term "Transmissible small nuclear ribonucleic acids" (TsnRNAs) was coined to identify those viroids that do not express a disease syndrome, but rather act as modifying agents of tree performance that result in desirable agronomic traits with potential economic advantages (Semancik et al., 1997; Semancik, 2003). In our study, we did not observe induction and regulation

of defense genes via the sRNA pathway in response to CDVd infection, a finding that concurs with the hypothesis that some viroid species might be considered as RNAs modifying cellular functions and plant performance.

CONCLUSION

The expression profile of CDVd-responsive miRNAs indicates that these miRNAs play a role in regulating important citrus tree growth and development processes that may participate in the cellular changes leading to the observed *C. sinensis* on *C. trifoliata* rootstock dwarf phenotype.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the NCBI Sequence Read Archive under accession numbers SRS8100788-SRS8100791 and PRJNA693870.

AUTHOR CONTRIBUTIONS

TD, IL-C, and GV conceived and designed the experiments, analyzed the data, and wrote the manuscript. TD and IL-C performed the experiments with the assistance of SB. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Citrus Research Board project "Citrus Dwarfing of Commercial Varieties using TsnRNAs" (project 5100-154) awarded to GV. Additional support was provided in part by the Citrus Research Board (project 6100), the USDA National Institute of Food and Agriculture, Hatch project 1020106.

ACKNOWLEDGMENTS

We would like to acknowledge the assistance of the dedicated personnel of the CCPP and the UC-ANR Lindcove Research and Extension Center. We would also like to acknowledge Prof. Danelle Seymour (University of California, Riverside) for providing advice regarding *Citrus trifoliata* genome. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the University of California. UC is an equal opportunity provider and employer.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.646273/full#supplementary-material

Supplementary Table 1 | Sequences submitted for stem-loop RT-qPCR of miRNAs.

Supplementary Table 2 | Specific primers used for RT-qPCR relative quantification of miRNA target genes.

Supplementary Table 3 | List of all conserved miRNAs found in *Citrus sinensis* stems.

Supplementary Table 4 | List of all conserved miRNAs found in *Citrus trifoliata* roots.

Supplementary Table 5 | List of all novel miRNAs found in stems.

Supplementary Table 6 | List of all novel miRNAs found in roots.

Supplementary Table 7 | List of predicted targets for conserved miRNAs in stems.

Supplementary Table 8 | List of predicted targets for novel miRNAs in stems.

Supplementary Table 9 | List of predicted targets for conserved miRNA in roots.

Supplementary Table 10 | List of predicted targets for novel miRNAs in roots.

Supplementary Table 11 | List of target genes of citrus dwarfing viroid-responsive conserved and novel miRNAs.

REFERENCES

Adkar-Purushothama, C. R., Brosseau, C., Giguère, T., Sano, T., Moffett, P., and Perreault, J.-P. (2015). Small RNA derived from the virulence modulating region of the potato spindle tuber viroid silences callose synthase genes of tomato plants. *Plant Cell* 27, 2178–2194. doi: 10.1105/tpc.15.00523

Adkar-Purushothama, C. R., Iyer, P. S., and Perreault, J.-P. (2017). Potato spindle tuber viroid infection triggers degradation of chloride channel protein CLC-b-like and Ribosomal protein S3a-like mRNAs in tomato plants. *Sci. Rep.* 7:8341.

Axtell, M. J., and Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell* 17, 1658–1673. doi: 10.1105/tpc.105.032185

Axtell, M. J., Snyder, J. A., and Bartel, D. P. (2007). Common functions for diverse small RNAs of land plants. *Plant Cell* 19, 1750–1769. doi: 10.1105/tpc.107.051706

Babcock, B. (2018). Economic impact of California's citrus industry. *Citrograph* 9, 36–39.

Bar-Joseph, M. (1993). Citrus viroids and citrus dwarfing in Israel. *Acta Hortic.* 349, 271–276. doi: 10.17660/actahortic.1993.349.45

Bolduc, F., Hoareau, C., St-Pierre, P., and Perreault, J.-P. (2010). In-depth sequencing of the siRNAs associated with peach latent mosaic viroid infection. *BMC Mol. Biol.* 11:16. doi: 10.1186/1471-2199-11-16

Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi: 10.1038/nrm4085

Broadbent, P., Forsyth, J. B., Hutton, R. J., and Bevington, K. B. (1992). "Guidelines for the commercial use of graft-transmissible dwarfing in Australia–potential benefits and risks," in *Proceedings of the International Society of Citriculture*, (Acireale: International Society of Citriculture), 697–701.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

Cai, Y., Zhuang, X., Gao, C., Wang, X., and Jiang, L. (2014). The Arabidopsis endosomal sorting complex required for transport III regulates internal vesicle formation of the prevacuolar compartment and is required for plant development. *Plant Physiol.* 165, 1328–1343. doi: 10.1104/pp.114.238378

Cardon, G., Höhmann, S., Klein, J., Nettesheim, K., Saedler, H., and Huijser, P. (1999). Molecular characterisation of the Arabidopsis SBP-box genes. *Gene* 237, 91–104. doi: 10.1016/s0378-1119(99)00308-x

Chen, H.-M., Chen, L.-T., Patel, K., Li, Y.-H., Baulcombe, D. C., and Wu, S.-H. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15269–15274. doi: 10.1073/pnas.1001738107

Chen, X., Zhang, Z., Liu, D., Zhang, K., Li, A., and Mao, L. (2010). SQUAMOSA promoter-binding protein-like transcription factors: star players for plant growth and development. *J. Integr. Plant Biol.* 52, 946–951. doi: 10.1111/j.1744-7909.2010.00987.x

Cohen, M. (1968). Exocortis virus as a possible factor in producing dwarf citrus trees. *Proc. Fla. State Hort. Soc.* 81, 115–119.

Czech, B., Munafò, M., Ciabrelli, F., Eastwood, E. L., Fabry, M. H., Kneuss, E., et al. (2018). piRNA-Guided genome defense: from biogenesis to silencing. *Annu. Rev. Genet.* 52, 131–157. doi: 10.1146/annurev-genet-120417-031441

da Graça, J. V., Douhan, G. W., Halbert, S. E., Keremane, M. L., Lee, R. F., Vidalakis, G., et al. (2016). Huanglongbing: an overview of a complex pathosystem ravaging the world's citrus. *J. Integr. Plant Biol.* 58, 373–387. doi: 10.1111/jipb.12437

Dadami, E., Boutla, A., Vrettos, N., Tzortzakaki, S., Karakasilioti, I., and Kalantidis, K. (2013). DICER-LIKE 4 but not DICER-LIKE 2 may have a positive effect on potato spindle tuber viroid accumulation in Nicotiana benthamiana. *Mol. Plant* 6, 232–234. doi: 10.1093/mp/sss118

Dadami, E., Dalakouras, A., and Wassenegger, M. (2017). "Viroids and RNA Silencing," in *Viroids and Satellites*, eds A. Hadidi, R. Flores, J. W. Randles, and P. Palukaitis (Boston: Academic Press), 115–124. doi: 10.1016/b978-0-12-801498-1.00011-5

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* 46, W49–W54.

Diermann, N., Matoušek, J., Junge, M., Riesner, D., and Steger, G. (2010). Characterization of plant miRNAs and small RNAs derived from potato spindle tuber viroid (PSTVd) in infected tomato. *Biol. Chem.* 391, 1379–1390.

Eamens, A. L., Smith, N. A., Dennis, E. S., Wassenegger, M., and Wang, M.-B. (2014). In Nicotiana species, an artificial microRNA corresponding to the virulence modulating region of Potato spindle tuber viroid directs RNA silencing of a soluble inorganic pyrophosphatase gene and the development of abnormal phenotypes. *Virology* 450-451, 266–277. doi: 10.1016/j.virol.2013.12.019

Farooq, M., Mansoor, S., Guo, H., Amin, I., Chee, P. W., Azim, M. K., et al. (2017). Identification and characterization of miRNA transcriptome in asiatic cotton (Gossypium arboreum) using high throughput sequencing. *Front. Plant Sci.* 8:969. doi: 10.3389/fpls.2017.00969

Ferdous, J., Sanchez-Ferrero, J. C., Langridge, P., Milne, L., Chowdhury, J., Brien, C., et al. (2017). Differential expression of microRNAs and potential targets under drought stress in barley. *Plant Cell Environ.* 40, 11–24. doi: 10.1111/pce.12764

Flores, R., Di Serio, F., Navarro, B., and Owens, R. A. (2017). "Viroid Pathogenesis," in *Viroids and Satellites*, eds A. Hadidi, R. Flores, J. W. Randles, and P. Palukaitis (Boston: Academic Press), 93–103. doi: 10.1016/b978-0-12-801498-1.00009-7

Flores, R., Gas, M.-E., Molina-Serrano, D., Nohales, M. -Á, Carbonell, A., Gago, S., et al. (2009). Viroid replication: rolling-circles, enzymes and ribozymes. *Viruses* 1, 317–334. doi: 10.3390/v1020317

Gandikota, M., Birkenbihl, R. P., Höhmann, S., Cardon, G. H., Saedler, H., and Huijser, P. (2007). The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. *Plant J.* 49, 683–693. doi: 10.1111/j.1365-313x.2006.02983.x

Gao, J., Yin, F., Liu, M., Luo, M., Qin, C., Yang, A., et al. (2015). Identification and characterisation of tobacco microRNA transcriptome using high-throughput sequencing. *Plant Biol.* 17, 591–598. doi: 10.1111/plb.12275

Gómez, G., Martínez, G., and Pallás, V. (2009). Interplay between viroid-induced pathogenesis and RNA silencing pathways. *Trends Plant Sci.* 14, 264–269. doi: 10.1016/j.tplants.2009.03.002

Gottwald (2010). Current epidemiological understanding of citrus huanglongbing. *Annu. Rev. Phytopathol.* 48, 119–139. doi: 10.1146/annurev-phyto-073009-114418

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176

Gullner, G., Komives, T., Király, L., and Schröder, P. (2018). Glutathione S-transferase enzymes in plant-pathogen interactions. *Front. Plant Sci.* 9:1836. doi: 10.3389/fpls.2018.01836

Huang, J.-H., Lin, X.-J., Zhang, L.-Y., Wang, X.-D., Fan, G.-C., and Chen, L.-S. (2019). MicroRNA sequencing revealed citrus adaptation to long-term boron toxicity through modulation of root development by miR319 and miR171. *Int. J. Mol. Sci.* 20:1422. doi: 10.3390/ijms20061422

Hutton, R. J., Broadbent, P., and Bevington, K. (2000). Viroid dwarfing for high density citrus planting. *Hortic Rev.* 24, 277–317. doi: 10.1002/9780470650776.ch6

Itaya, A., Zhong, X., Bundschuh, R., Qi, Y., Wang, Y., Takeda, R., et al. (2007). A structured viroid RNA serves as a substrate for dicer-like cleavage to produce biologically active small RNAs but is resistant to RNA-induced silencing complex-mediated degradation. *J. Virol.* 81, 2980–2994. doi: 10.1128/jvi.02339-06

Jagadeeswaran, G., Zheng, Y., Sumathipala, N., Jiang, H., Arrese, E. L., Soulages, J. L., et al. (2010). Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. *BMC Genomics* 11:52. doi: 10.1186/1471-2164-11-52

Jia, L., Zhang, D., Qi, X., Ma, B., Xiang, Z., and He, N. (2014). Identification of the conserved and novel miRNAs in Mulberry by high-throughput sequencing. *PLoS One* 9:e104409. doi: 10.1371/journal.pone.0104409

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., et al. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46, D335–D342.

Kanaoka, M. M., Urban, S., Freeman, M., and Okada, K. (2005). An Arabidopsis rhomboid homolog is an intramembrane protease in plants. *FEBS Lett.* 579, 5723–5728. doi: 10.1016/j.febslet.2005.09.049

Kawahara, Y., Endo, T., Omura, M., Teramoto, Y., Itoh, T., Fujii, H., et al. (2020). Mikan Genome Database (MiGD): integrated database of genome annotation, genomic diversity, and CAPS marker information for mandarin molecular breeding. *Breed. Sci.* 70, 200–211. doi: 10.1270/jsbbs.19097

Kou, S.-J., Wu, X.-M., Liu, Z., Liu, Y.-L., Xu, Q., and Guo, W.-W. (2012). Selection and validation of suitable reference genes for miRNA expression normalization by quantitative RT-PCR in citrus somatic embryogenic and adult tissues. *Plant Cell Rep.* 31, 2151–2163. doi: 10.1007/s00299-012-1325-x

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39, D152–D157.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162.

Lambin, E. F. (2012). Global land availability: malthus versus Ricardo. *Glob. Food Sec.* 1, 83–87. doi: 10.1016/j.gfs.2012.11.002

Lavagi-Craddock, I., Campos, R., Pagliaccia, D., Kapaun, T., Lovatt, C., and Vidalakis, G. (2020). Citrus dwarfing viroid reduces canopy volume by affecting shoot apical growth of navel orange trees grown on trifoliate orange rootstock. *J. Citrus Pathol.* 7, 1–6.

Lei, J., and Sun, Y. (2014). miR-PREFeR: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-Seq data. *Bioinformatics* 30, 2837–2839. doi: 10.1093/bioinformatics/btu380

Leitner-Dagan, Y., Ovadis, M., Shklarman, E., Elad, Y., Rav David, D., and Vainstein, A. (2006). Expression and functional analyses of the plastid lipid-associated protein CHRC suggest its role in chromoplastogenesis and stress. *Plant Physiol.* 142, 233–244. doi: 10.1104/pp.106.082404

Liang, W.-W., Huang, J.-H., Li, C.-P., Yang, L.-T., Ye, X., Lin, D., et al. (2017). MicroRNA-mediated responses to long-term magnesium-deficiency in *Citrus sinensis* roots revealed by Illumina sequencing. *BMC Genomics* 18:657. doi: 10.1186/s12864-017-3999-5

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Lu, Y.-B., Qi, Y.-P., Yang, L.-T., Guo, P., Li, Y., and Chen, L.-S. (2015). Boron-deficiency-responsive microRNAs and their targets in Citrus sinensis leaves. *BMC Plant Biol.* 15:271. doi: 10.1186/s12870-015-0642-y

Ma, C.-L., Qi, Y.-P., Liang, W.-W., Yang, L.-T., Lu, Y.-B., Guo, P., et al. (2016). MicroRNA regulatory mechanisms on citrus sinensis leaves to magnesium-deficiency. *Front. Plant Sci.* 7:201. doi: 10.3389/fpls.2016.00201

Ma, Z., Hu, X., Cai, W., Huang, W., Zhou, X., Luo, Q., et al. (2014). Arabidopsis miR171-targeted scarecrow-like proteins bind to GT cis-elements and mediate gibberellin-regulated Chlorophyll biosynthesis under light conditions. *PLoS Genet.* 10:e1004519. doi: 10.1371/journal.pgen.1004519

Macovei, A., Vaid, N., Tula, S., and Tuteja, N. (2012). A new DEAD-box helicase ATP-binding protein (OsABP) from rice is responsive to abiotic stress. *Plant Signal. Behav.* 7, 1138–1143. doi: 10.4161/psb.21343

Mafra, V., Kubo, K. S., Alves-Ferreira, M., Ribeiro-Alves, M., Stuart, R. M., Boava, L. P., et al. (2012). Reference genes for accurate transcript normalization in citrus genotypes under different experimental conditions. *PLoS One* 7:e31263. doi: 10.1371/journal.pone.0031263

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Mendel, K. (1968). "Interrelations between tree performance and some virus diseases," in *Proceedings of the 4th Conference IOCV*, (Riverside, CA: IOCV), 310–313.

Minoia, S., Carbonell, A., Di Serio, F., Gisel, A., Carrington, J. C., Navarro, B., et al. (2014). Specific argonautes selectively bind small RNAs derived from potato spindle tuber viroid and attenuate viroid accumulation in vivo. *J. Virol.* 88, 11933–11945. doi: 10.1128/jvi.01404-14

Mishra, A. K., Duraisamy, G. S., Matoušek, J., Radisek, S., Javornik, B., and Jakse, J. (2016). Identification and characterization of microRNAs in Humulus lupulus using high-throughput sequencing and their response to Citrus bark cracking viroid (CBCVd) infection. *BMC Genomics* 17:919. doi: 10.1186/s12864-016-3271-4

Moriguchi, T., Kita, M., Endo-Inagaki, T., Ikoma, Y., and Omura, M. (1998). Characterization of a cDNA homologous to carotenoid-associated protein in citrus fruits1The nucleotide sequence reported in this paper has been submitted to DDBJ under accession No. AB011797 (CitPAP). Contribution No. 1112 of the NIFTS.1. *Biochim. Biophys. Acta* 1442, 334–338.

Motameny, S., Wolters, S., Nürnberg, P., and Schumacher, B. (2010). Next generation sequencing of miRNAs - strategies, resources and methods. *Genes* 1, 70–84. doi: 10.3390/genes1010070

Moxon, S., Jing, R., Szittya, G., Schwach, F., Rusholme Pilcher, R. L., Moulton, V., et al. (2008). Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.* 18, 1602–1609. doi: 10.1101/gr.080127.108

Murcia, N., Bernad, L., Serra, P., Hashemian, S. M. B., and Duran-Vila, N. (2009). Molecular and biological characterization of natural variants of Citrus dwarfing viroid. *Arch. Virol.* 154, 1329–1334. doi: 10.1007/s00705-009-0430-9

Navarro, B., Gisel, A., Rodio, M. E., Delgado, S., Flores, R., and Di Serio, F. (2012). Small RNAs containing the pathogenic determinant of a chloroplast-replicating viroid guide the degradation of a host mRNA as predicted by RNA silencing. *Plant J.* 70, 991–1003. doi: 10.1111/j.1365-313x.2012.04940.x

Navarro, B., Pantaleo, V., Gisel, A., Moxon, S., Dalmay, T., Bisztray, G., et al. (2009). Deep sequencing of viroid-derived small RNAs from grapevine provides new insights on the role of RNA silencing in plant-viroid interaction. *PLoS One* 4:e7686. doi: 10.1371/journal.pone.0007686

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., et al. (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* 25, 1396–1405. doi: 10.1038/sj.emboj.7600970

Owens, R. A., Tech, K. B., Shao, J. Y., Sano, T., and Baker, C. J. (2012). Global analysis of tomato gene expression during Potato spindle tuber viroid infection reveals a complex array of changes affecting hormone signaling. *Mol. Plant. Microbe. Interact.* 25, 582–598. doi: 10.1094/mpmi-09-11-0258

Platt, R. G. (1973). *Treatment of Frost-Inured Citrus, Avocados. California Citrograph*. Available online at: http://agris.fao.org/agris-search/search.do?recordID=US201303229376 (accessed November 15, 2020).

Ramesh, S. V., Yogindran, S., Gnanasekaran, P., Chakraborty, S., Winter, S., and Pappu, H. R. (2020). Virus and viroid-derived small RNAs as modulators of host gene expression: molecular insights into pathogenesis. *Front. Microbiol.* 11:614231. doi: 10.3389/fmicb.2020.614231

Reis, R. S., Eamens, A. L., Roberts, T. H., and Waterhouse, P. M. (2015). Chimeric DCL1-partnering proteins provide insights into the MicroRNA pathway. *Front. Plant Sci.* 6:1201. doi: 10.3389/fpls.2015.01201

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., and Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell* 110, 513–520.

Riese, M., Höhmann, S., Saedler, H., Münster, T., and Huijser, P. (2007). Comparative analysis of the SBP-box gene families in P. patens and seed plants. *Gene* 401, 28–37. doi: 10.1016/j.gene.2007.06.018

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73

Semancik, J. S. (2003). "Considerations for the introduction of viroids for economic advantage," in *Viroids*, eds A. Hadidi, R. Flores, J. W. Randles, and J. S. Semancik (Clayton VIC: CSIRO Publishing), 357–362.

Semancik, J. S., Rakowski, A. G., Bash, J. A., and Gumpf, D. J. (1997). Application of selected viroids for dwarfing and enhancement of production of "Valencia" orange. *J. Hortic. Sci.* 72, 563–570. doi: 10.1080/14620316.1997.11515544

Shi, M., Hu, X., Wei, Y., Hou, X., Yuan, X., Liu, J., et al. (2017). Genome-wide profiling of small RNAs and degradome revealed conserved regulations of miRNAs on auxin-responsive genes during fruit enlargement in peaches. *Int. J. Mol. Sci.* 18:2599. doi: 10.3390/ijms18122599

Taliansky, M., Samarskaya, V., Zavriev, S. K., Fesenko, I., Kalinina, N. O., and Love, A. J. (2021). RNA-Based technologies for engineering plant virus resistance. *Plants* 10, 82–101. doi: 10.3390/plants10010082

Treiber, T., Treiber, N., and Meister, G. (2019). Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat. Rev. Mol. Cell Biol.* 20, 5–20. doi: 10.1038/s41580-018-0059-1

Tsushima, T., Murakami, S., Ito, H., He, Y.-H., Charith Raj, A. P., and Sano, T. (2011). Molecular characterization of Potato spindle tuber viroid in dahlia. *J. Gen. Plant Pathol.* 77, 253–256.

Tucker, D. P. H., and Wheaton, T. A. (1978). Trends in higher citrus planting densities. *Proc. Fla. State Hort. Soc.* 91, 36–40.

USDA-NASS (2019). *Citrus Fruits 2019 Summary August 2019*. Available online at: https://www.nass.usda.gov/Publications/Todays_Reports/reports/cfrt0819.pdf. (accessed November 15, 2020).

Van Duyn, M. A., and Pivonka, E. (2000). Overview of the health benefits of fruit and vegetable consumption for the dietetics professional: selected literature. *J. Am. Diet. Assoc.* 100, 1511–1521. doi: 10.1016/s0002-8223(00)00420-x

Verburg, P. H., Mertz, O., Erb, K.-H., Haberl, H., and Wu, W. (2013). Land system change and food security: towards multi-scale land system solutions. *Curr. Opin. Environ. Sustain.* 5, 494–502. doi: 10.1016/j.cosust.2013.07.003

Vernière, C., Perrier, X., Dubois, C., Dubois, A., Botella, L., Chabrier, C., et al. (2006). Interactions between citrus viroids affect symptom expression and field performance of clementine trees grafted on trifoliate orange. *Phytopathology* 96, 356–368. doi: 10.1094/phyto-96-0356

Vidalakis, G., Gumpf, D. J., Bash, J. A., and Semancik, J. S. (2004). Finger imprint of Poncirus trifoliata: a specific interaction of a viroid, a host, and irrigation. *Plant Dis.* 88, 709–713. doi: 10.1094/pdis.2004.88.7.709

Vidalakis, G., Pagliaccia, D., Bash, J. A., Afunian, M., and Semancik, J. S. (2011). Citrus dwarfing viroid: effects on tree size and scion performance specific to Poncirus trifoliata rootstock for high-density planting. *Ann. Appl. Biol.* 158, 204–217. doi: 10.1111/j.1744-7348.2010.00454.x

Vogt, T., and Jones, P. (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.* 5, 380–386. doi: 10.1016/s1360-1385(00)01720-9

Wang, J., Mei, J., and Ren, G. (2019). Plant microRNAs: biogenesis. *Front. Plant Sci.* 10:360. doi: 10.3389/fpls.2019.00360

Wang, M.-B., Bian, X.-Y., Wu, L.-M., Liu, L.-X., Smith, N. A., Isenegger, D., et al. (2004). On the role of RNA silencing in the pathogenicity and evolution of viroids and viral satellites. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3275–3280. doi: 10.1073/pnas.0400104101

Wang, Y., Shibuya, M., Taneda, A., Kurauchi, T., Senda, M., Owens, R. A., et al. (2011). Accumulation of Potato spindle tuber viroid-specific small RNAs is accompanied by specific changes in gene expression in two tomato cultivars. *Virology* 413, 72–83. doi: 10.1016/j.virol.2011.01.021

Wassenegger, M., Heimes, S., and Sänger, H. L. (1994). An infectious viroid RNA replicon evolved from an in vitro-generated non-infectious viroid deletion mutant via a complementary deletion in vivo. *EMBO J.* 13, 6172–6177. doi: 10.1002/j.1460-2075.1994.tb06964.x

Wu, G., and Poethig, R. S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development* 133, 3539–3547. doi: 10.1242/dev.02521

Xiang, L., Etxeberria, E., and Van den Ende, W. (2013). Vacuolar protein sorting mechanisms in plants. *FEBS J.* 280, 979–993. doi: 10.1111/febs.12092

Xie, K., Wu, C., and Xiong, L. (2006). Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol.* 142, 280–293. doi: 10.1104/pp.106.084475

Xie, R., Zhang, J., Ma, Y., Pan, X., Dong, C., Pang, S., et al. (2017). Combined analysis of mRNA and miRNA identifies dehydration and salinity responsive key molecular players in citrus roots. *Sci. Rep.* 7:42094.

Yao, L. H., Jiang, Y. M., Shi, J., Tomás-Barberán, F. A., Datta, N., Singanusong, R., et al. (2004). Flavonoids in food and their health benefits. *Plant Foods Hum. Nutr.* 59, 113–122.

Yu, S., Galvão, V. C., Zhang, Y.-C., Horrer, D., Zhang, T.-Q., Hao, Y.-H., et al. (2012). Gibberellin regulates the Arabidopsis floral transition through miR156-Targeted SQUAMOSA PROMOTER BINDING–LIKE transcription factors. *Plant Cell* 24, 3320–3332. doi: 10.1105/tpc.112.101014

Zavallo, D., Debat, H. J., Conti, G., Manacorda, C. A., Rodriguez, M. C., and Asurmendi, S. (2015). Differential mRNA accumulation upon early *Arabidopsis thaliana* infection with ORMV and TMV-Cg Is associated with distinct endogenous small RNAs level. *PLoS One* 10:e0134719. doi: 10.1371/journal.pone.0134719

Zhang, B., Pan, X., and Stellwag, E. J. (2008). Identification of soybean microRNAs and their targets. *Planta* 229, 161–182. doi: 10.1007/s00425-008-0818-x

Zhang, L., Chia, J.-M., Kumari, S., Stein, J. C., Liu, Z., Narechania, A., et al. (2009). A genome-wide characterization of microRNA genes in maize. *PLoS Genet.* 5:e1000716. doi: 10.1371/journal.pgen.1000716

Zhang, R., Marshall, D., Bryan, G. J., and Hornyik, C. (2013). Identification and characterization of miRNA transcriptome in potato by high-throughput sequencing. *PLoS One* 8:e57233. doi: 10.1371/journal.pone.0057233

Zhang, T., Hu, S., Yan, C., Li, C., Zhao, X., Wan, S., et al. (2017). Mining, identification and function analysis of microRNAs and target genes in peanut (Arachis hypogaea L.). *Plant Physiol. Biochem.* 111, 85–96.

Zhang, Y., Wang, Y., Gao, X., Liu, C., and Gai, S. (2018). Identification and characterization of microRNAs in tree peony during chilling induced dormancy release by high-throughput sequencing. *Sci. Rep.* 8:4537.

Zhou, Y., Luo, S., Hameed, S., Xiao, D., Zhan, J., Wang, A., et al. (2020). Integrated mRNA and miRNA transcriptome analysis reveals a regulatory network for tuber expansion in Chinese yam (Dioscorea opposita). *BMC Genomics* 21:117. doi: 10.1186/s12864-020-6492-5

Zhu, L., Ow, D. W., and Dong, Z. (2018). Transfer RNA-derived small RNAs in plants. *Sci. China Life Sci.* 61, 155–161. doi: 10.1007/s11427-017-9167-5

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.

Copyright © 2021 Dang, Lavagi-Craddock, Bodaghi and Vidalakis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

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