CHARACTERIZATION OF PLANT VIROMES USING DIFFERENT VIRAL NUCLEIC ACIDS ENRICHMENT STRATEGIES AND HIGH-THROUGHPUT SEQUENCING PLATFORMS

Anja Pecman

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Maja Ravnikar, PhD, National Institute of Biology Ljubljana, Slovenia
Co-Supervisor: Ion Gutierrez Aguirre, PhD, National Institute of Biology Ljubljana, Slovenia

Evaluation Board:
Prof. Kristina Gruden, PhD, Chair, National Institute of Biology Ljubljana, Slovenia
Prof. Jernej Jakše, PhD, Member, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia
Prof. Sebastien Massart, PhD, Member, University of Liège, Liège, Belgium

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Anja Pecman

CHARACTERIZATION OF PLANT VIROMES USING DIFFERENT VIRAL NUCLEIC ACIDS ENRICHMENT STRATEGIES AND HIGH-THROUGHPUT SEQUENCING PLATFORMS

Doctoral Dissertation

KARAKTERIZACIJA RASTLINSKEGA VIROMA Z UPORABO RAZLIČNIH STRATEGIJ OBOGATITVE VIRUSNIH NUKLENSKIH KISLIN IN PLATFORM VISOKO ZMOGLJIVEGA SEKVENCIRANJA

Doktorska disertacija

Supervisor: Prof. Maja Ravnikar, PhD
Co-Supervisor: Ion Gutierrez Aguirre, PhD

Ljubljana, Slovenia, September 2022

Acknowledgments

The research presented in this thesis was supported by the funding of the Slovenian Research Agency (ARRS) through a PhD grant, core financing grants (P4-0407, P4-0165) and projects NanoPhyto (L7-2632), Euphresco NGS-Detect, and Euphresco VIRFAST and by Ministry of Agriculture, Forestry and Food of the Republic of Slovenia through the ERA-NET ArimNet2 project (EMERAMB). The short-term scientific missions (STSM) were supported by COST Action FA1407 (DIVAS) and COST Action CA15223.

Firstly, I would like to express my sincere gratitude to my mentors Maja Ravnikar and Ion Gutierrez-Aguirre for their continuous support, patience, motivation, great advice, and research enthusiasm. Besides my mentors, I am very thankful to Denis Kutnjak for teaching me many skills in the lab, for introducing to me the high throughput sequencing technology, and for teaching me about its bioinformatics analysis.

This thesis, however, is a result of a great collaboration with colleagues from FERA ltd (York, UK), so my sincere thanks go to Neil Boonham, Ian Adams, and Adrian Fox. I had the privilege to spend part of my PhD in their laboratory on short-term scientific missions. They hosted me in 2016, 2017, 2018, and 2019; as a result of all these collaborations, three publications included in this thesis were produced. Thanks also to Ummey Hany for help with the library preparation and sequencing and Sam McGreig for help with bioinformatic analysis. Big thanks, furthermore, to all of you for teaching me sequencing and analyzing skills and introducing me to your laboratories at FERA ltd (York, UK). Thanks to Tomaž Curk (University of Ljubljana, Faculty of Computer and Information Science) for providing servers for the data analysis of high-throughput sequences.
Many thanks also to members of the dissertation committee, Kristina Gruden, Jernej Jakše, and Sebastien Massart.

Thanks to IPSP (Institute for Sustainable Plant Protection (Italy), Marina Ciuffo) for providing the henbane mosaic virus sample material, the University of Ljubljana Botanical Garden for providing seeds of H. niger, Heiko Ziebell for providing the pea necrotic yellow dwarf nanovirus sample material, and Mike Rott for seeds of Phaseolus vulgaris (Black Turtle) infected with endornaviruses.

I would also like to thank Larisa Gregur for teaching me the diagnostic part of our laboratories and Nataša Mehle for giving me this opportunity. Thanks to Juš Smole, Katarina Bačnik, and Lidija Matičič for technical assistance with the host range experiments. Thanks to Polona Kogovšek, Tjaša Jakomin, Jakob Brodarič and Paul Rivarez for in-office discussions, information, and company. Thanks to virology group and to the entire FITO group. In all my years working with all of you, I am more than grateful for a friendly and supportive environment.

The last but not least is the thanks to my friends and all my family members for the support from the very beginning until today and for helping me to become the person who and where I am today.

Abstract

Plant viruses are very important plant pathogens, causing economic losses by infecting cultivated plants, causing diseases, and consequently reducing crop quality and quantity. In recent years, the development of high throughput sequencing (HTS) technologies has dramatically broadened the possibilities for plant virus research and diagnostics, enabling the discovery of new or obscure viruses and virus strains and the rapid sequencing of their genomes. HTS enables the sequencing of all nucleic acids in the sample; however, due to its generic approach, it may detect the more abundant host nucleic acids while overlooking less abundant yet important viral ones. To overcome this problem, different laboratories use different sequencing platforms and sample preparation protocols. In the first study, we compared the two most widely used sample preparation protocols for viral nucleic acid enrichment (small RNAs vs. ribosomal RNA-depleted total RNA) to perform the generic detection of plant viruses and viroids on the Illumina platform. In the study, we included viruses with different genome organizations and viroids. All selected viruses/viroids were detected with both protocols; however, a putative novel cytorhabdovirus, discovered in this study, was only detected by analyzing the data generated from ribosomal RNA-depleted total RNA and not from the small RNA dataset. The obtained knowledge was then used to investigate tomato sample with unknown etiology symptoms. HTS results identified a mixed infection of three different viruses: potato virus M (*Carlavirus, Betaflexiviridae*), southern tomato virus (*Amalgavirus, Amalgamaviridae*), and (for the first time in tomato and in Slovenia) a new strain of henbane mosaic virus (*Potyvirus, Potyviridae*). In the next step, the complete genomic sequence of henbane mosaic virus (HMV) was assembled from the HTS reads for the first time. By re-inoculation of the infected material on selected test plants, HMV was isolated and a host range analysis was performed, demonstrating that the virus was able to infect and cause symptoms in several plant species, including tomato. Thus, we demonstrated the usability and added value of high throughput sequencing as a diagnostic technique. With this in mind and with the increasing need for fast HTS analysis, the third study was conducted. We compared the most established HTS platform (MiSeq, Illumina) using ribosomal RNA-depleted total RNA as RNA input, with different library preparation protocols using a smaller, more affordable MinION sequencer (Oxford Nanopore Technologies) for the detection of plant viruses and viroids. Protocol comparisons were performed on five selected samples, containing eleven plant viruses with different genome organizations and two viroids, which are the first example of a circular dsRNA pathogen being sequenced using nanopore technology. The results of this study suggested that, upon selection of appropriate library preparation protocols, nanopore MinION sequencing can be used for the detection of plant viruses and viroids with similar performance as Illumina sequencing; this opens the door for implementation in official diagnostics.

Povzetek

Rastlinski virusi so zelo pomembni patogeni mikrobi, saj preko okužb kulturnih rastlin in povzročene bolezni vplivajo na povečanje gospodarske škode in posledično na zmanjševanje kakovosti in količine pridelka. V zadnjih letih so se zaradi razvoja tehnologije visoko zmogljivega sekvenciranja (high throughput sequencing - HTS) povečale možnosti za raziskave in diagnostiko rastlinskih virusov. Slednje pa je omogočilo odkrivanje novih ali manj raziskanih virusov in virusnih različkov ter hitro sekvenciranje njihovih genomov. HTS omogoča sekvenciranje vseh nukleinskih kislin v vzorcu, vendar pa lahko z njim, zaradi popolnoma generičnega pristopa, včasih zaznamo v glavnem bolj zastopane nukleinske kisline gostitelja, obenem pa spregledamo tiste (npr. virusen), ki so manj zastopane, a še vedno pomembne. Da bi rešili to težavo, v različnih laboratorijih uporabljajo različne platforme za sekvenciranje in različne postopke priprave vzorcev. V prvi študiji smo se odločili primerjati dva najpogosteje uporabljena protokola priprave vzorcev za obogatitev virusnih nukleinskih kislin (sekvenciranje malih RNA ter celokupne RNA z odstranjeno ribosomalno RNA) z namenom generičnega zaznavanja rastlinskih virusov in viroidov na platformi Illumina. V študijo smo vključili viruse z različnimi organizacijami genoma in viroide. Vse izbrane viruse/viroide smo zaznali z obema protokoloma; le domnevno nov cytorhabdovirus, ki smo ga zaznali v tej študiji, smo odkrili le z analizo podatkov pridobljenih iz celokupne RNA z odstranjeno ribosomalno RNA. Pridobljeno znanje smo nato uporabili pri raziskavi, v kateri smo iskali viruse v vzorcu paradižnika z neznano etiologijo. Rezultati HTS tega vzorca so razkrili mešano okužbo rastline s tremi različnimi virusi: virusom M krompirja (*Carlavirus, Betaflexiviridae*), južnim virusom paradižnika (*Amalgavirus, Amalgamaviridae*) in (prvič na paradižniku in v Sloveniji) novim različkom virusa mozaika zobnika (*Potyvirus, Potyviridae*). Z uporabo HTS smo v naslednjem koraku prvič sestavili celoten genom virusa mozaika zobnika (HMV). S ponovno inokulacijo okuženega materiala na izbrane testne rastline smo HMV osamili in izvedli analizo nabora gostiteljev. Le ta je pokazala, da virus lahko okuži in povzroči bolezenske znake pri več rastlinskih vrstah, vključno s paradižnikom. S tem smo pokazali uporabnost in dodano vrednost HTS kot diagnostične tehnike. Z naraščajočo potrebo po hitri analizi HTS smo izvedli tretjo študijo, v kateri smo primerjali najbolj uveljavljeno platformo HTS (MiSeq, Illumina) z različnimi protokoli priprave knjižnic ter uporabo manjšega, cenovno ugodnejšega sekvenatorja MinION (Oxford Nanopore Technologies) za detekcijo rastlinskih virusov in viroidov. Primerjave protokolov smo izvedli na petih izbranih vzorcih, ki so vsebovali enajst rastlinskih virusov z različnimi organizacijami genoma in dva viroida. Pri tem smo prvič določili nukleotidno zaporedje viroidov z uporabo sekvenciranja s pomočjo nanopor. Rezultati te študije so pokazali, da se lahko, v primeru izbire primerne priprave knjižnice, sekvenciranje s pomočjo nanopor, uporabi za odkrivanje rastlinskih virusov in viroidov, s podobno učinkovitostjo kot sekvenciranje s tehnologijo Illumina; to pa odpira vrata za implementacijo sekvenciranja s pomočjo nanopor v uradno diagnostiko.

Contents

Abbreviations

AMV	...	alfalfa mosaic virus
BLASTx	...	Basic Local Alignment Search Tool for searching protein databases using a translated nucleotide query
CaMV	...	cauliflower mosaic virus
cDNA	...	complementary DNA
CCyV1	...	cabbage cytorhabdovirus 1
CLVd	...	Columnea latent viroid
CSNV	...	Chrysanthemum stem necrosis virus
DNA	...	deoxyribonucleic acid
dsDNA +/−	...	positive/negative double-stranded DNA virus genome organization
dsDNA-RT +/−	...	positive/negative genome organization of double-stranded DNA virus that replicates through an RNA intermediate
dsRNA	...	double-stranded RNA
dsRNA +/−	...	positive/negative double-stranded RNA virus genome organization
HMV	...	henbane mosaic virus
HTS	...	high throughput sequencing
NGS	...	next generation sequencing, recently rephrased to high throughput sequencing
PCR	...	polymerase chain reaction
PepMV	...	pepino mosaic virus
PLMVd	...	peach latent mosaic viroid
PMV	...	potato virus M
PNYDV	...	pea necrotic yellow dwarf virus
PVeV 1,2,3	...	Phaseolus vulgaris alphaendornavirus 1,2,3
PVY	...	potato virus Y
RNA	...	ribonucleic acid
rRNA	...	ribosomal RNA
rRNA-depleted totRNA	...	ribosomal RNA-depleted total RNA
sRNA	...	small RNA
ssDNA +	...	positive single-stranded DNA virus genome organization
ssRNA -	...	negative sense single-stranded RNA virus genome organization
ssRNA +	...	positive sense single-stranded RNA virus genome organization
ssRNA-RT +	...	positive sense genome organization of single-stranded RNA viruses that replicate through a DNA intermediate
STV	...	southern tomato virus

TASVd	...	tomato apical stunt viroid
TMV	...	tobacco mosaic virus
ToCV	...	tomato chlorosis virus
ToMV	...	tomato mosaic virus
totRNA	...	total RNA
TSWV	...	tomato spotted wilt orthotospovirus
TYLCV	...	tomato yellow leaf curl virus
VANA	...	virion-associated nucleic acids

Chapter 1

Introduction

1.1 Detection and Identification of Plant Viruses and Viroids – from the Early Beginnings to HTS-Based Diagnostics

Scientific investigation of plant viruses started in late 19th/beginning of 20th century and was mainly based on the visual inspection of viral symptoms on hosts, inoculated test plants, and later on (1930) on the use of electron microscopy (Boonham et al., 2014; Hull, 2014). The main drawback of such non-specific methods was the inability to identify viral infection to the species level. Later with the introduction of virus specific serological (Clark et al., 1977) and molecular (Vunsh et al., 1990) methods, a new era of virus diagnostics began. However, specific methods are not without drawbacks as they can only identify the pathogen being targeted and potentially miss unexpected ones. Furthermore, plant viruses have high sequence diversity and lack of conserved genes across taxa, so they cannot be detected by universal methods, such as DNA metabarcoding of 16srRNA and Internal transcribed spacer, which are typically applied to bacteria and fungi, respectively. Consequently, viruses are considered one of the most challenging entities to detect. In 2009, next generation sequencing (NGS) (Adams et al., 2009; Al Rwahnih et al., 2009; Donaire et al., 2009; Kreuze et al., 2009), recently rephrased to high-throughput sequencing (HTS), was introduced into plant virology and revolutionized the detection and discovery of (new) viruses due to its generic sequencing approach (Adams & Fox, 2016; Boonham et al., 2014). The method does not require any prior knowledge of the pathogen being tested yet delivers a strain-specific result (Adams & Fox, 2016). In recent years, HTS has become an important tool in plant virology not only for the discovery of new viruses and virus strains but also for the rapid generation of viral complete genome sequences (reviewed in Barba et al., 2013; Massart et al., 2014; Rivarez et al., 2021; Roossinck et al., 2015; Villamor et al., 2019). Furthermore, due to the high performance of the technology for plant virus detection and discovery, as well as due to increase in sequencing throughput and decrease in prize per sample, HTS applications are being progressively incorporated into plant virus diagnostics environments (Bester et al., 2021; Malapi-Wight et al., 2021; Maree etal., 2018; Mehle et al., 2019; Mehle et al., 2018; Rivarez et al., 2021; Rott et al., 2017). Thus, in general, the detection of plant virus/viroids has greatly benefited from the emergence of HTS. However, in relation to the remarkably larger number of known/new virus sequences being detected, new bottlenecks have raised. For instance, the new viruses detected by HTS are often characterized by their genome sequence only, lacking biological characterisation (Rivarez et al., 2021) and consequently phytosanitary risk analysis. To address these challenges, a framework proposing guidelines for researchers has been proposed in 2017 (Massart et al., 2017), and it is currently updated (Massart et al., 2022).

1.2 Viral Nucleic Acid Preparation Strategies for High-Throughput Sequencing of Plant Viruses

HTS process consist of several steps, combining wet-lab steps: sampling, nucleic acid extraction, library preparation, sequencing, and dry lab steps: analysis of raw reads, identification of targets, analysis of controls and target confirmation, interpretation and reporting (Lebas et al., 2022). In recent years, different sample preparation protocols (nucleic acid extraction/library preparation) have been developed. The main goal of some protocols was to increase the sensitivity of the HTS by enriching viral nucleic acids over the host ones. The chosen enrichment strategy should take into account the genome organization and replication strategy of all virus groups. Based on these two features, the Baltimore classification groups viruses into: Group I: double-stranded DNA (dsDNA +/−), Group II: single-stranded DNA (ssDNA +), Group III: double-stranded RNA (dsRNA +/−), Group IV: positive sense single-stranded RNA (ssRNA +), Group V: negative sense single-stranded RNA (ssRNA −), Group VI: positive sense single-stranded RNA viruses that replicate through a DNA intermediate (ssRNA-RT +), and Group VII: double-stranded DNA viruses that replicate through an RNA intermediate (dsDNA-RT +/−) (Baltimore, 1971).

The most common nucleic acid fractions prepared as inputs in HTS for plant virus detection are: total RNA (totRNA), ribosomal RNA-depleted total RNA (rRNA-depleted totRNA), double-stranded RNA (dsRNA), small RNA (sRNA), RNA from purified or partially purified viral particles (VANA) and poly(A) RNA (reviewed in Adams & Fox, 2016; Roossinck et al., 2015; Wu et al., 2015). The above listed viral nucleic acid preparation protocols differ in their efficiency, and each has its specific advantages and disadvantages (Roossinck et al., 2015). The main common disadvantage of some protocols is the bias in detecting a particular group of viruses. For example, an enrichment strategy for poly(A) RNA could miss viruses without polyA tail, while using RNA from viral particles is not sensitive enough for non-encapsidated viruses; furthermore, it requires sample-specific processing because it is unlikely that all viruses could be captured by a single viral particle purification protocol (Wu et al., 2015; Roossinck et al., 2015). The dsRNA enrichment strategy is mainly used for sequencing of ssRNA + and dsRNA +/- viruses; however, it is considered problematic for negative-stranded RNA (they do not accumulate large amounts of dsRNA during replication) and DNA viruses (Wu et al., 2015; Roossinck et al., 2015); nevertheless, both, RNA – and ssDNA + viruses have been detected using this protocol (Al Rwahnih et al., 2013; Gaafar et al., 2020).

In contrast, total RNA (totRNA), particularly its improved version consisting of ribosomal RNA-depleted total RNA (rRNA-depleted totRNA) (Adams & Fox, 2016) and small RNA (sRNA) seem to be more generically applicable to viruses with different genome types and replication strategies.

1.3 Towards On-Site High-Throughput Sequencing-Based Detection of Plant Viruses/Viroids

Since the introduction of HTS technology, different platforms have been developed: the Illumina technology “utilizes reversible florescent dideoxy terminators to sequence DNA clusters amplified on the surface of flow cells”; the Ion Torrent platform utilizes an electronic microchip to detect pH variation caused by the addition of bases to a DNA template; Pacific Bioscience released a platform for single molecule real time sequencing (SMRT). In the latter case, fluorescently labelled nucleotides are incorporated into the

DNA molecule in small wells and the incorporated fluorescence is measured (Adams and Fox, 2016). In parallel to this development, the cost efficiency and throughput of the methods has markedly improved, enabling the simultaneous sequencing of a high number of samples at affordable price. Illumina (https://www.illumina.com/) has become one of the most widely used platform as it provides the highest throughput, has the lowest error rate, and is the most cost effective (Villamor et al., 2019). However, such high throughput is not always the most convenient, e.g., in small laboratories that analyze small number of samples, as the cost efficiency becomes evident only when high numbers of samples are processed on a daily basis. Consequently, such laboratories often have to outsource sequencing of their sample to commercial service providers, which prolongs the turnaround time and limits the possibilities for quick results. The globalization of agriculture and international trade are generating an increased spread of plant viruses and viroids to new geographic regions with unexpected consequences for food production and natural ecosystems (Jones & Naidu, 2019). To combat the negative impact of such increased spread on crop production and food safety, rapid generic, and preferably onsite deployable plant virus/viroid detection technologies are needed. The HTS-based solution for fast, generic, and portable surveillance method became commercially available in 2014/2015, namely nanopore sequencing from Oxford Nanopore Technologies. The nanopore sequencing technology works by "monitoring changes of an electrical current as nucleic acids are passed through a protein nanopore. The resulting signal is then decoded to provide the specific DNA or RNA sequence" (https://nanoporetech.com/). Among others, the company developed the MinION sequencer, which can also be used in small laboratories, releasing the results near real time (Branton & Deamer, 2018). The technology is fast and affordable, enabling quick health status assessment of plant material. Since 2015, the technology has been constantly developing further in terms of hardware and software equipment and has been used for different surveillances, i.e., EBOLA (Quick et al., 2016), SARS-CoV-2 (Meredith et al., 2020) and plant virus detection (Bronzato Badial et al., 2018; Chalupowicz et al., 2019; Fellers et al., 2019; Filloux et al., 2018; Liefting et al., 2021). One of the main drawback of nanopore technology i.e. high error rate, which initially reached up to 15% (Ip et al., 2015; van Dijk et al., 2018), has been improved to around 6% (Delahaye & Nicolas, 2021). The nanopore technology however enables different library preparation and sample inputs, where different options can be divided according to the type of the input material (DNA, RNA), type of library preparation (PCR step, barcoding option, etc.) and the primary scope of the analysis (i.e., generic detection, targeted sequencing, amplicon sequencing, whole genome sequencing etc.), which supports the practicality and affordability of the technology.

1.4 Aims of the Dissertation

1.4.1 Comparing of two HTS protocols (sRNA and rRNA-depleted totRNA) for detection and discovery of plant viruses/viroids

For the detection of plant viruses and viroids by HTS different nucleic acid preparation protocols are used. Our main aim was to compare the performance of two most generic nucleic acid preparation protocols, sRNA and rRNA-depleted totRNA, for the detection and identification of known and unknown plant viruses and viroids in real diagnostic plant samples with either single or mixed infection. The samples selected for the comparison included plant viruses from different Baltimore classification groups and viroids from two different families. More specifically, we aimed to compare for each protocol i) the recovery

(i.e. identification and selection) of reads from viruses with different genome organization, ii) the recovery of reads from viroids from the two different viroid families, and iii) the performance of the HTS-based detection with each protocol by assessing detection efficiency through mapping and de novo assembly analysis of differently sized subsamples with decreasing numbers of nucleotides.

1.4.2 Utilizing HTS for detection and phylogenetic analysis of a virus, detected for the first time in tomato in Slovenia, followed by its biological characterization

The use of high throughput sequencing in plant viral diagnostic has enabled many unexpected or new virus discoveries. Our second aim was to use HTS to explore the etiology of uncharacterized symptoms in diagnostic plant tissue sample, including detection and genomic and phylogenetic characterization of a new strain of henbane mosaic virus and its biological characterization. Firstly, by using sRNA and rRNA-depleted totRNA HTS protocols, we assembled the complete genome of the virus. In order to perform phylogenetic analysis across all available henbane mosaic virus strains, four other strains were sequenced using sRNA HTS protocol and their complete genomes were assembled. Additionally, genome annotation of the new strain of henbane mosaic virus, detected and identified for the first time both in tomato and in Slovenia, was performed. Lastly, host range and biological analysis was accomplished using test plant bioassay.

1.4.3 Assessing the applicability of HTS sequencing using nanopore technology for plant viruses/viroids detection

Our third, and last, aim was to compare different library preparation protocols for plant viruses/viroids detection using nanopore technology (MinION sequencer, Oxford Nanopore Technologies). For this purpose, five plant samples infected with plant viruses/viroids that differ in their genome organization, concentration and mixed/single infection status, were selected. For all samples, we performed sequencing with Illumina technology (MiSeq sequencer) as the HTS gold standard, using rRNA-depleted totRNA. All samples were then sequenced with nanopore technology, with the MinION sequencer (Oxford Nanopore Technologies) using the direct RNA sequencing of totRNA. Moreover, for one of the samples, which contained five different plant viruses and a viroid, additional variations of library preparation for MinION sequencing were assessed: direct RNA sequencing of rRNA-depleted totRNA, cDNA-PCR sequencing of totRNA, and cDNA-PCR sequencing of rRNA-depleted totRNA. From the obtained HTS data, we then calculated and compared for each protocol i) the recovery of different viruses/viroid reads, and ii) the performance of the HTS-based detection with each protocol by assessing detection efficiency through mapping and de novo assembly analysis of different size subsamples with decreasing number of nucleotides.

1.5 Hypothesis

1. HTS will enable the detection of known and unknown plant viruses and viroids with both sequencing (sRNA and rRNA-depleted totRNA) protocols on the Illumina platform.
2. HTS will enable the identification of the viral agent behind a symptomatology of unknown etiology in a symptomatic tomato plant.
3. HTS will enable the assembly and annotation of the henbane mosaic virus for the first time, followed by its biological characterization.
4. Sequencing of different infected plant samples with nanopore technology can detect all plant viruses and viroids included in the study, in a reduced time and with potential for on-site deployment in comparison to Illumina technology

1.6 Publications Included and Candidate's Contributions

In the first paper (Next Generation Sequencing for Detection and Discovery of Plant Viruses and Viroids: Comparison of Two Approaches), two different HTS (in the manuscript, called "next generation sequencing") protocols were compared in terms of plant virus detection and discovery. The study was the first in depth protocol comparison and included plant samples naturally infected with viruses belonging to all different Baltimore classification groups and viroids belonging to both viroid families. The comparison included the detection and identification of known viruses, the recovery of virus/viroid reads, and the detection efficiency of new/unknown virus. The PhD candidate is the first author of the publication. She was involved in experimental design, she carried out the laboratory work (i.e., sample selection, RNA isolation and DNase treatment, rRNA-depleted totRNA Illumina library preparation), analyzed the data (sRNA detection pipeline, comparative bioinformatics analysis) and wrote the draft of the manuscript.

In the second paper (High-Throughput Sequencing Facilitates Characterization of a "Forgotten" Plant Virus: The Case of a Henbane Mosaic Virus Infecting Tomato), HTS was used for detection and identification of a new strain of henbane mosaic virus detected for the first time on a new host and new geographical location (a tomato species in Slovenia). The sequencing outputs obtained with both sRNA and rRNA-depleted totRNA enrichment approaches enabled first complete genome assembly of henbane mosaic virus. In the next step, four additional historical isolates of henbane mosaic virus were sequenced using the sRNA protocol. The complete genomes of all of them were assembled, and phylogenetic analysis was conducted. Finally, the host range analysis of a new henbane mosaic virus strain was performed. The PhD candidate is the first author of the publication. She was involved in the experimental design, performed most of the laboratory work (RNA isolation, PCR primer design for Sanger sequencing, HMV isolation on host plant, host range analysis, rRNA-depleted totRNA library preparation for Illumina sequencing) apart from the symptoms discovery, first screening and transmission electron microscopy. She also analyzed the data (sRNA detection pipeline, sRNA whole genome assembly and phylogenetic analysis) and wrote the draft of the manuscript.

In the last, third paper (Systematic comparison of nanopore and Illumina sequencing for the detection of plant viruses and viroids using total RNA sequencing approach), the comparison between two different HTS platforms i.e., nanopore and Illumina was performed for the first time with plant viruses in such an extensive format. A comparison was made between Illumina as the HTS gold standard and nanopore as a fast-developing

sequencing technology, offering rapid and possibly on-site (in small facilities) applications. For the comparisons, five plant samples were selected, containing eleven plant viruses belonging to different Baltimore classification groups. In addition, the samples were infected by two viroids, which were sequenced for the first time using nanopore technology. The bioinformatics analysis compared all different protocols in terms of suitability for plant viruses and viroids detection. The PhD candidate is the first author of the publication. She was involved in the experimental design, she carried out the laboratory work (sample selection, RNA isolation and DNase treatment, ribosomal RNA depletion, nanopore library preparation and sequencing using MinION), analyzed the data (nanopore detection pipeline, Illumina rRNA-depleted totRNA detection pipeline and comparative bioinformatics analysis, apart from optimizing parameters for de novo assembly of nanopore reads and creating script for calculating the cumulative yields of reads), and wrote the draft of the manuscript.

Chapter 2

Scientific Publications

2.1 Next Generation Sequencing for Detection and Discovery of Plant Viruses and Viroids: Comparison of Two Approaches

Anja Pecman, Denis Kutnjak, Ion Gutiérrez-Aguirre, Ian Adams, Adrian Fox, Neil Boonham and Maja Ravnikar

Frontiers in Microbiology, 2017, 8:1998. doi: 10.3389/fmicb.2017.01998

In this paper, we present a comparison of two HTS protocols, sequencing of sRNA and sequencing of rRNA-depleted totRNA. Both protocols differ in two successive steps within the HTS process: nucleic acid preparation and library preparation. The protocols were compared for the detection and identification of twelve plant viruses, which differ in their genome organization, and three viroids from both known families, using Illumina sequencing. The above mentioned two sequence protocols were selected and compared because theoretically they seem as the most generically applicable to viruses with different genome types and replication strategies. For the comparison, we selected nine samples of different plant species (*Solanum tuberosum, Solanum lycopersicum, Brassica oleracea, Nicotiana tabaccum, Pisum sativum*). The results showed that both protocols enabled the detection and identification of a wide array of known plant viruses/viroids in the analyzed samples. In general, the percentage of viral/viroid sequences was dependent on the genome organization of the virus/viroid and on the amount of viral reads in the sequencing output data. The results of our comparison showed that sRNA sequencing generated a higher fraction of viral/viroid sequences than rRNA-depleted totRNA for TASVd, ToCV, CLVd, TYLCV, PNYDV, PLMVd, and PVY, while rRNA-depleted totRNA protocol, generated more viral sequences for a novel cytorhabdovirus (CCyV1), PepMV (two isolates), CaMV, AMV, CSNV and TMV. In the cases of STV and ToMV, the percentage of virus sequences were extremely low (below 0.1%) regardless of the protocol. Subsampling analysis showed that for the read mapping approach 10 million of nucleotide would be enough to cover near complete viral genome of almost all investigated viruses/viroids using either of the protocol. However, in the cases of STV and ToMV, none of the viral genomes were completely covered by the reads even at the highest subsample size (50 million). In the assessment of the de novo assembly approach, rRNA-depleted totRNA protocol performed better than the sRNA protocol. For instance, the contigs assembled from rRNA-depleted totRNA data covered at least a fraction of the consensus genome also when the percentage of virus/viroid nucleotides was lower than 0.1% (i.e., ToMV and STV). In those cases, no viral contigs were assembled using sRNA datasets, probably due to a combination of low amount and small sizes of viral reads. A similar scenario was also observed in the case of a putative novel cytorhabovirus, for which very low recovery of viral reads in the sRNA dataset resulted in no assembled contigs corresponding to this virus. A putative new cytorhabdovirus was, however, identified from the rRNA-depleted totRNA dataset, whilst

the virus reads could only be found in the sRNA sequence data post-hoc (i.e. 0.05% of sRNA reads mapped to the genome assembled by rRNA-depleted totRNA sequencing). The obtained results significantly contributed to the optimization of the NGS (currently rephrased to HTS) technology for detection and identification of known and new plant viruses and viroids.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 October 2017
doi: 10.3389/fmicb.2017.01998

Next Generation Sequencing for Detection and Discovery of Plant Viruses and Viroids: Comparison of Two Approaches

Anja Pecman[1,2], Denis Kutnjak[1]*, Ion Gutiérrez-Aguirre[1], Ian Adams[3], Adrian Fox[3], Neil Boonham[3,4] and Maja Ravnikar[1]*

[1] Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia, [2] Jožef Stefan International Postgraduate School, Ljubljana, Slovenia, [3] Fera Science Ltd., York, United Kingdom, [4] Institute for Agri-Food Research and Innovation, Newcastle University, Newcastle upon Tyne, United Kingdom

Next generation sequencing (NGS) technologies are becoming routinely employed in different fields of virus research. Different sequencing platforms and sample preparation approaches, in the laboratories worldwide, contributed to a revolution in detection and discovery of plant viruses and viroids. In this work, we are presenting the comparison of two RNA sequence inputs (small RNAs vs. ribosomal RNA depleted total RNA) for the detection of plant viruses by Illumina sequencing. This comparison includes several viruses, which differ in genome organization and viroids from both known families. The results demonstrate the ability for detection and identification of a wide array of known plant viruses/viroids in the tested samples by both approaches. In general, yield of viral sequences was dependent on viral genome organization and the amount of viral reads in the data. A putative novel *Cytorhabdovirus,* discovered in this study, was only detected by analysing the data generated from ribosomal RNA depleted total RNA and not from the small RNA dataset, due to the low number of short reads in the latter. On the other hand, for the viruses/viroids under study, the results showed higher yields of viral sequences in small RNA pool for viroids and viruses with no RNA replicative intermediates (single stranded DNA viruses).

Keywords: next generation sequencing, small RNA, ribosomal RNA depleted total RNA, detection, plant viruses, plant viroids

OPEN ACCESS

Edited by:
Guenther Witzany,
Telos - Philosophische Praxis, Austria

Reviewed by:
Claudio Ratti,
Università di Bologna, Italy
Carmen Hernandez,
Instituto de Biología Molecular y Celular de Plantas (CSIC), Spain

****Correspondence:***
Anja Pecman
anja.pecman@nib.si
Denis Kutnjak
denis.kutnjak@nib.si

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

Received: *17 July 2017*
Accepted: *28 September 2017*
Published: *13 October 2017*

Citation:
Pecman A, Kutnjak D, Gutiérrez-Aguirre I, Adams I, Fox A, Boonham N and Ravnikar M (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

INTRODUCTION

Plant viruses and viroids are important plant pathogens, causing economic losses by reducing crop quality and quantity all over the world (Loebenstein, 2008; Soliman et al., 2012). Thus, their reliable detection is of a crucial importance for plant protection. Classical methods in plant virus diagnostics can be roughly divided into specific (serological/molecular tests) and non-specific (indicator test plants, electron microscopy) approaches. Specific methods are usually targeted to one or a few viral species and require *a priori* knowledge of the pathogens being tested, whilst non-specific approaches do not require specific knowledge of the pathogens, however, frequently only classify viruses at a genus level based on the shared physical/biological characters. Discovery of new viruses/viroids and new hosts has increased rapidly after the introduction of next generation

sequencing (NGS). NGS technologies allow a generic approach (non-specific method) to virus identification that does not require any prior knowledge on the targeted pathogens but can deliver a species/strain specific result (Adams and Fox, 2016.) It was first employed for plant virus detection in 2009 (Adams et al., 2009; Al Rwahnih et al., 2009; Kreuze et al., 2009). Since 2009, different sample preparation methods have been developed, relying on different nucleic acid inputs, most commonly: total RNA (totRNA); ribosomal RNA depleted total RNA (rRNA depleted totRNA); double stranded RNA (dsRNA); virus derived small interfering RNA (sRNA); RNA from purified or partially purified viral particles; polyadenylated RNA (poly(A) RNA); and RNA after subtractive hybridization with healthy plant RNA. Applications of different sample preparation methods are reviewed in Roossinck et al. (2015); Wu et al. (2015), and Adams and Fox (2016). Viruses have diverse genome organizations and use different replication strategies. Based on these two characteristics they can be classified into 7 groups (the Baltimore classification): double stranded DNA (Group I, dsDNA +/−), single stranded DNA (Group II, ssDNA +,) double stranded RNA (Group III, dsRNA +/−), positive sense single stranded RNA (Group IV, ssRNA +), negative sense single stranded RNA (Group V, ssRNA −) viruses, positive sense single stranded RNA viruses that replicate through a DNA intermediate (Group VI, ssRNA-RT +), and double stranded DNA viruses that replicate through a RNA intermediate (Group VII, dsDNA-RT +/−( )Baltimore, 1971). Viroids are classified into two families: members of *Avsunviroidae* family replicate in chloroplast, whereas members of *Pospiviroidae* family replicate in nucleus (Flores et al., 2014). Considering the diversity of viruses and viroids, with different genome organizations in mind, it is conceivable that using different nucleic acid inputs for NGS could affect their overall detection.

Sample preparation methods (i.e., different nucleic acid inputs), used before NGS, can differ in their efficiency and can have specific advantages and disadvantages. For example, subtractive hybridization of the host plant nucleic acids, using tomato (*Solanum lycopersicum*) and *Pepino mosaic virus* (PepMV, RNA +, *Potexvirus, Alphaflexiviridae*) as a model system, resulted in three times more PepMV sequences in subtracted sample (Adams et al., 2009), but as it is a time consuming procedure, which requires a healthy plant of the same species as the sample to be tested (Adams and Fox, 2016), subtractive hybridization is not well suited in a high-throughput diagnostic settings. Some sample preparation methods may cause bias in the detection of a particular group of viruses. Sequencing of dsRNA was mainly used for detection of RNA + and RNA +/− viruses, since RNA—and DNA viruses could be missed (Roossinck et al., 2015) using this approach; nevertheless, a new geminivirus (DNA +) was identified using dsRNA sequencing (Al Rwahnih et al., 2013). RNA isolated from purified viral particles has been successfully used for sequencing different viruses (reviewed in Roossinck et al., 2015; Wu et al., 2015.) A comparison between deep sequencing of sRNAs and RNA isolated from viral particles showed higher efficiency of the latter for the reconstruction of complete consensus *Potato virus Y* (RNA +, *Potyvirus, Potyviridae*) genomes (Kutnjak et al., 2015). However, virus purification is not applicable for un-encapsidated viruses and requires sample specific processing since it is unlikely that all viruses could be captured by a single protocol for viral particles purification (Roossinck et al., 2015; Wu et al., 2015). Poly(A) RNA based enrichment strategy has been also used for both RNA and DNA viruses but it is not applicable for the detection of viruses without a poly(A) tail (Wu et al., 2015). Data from sequencing poly(A) RNA showed a lower degree of virus genome coverage in comparison to saturated genome coverage reached with sRNA data for *Grapevine leafroll-associated virus 3* (RNA +, *Ampelovirus, Closteroviridae*), yet a comparison between poly(a)RNA and sRNA data for *Hop stunt viroid*(,*Pospiviroidae* family) showed comparable outcomes (high genome coverage) for both approaches (Visser et al., 2016.)

In this study, we focused the comparison (with the detection and identification of plant viruses and viroids in mind) on the two types of RNA inputs: sequencing of sRNA and sequencing of rRNA depleted totRNA. Those two approaches seem to be the most generically applicable to viruses with different genome types and replication strategies and could be relatively easily integrated in workflows of diagnostic labs.

Sequencing and assembly of viral sRNA (Kreuze et al., 2009) has been successfully used for detection and identification of several plant viruses and viroids and their complete genome assembly (reviewed in Boonham et al., 2014; Kreuze, 2014). It has been speculated that this approach could be problematic if used to detect viruses that either do not trigger silencing responses or that express silencing suppressors (Roossinck et al., 2015). Also, de novo assembly of longer viral contigs could be complicated due to short reads lengths (Boonham et al., 2014; Roossinck et al., 2015; Adams and Fox, 2016). On the other hand, the approach is very generic, using the same protocol of sample preparation for many different plant species and doesn't require high quality of RNA input (Kutnjak et al., 2017.)

Sequencing of plant viruses using total RNA as an input was first described by Adams et al. (2009) and Al Rwahnih et al. (2009), followed by several successful studies (reviewed in Boonham et al., 2014). It is also a very generic approach, however, a potential shortcoming of that method can be the low viral RNA titer within the background plant RNA. To overcome this, removal of the highly abundant plant ribosomal RNA from the total RNA pool (rRNA depleted tot RNA) has been explored, which can results in a 10-fold enrichment of viral RNA (Adams and Fox, 2016.)

Recent comparison (Visser et al., 2016) of sRNA and rRNA depleted totRNA for *Citrus tristeza virus* (RNA +, *Closterovirus, Closteroviridae*) and *Citrus dwarfing viroid* (*Pospiviroidae* family) implied a preferential use of rRNA depleted totRNA for *de novo* assembly of viral genome sequences from NGS data. No wider comparison of these two approaches (including viruses with different genome characteristics) has been reported. With this in mind, our aim was to compare the two approaches, including plant viruses with different genome structures and replication strategies (belonging to different Baltimore classification groups) and viroids from different families into comparison. The aims were to compare the two approaches in terms of: (1) known virus detection and identification (2) recovery of virus/viroid reads

and (3) effectiveness of detection of new/unknown viruses by reconstruction of longer viral contigs by *de novo* assembly and read mapping analysis approaches.

MATERIALS AND METHODS

Description of Samples

Nine virus-infected plant samples were included in this study. The selection included samples of different plant species, infected with a range of plant viruses in single or mixed infections with at least one representative from each group of the Baltimore viral classification containing plant viruses, and viroids from both families (**Table 1**.)

Sample Preparation and Sequencing

Total RNA was isolated from plant samples using TRIzol reagent (Life technologies, USA) following the manufacturer's instructions. Isolated total RNA was then divided in half for comparative purposes. One half was sent to Seqmatic LLC (USA) for sRNA library preparation (TailorMix miRNA Sample Preparation Kit V2, SeqMatic LLC, USA) and sequencing. The samples were multiplexed in one lane of a HiSeq 2500 (Illumina, USA) in 1 × 50 bp mode. The remaining total RNA was further purified using an RNeasy protocol including DNase treatment following the manufacturer's protocols (RNA Cleanup protocol; RNeasy Mini Kit; Qiagen, Netherlands). Ribosomal RNA was depleted from the purified total RNA and sequencing libraries were prepared using the ScriptSeq™ Complete Kit (plant leaf) (Illumina, USA). The libraries were sequenced using MiSeq (Illumina, USA) in 2 × 300 bp (V3) mode. Number and average length of sequencing reads for every sample sequenced by both approaches are in Supplementary Table 7.

Detection of Viruses in NGS Data

Reads obtained by both sequencing procedures were trimmed, filtered and further analyzed to confirm the presence of viruses and viroids. Bioinformatics pipelines used for virus detection from NGS data are detaily described in Supplementary Data 1.1. In both cases, the presence of suspected viral sequences was confirmed by mapping the reads to the complete viral genome sequences of the most similar viral isolates from the NCBI GenBank database, followed by visual inspection of individual mappings.

Confirmatory Testing

The presence of virus in each case was also confirmed by using ELISA, RT-PCR, and RT-qPCR methods (**Table 1**). ELISA was performed using polystyrene microtiter plates (nunc-Immuno™, Sigma-Aldrich Inc., USA) and kits containing virus specific reagents as follows, AMV: Cat No. 07001S (Loewe Biochemica GmbH, Germany), CaMV: Cat No. 07086 (Loewe Biochemica GmbH, Germany), PVY: Cat No. 1105 (Bioreba AG, Switzerland) and TYLCV: Cat. No. 1072 (Neogen Europe Ltd., UK). The assays were performed following the manufacturer's instructions. In each case a negative control corresponding to the same species as the test sample was used. The result was considered positive when the optical density (OD) A_{405} value after 2 h for a given sample was greater than 2× the mean OD value of the corresponding negative control. For reverse transcription quantitative PCR (RT-qPCR) and reverse transcription conventional PCR (RT-PCR), total RNA was extracted from fresh or lyophilized plant material using the RNeasy Plant Mini Kit (Qiagen), following the manufacturer instructions. RT-qPCR was performed using published methods for PepMV (Gutiérrez-Aguirre et al., 2009) and for ToMV (Boben et al., 2007). Conventional RT-PCR was performed for PNYDV (Gaafar and Ziebell, 2016), STV (Sabanadzovic et al., 2009), ToCV (Dovas et al., 2002), TMV (Kumar et al., 2011), PLMVd (Loreti et al., 1999) TASVd and CLVd (Verhoeven et al., 2004). PCR primers designed specifically to confirm the presence of novel CCyV1 were as follows: CCyV1-fw (5′-GTCTCTCTTGCGTTGAGCCA-3′) and CCyV1-rev (5′-GGTTGCGGATAGCTCTTCCT-3′). All the amplicons obtained by RT-PCR were purified and sent for Sanger sequencing (GATC Biotech AG, Germany). The Sanger sequences were aligned against the genomes of detected viral species and their identity was confirmed in all of the cases.

Construction of Consensus Viral/Viroid Genome Sequences

For every identified virus/viroid the consensus viral genomes were extracted from the sRNA read mappings (see section Detection of Viruses in NGS Data) to obtain a corrected consensus genome. Validation of each corrected consensus genome was performed by mapping the *de-novo* generated contigs obtained by both NGS approaches to corresponding corrected consensus genome. Both mapping results were visually inspected for possible differences between the *de-novo* contigs and corrected consensus genome sequence. Observed conflicts were further investigated by inspecting the read mapping results. Finally, few of the observed differences were explained as polymorphisms in viral populations. In sample III, two divergent strains (80% nucleotide identity) of PepMV were detected (PepMV-EU and PepMV-CH2). In this case, the complete genome sequences of the two most similar isolates from NCBI GenBank were used in subsequent comparisons (KF718832.1 and JX866666.1), without the corrections after reads and contigs mapping as described previously.

Comparison of sRNA and rRNA Depleted totRNA Inputs

For comparisons, all raw reads were trimmed and filtered in CLC Genomic Workbench 9 (Qiagen). For rRNA depleted totRNA datasets, reads shorter than 100 nucleotides were discarded. Then, reads were trimmed using quality scores, setting the limit to 0.05 (see CLC Genomics Workbench User Manual, Chapter 23, for explanation). For sRNA reads, first, adaptor trimming was performed, then reads shorter than 20 and longer than 24 nucleotides were discarded.

First, the viral fraction of the total nucleotides sequenced (from now on called percentage of virus/viroid nucleotides) in each of the datasets for each of the detected viruses was calculated by mapping the trimmed and filtered reads (of the corresponding dataset) to the consensus viral/viroid genomes generated in the

TABLE 1 | Samples included in the comparison with corresponding results from: NGS (viruses/viroids listed in the table were detected in corresponding samples by NGS) and other diagnostic methods (ELISA, RT-PCR and RT-qPCR).

Sample number	Virus, genus, family	Baltimore classification	Genome organization	Abbreviations	Host	Initial detection with NGS		Results of confirmatory testing	NCBI GenBank accession number	NCBI SRA accession number (sRNA/rRNA depleted totRNA)
						sRNA	rRNA depleted totRNA			
I	*Potato virus Y, Potyvirus, Potyviridae	Group IV (ssRNA +)	Linear	PVY	*Solanum tuberosum*	+	+	+[a]	KY810782	SRR5377154/SRR5377146
II	*Cauliflower mosaic virus, Caulimovirus, Caulimoviridae;	Group VII (dsDNA-RT +/–)	Circular	CaMV	*Brassica oleracea*	+	+	+[a]	KY810770	SRR5377153/SRR5377145
	Novel cabbage cytorhabdovirus 1, Cytorhabdovirus, Rhabdoviridae	Group V (ssRNA –)	Linear	Novel CCyV1	*Brassica oleracea*	–	+	+[b]	KY810772	
III	*Tomato Yellow Leaf Curl Virus, Begomovirus, Geminiviridae;	Group II (ssDNA +)	Circular	TYLCV	*Solanum lycopersicum*	+	+	+[a]	KY810789	SRR5377152/SRR5377144
	Tomato chlorosis virus, Crinivirus, Closterovirdae;	Group IV (ssRNA +)	Linear, segmented	ToCV	*Solanum lycopersicum*	+	+	+[b]	KY810786 KY810787	
	Pepino mosaic virus, Potexvirus, Alphaflexiviridae;	Group IV (ssRNA +)	Linear	PepMV	*Solanum lycopersicum*	+	+	+[c]		
	Tomato mosaic virus, Tobamovirus, Virgaviridae;	Group IV (ssRNA +)	Linear	ToMV	*Solanum lycopersicum*	+	+	+[c]	KY810788	
	Southern tomato virus, Amalgavirus, Amalgaviridae;	Group III (dsRNA +/–)	Linear	STV	*Solanum lycopersicum*	+	+	+[b]	KY810783	
	Columnea latent viroid, Pospiviroid, Pospiviroidae	viroid	Circular	CLVd	*Solanum lycopersicum*	+	+	+[b]	KY810771	
IV	*Alfalfa mosaic virus, Alfamovirus, Bromoviridae	Group IV (ssRNA +)	Linear, segmented	AMV	*Nicotiana tabacum*	+	+	+[a]	KY810767 KY810768 KY810769	SRR5377151/SRR5377143
V	*Pea necrotic yellow dwarf virus, Nanovirus, Nanoviridae	Group II (ssDNA +)	Circular, segmented	PNYDV	*Pisum sativum*	+	+	+[b]	KY810774 KY810775 KY810776 KY810777 KY810778 KY810779 KY810780 KY810781	SRR5377150/SRR5377142
VI	*Tobacco mosaic virus, Tobamovirus, Virgaviridae	Group IV (ssRNA +)	Linear	TMV	*Nicotiana* sp.	+	+	+[b]	KY810785	SRR5377149/SRR5377141
VII	*Peach latent mosaic viroid, Pelamoviroid, Avsunviroidae	viroid	Circular	PLMVd	*Prunus* sp.	+	+	+[b]	KY810773	SRR5377148/SRR5377140
VIII	*Tomato apical stunt viroid, Pospiviroid, Pospiviroidae	viroid	Circular	TASVd	*Solanum lycopersicum*	+	+	+[b]	KY810784	SRR5377147/SRR5377139
IX	*Chrysanthemum stem necrosis virus, Tospovirus, Tospoviridae	Group V (ssRNA –)	Linear, segmented	CSNV	*Nicotiana benthamiana*	+	+	+[c]	MF093683 MF093684 MF093685	SRR5630913/SRR5630912

*Taxonomic classification, Baltimore classification and genome organization of detected viruses are given in separate columns. Host plant information is given in the separate column. NA, not applicable; +, detected; –, not detected; *, viruses/viroids which were known to be present in the sample before NGS analysis.*

[a] Confirmatory testing has been done using ELISA assay.

[b] Confirmatory testing has been done using RT-PCR assay.

[c] Confirmatory testing has been done using RT-qPCR assay.

previous step. Mapping parameters are listed in Supplementary Tables 1, 4.

To further compare the effectiveness of both approaches for detection and discovery of selected viruses, we then performed a normalization by subsampling the data from each sample (for both sRNA and rRNA depleted totRNA) to the same number of nucleotides. Random subsampling was performed to different subsample sizes: 1, 10, 30, and 50 million nucleotides. This was repeated ten times for each sample/size combination, yielding in total 360 datasets (9 samples × 4 subsample sizes × 10 replicates of subsampling). For those, the following analyses were implemented: (1) reads were mapped to the corresponding consensus viral/viroid genomes and the fraction of viral/viroid genome covered by reads (from now on: genome coverage (reads)) and the average depth of sequencing (number of times a nucleotide in a reference is covered by reads averaged for the complete genome) were calculated; (2) *de novo* assembly of reads was performed using CLC Genomics Workbench 9, followed by mapping the resulting contigs to the corresponding consensus viral/viroid genomes and calculation of the fraction of viral/viroid genome covered by the *de-novo* contigs (from now on: genome coverage (contigs)). Results of these comparisons are jointly shown in **Figure 2** and visualized as dots connected with solid lane (representing rRNA depleted totRNA results) and triangles connected with dashed lines (representing sRNA results). The mapping and *de novo* assembly parameters are listed in Supplementary Tables 1–4.

RESULTS

Sample Characterization

Twelve different viruses (among those, one viral species with two divergent strains) and three viroid species were detected using NGS in the nine samples included in the analysis (**Table 1**.) Nine were known to be present in the samples before the NGS analysis (marked with * in **Table 1**), whilst six virus/viroid species were detected using NGS during the study and their presence was confirmed as described in section Materials and Methods (**Table 1**). Both methods revealed the presence of 14 viral/viroid species whilst 1 virus (a putative novel viral species from the genus *Cytorhabdovirus*: CCyV1) could only be detected using the rRNA depleted totRNA approach. Seven samples (I, IV-IX) contained single viral/viroid infections, one sample (II) was infected with two viruses. Sample III was infected with five viruses and one viroid. All of the viruses and viroids detected and included in the study are listed in the **Table 1**.

Percentage of Virus/Viroid Reads Differs For Different Viruses

First, we estimated what percentage of the total sequenced nucleotides were viral/viroid nucleotides (of the complete cleaned NGS datasets) for different viral species for each of the two approaches. The percentage of viral/viroid nucleotides was in some cases higher using sRNA input and in other cases higher using rRNA depleted totRNA input (**Figure 1**.) Specifically, the results showed that for 6 viruses/viroids the sRNA approach generated a higher fraction of viral/viroid sequences: TASVd, ToCV, CLVd, TYLCV, PNYDV, PLMVd, and PVY (**Figure 1**: the viruses located below the diagonal line). For the sRNA approach, the highest percentage of viral sequences was observed for PVY (50%, **Figure 2A**). The rRNA depleted totRNA approach generated more viral sequences for 6 viruses: a novel *Cytorhabdovirus*, PepMV (two isolates), CaMV, AMV, CSNV and TMV (**Figure 1**, the viruses located above the diagonal line), with the highest viral sequences fractions for TMV (83,%) AMV (56%), CSNV (48%), and CaMV (48%) (**Figures 1**, **2A**). In two cases (STV and ToMV), the percentage of virus sequences were extremely low regardless of the RNA inputs (**Figures 1**, **2A**.)

Comparison on Normalized Subsamples

To be able to compare the two approaches in a greater detail, we subsampled all of the datasets to the same number of nucleotides. Ten replicates of four different sizes of subsamples (1, 10, 30, and 50 million nucleotides) were generated for each dataset to enable an assessment of the impact of data rarefaction and data variability on the performance of tested parameters.

First, average depth was evaluated (**Figure 2B**). In all cases, average depth increased with the increase of subsample sizes and followed the patterns observed when comparing the fractions of viral sequences nucleotides recovered by the two approaches. Results from 10 independent replicates for each subsample size showed a low variability for PVY, ToCV, PepMV, AMV, TMV, CSNV, and CaMV. Variability between the subsamples in average depth was higher for all other viruses/viroids (Supplementary Table 5).

Secondly, we investigated how effectively the reads cover the genomes of different viruses by calculating the fraction of the genome covered by reads [genome coverage (reads)] (**Figure 2C**.) Results of the analysis showed low variability between replicates of subsamples, except when mapping rRNA depleted totRNA reads to ToMV, STV, TYLCV, TASVd, and PLMVd where variation was very high (Supplementary Table 5, **Figure 2C**). In all cases, as expected, better genome coverage was achieved with the increasing subsample sizes. For the sRNA approach, complete genomes (100%) were covered for majority of the viruses/viroids at subsample size of 30 million nucleotides. The exceptions were ToMV, STV and the putative novel *Cytorhabdovirus*. For those, even at 50 million nucleotides, genome coverage was 70% or less.

For the rRNA depleted totRNA approach, for half of the viruses (PVY, PepMV, AMV, TMV, novel CCyV1, CSNV, CaMV, CLVd, and TASVd) complete genomes were covered at 10 million nucleotides. However, for some viruses/viroids (ToCV, TYLCV, PNYDV, and PLMVd) relatively low genome coverage was achieved at smaller subsample sizes (1 and 10 million nts) and even at the largest subsample size (50 million nts) the coverage did not reach 100% (**Figure 2C**). The genomes of ToMV and STV, for which very low numbers of reads were recovered (**Figures 1**, **2A**), were poorly covered even at high subsampling depths, for example, even with 50 million nucleotides, coverage remained below 50% (**Figure 2C**.)

Reads from normalized datasets were *de novo* assembled into contigs, which were then mapped to the corresponding

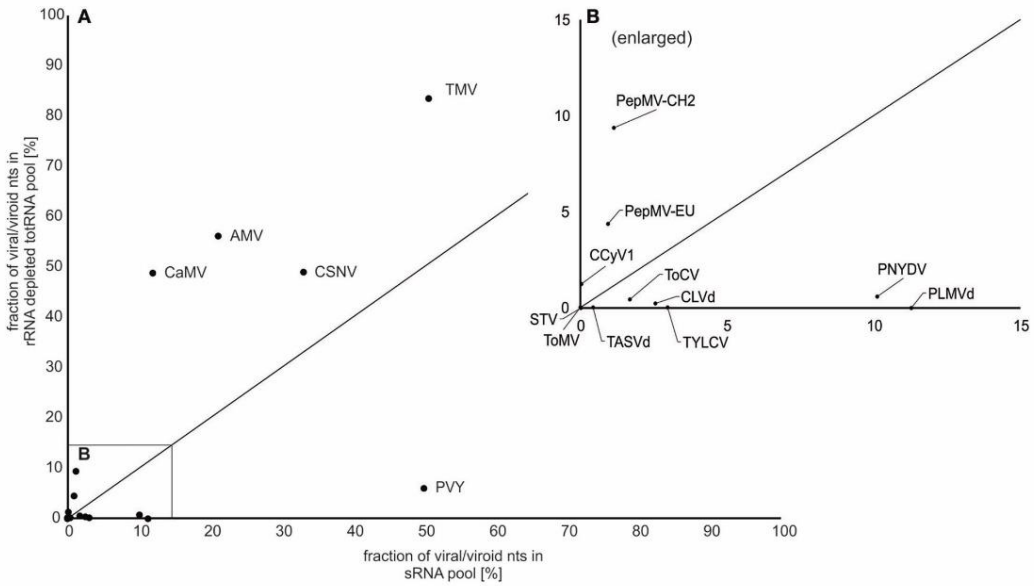

FIGURE 1 | Fraction of virus/viroid nucleotides (nt) in NGS datasets for small RNA and rRNA depleted totRNA approaches. **(A)** The plots show the percentage of nucleotides (fraction of total) that mapped to the corresponding consensus viral/viroid genome for both sRNA (x-axis) and rRNA depleted totRNA (y-axis) inputs. Dots represent the value for each virus/viroid (also listed in **Figure 2A**), with viral/viroid species abbreviations (see **Table 1**) next to the dots. The diagonal solid line represents theoretical scenario in which percentage of virus/viroid nucleotides would be equivalent for both approaches; dots above the line represent cases for which fraction of virus/viroid nucleotides was higher using rRNA depleted totRNA input, dots below the line represent cases for which fraction of viruses/viroids nucleotides was higher using sRNA input. **(B)** Enlarged part of **(A)** (0–15%).

consensus viral genomes in order to calculate the fraction of the viral genomes covered by the *de novo* assembled contigs [genome coverage (contigs)] (**Figure 2D**). The analysis of subsample replicates showed in general lower variability for sRNA datasets than rRNA depleted totRNA datasets (Supplementary Table 5). For the majority of the viruses, the coverage by contigs increased with subsample size, however, conversely, in several cases, it dropped at larger subsample sizes, i.e., TMV and PLMVd for sRNA and PepMV, CSNV, CaMV and CLVd for rRNA depleted totRNA approach (**Figure 2D**). Contigs, assembled *de novo* from rRNA depleted totRNA datasets covered higher fractions of viral genomes for almost all viruses at all subsample sizes (coverage reached 95% at 10 million nts for majority of viruses), in comparison to sRNA derived contigs (95% coverage at 10 million nts was achieved only for PVY, TMV, and CLVd). Two exceptions to this observation were TYLCV and CLVd, for which sRNA derived *de novo* contigs cover higher genome fraction than rRNA depleted totRNA contigs, for all subsample sizes.

The comparison of the *de novo* assemblies for STV and ToMV revealed that when very low numbers of viral reads are recovered, the rRNA depleted totRNA approach is more effective, since in the case of the sRNA approach, no corresponding viral contigs were generated (**Figure 2D**). A similar scenario was observed also for the putative novel *Cytrohabovirus*, where very low recovery of viral reads in the sRNA dataset resulted in no assembled contigs corresponding to this virus (**Figure 2D**.)

DISCUSSION

In this study we compared the effectiveness of two NGS approaches that have been widely adopted for plant virus detection: sRNA deep sequencing and deep sequencing of rRNA depleted totRNA. When comparing the amount of virus/viroid reads recovered by one or the other approach, we observed different results for different viruses/viroids: in some cases, more viral/viroid nucleotides were recovered using sRNA and in other by rRNA depleted totRNA sequencing.

Detailed inspection of the results of the read mapping suggested higher recovery of virus reads for ssDNA viruses and viroids when using sRNA approach than when using rRNA depleted totRNA approach. For viroids, this could be the consequence of induced RNA silencing (Itaya et al., 2001; Papaefthimiou et al., 2001; Martínez de Alba et al., 2002) and, at the same time, the absence of the messenger RNA production, because, in the case of viroids, "long" RNAs are generated solely for the purpose of replication. Similarly, in the case of viruses with a circular ssDNA genome organization, a smaller fraction of viral nucleotides was recovered using rRNA depleted totRNA. In contrast with viruses with RNA genomes, for ssDNA viruses, RNA molecules are generated only during the transcription step, as messenger RNAs, which could be the reason for the lower recovery of viral nucleotides in this pool. Moreover, small RNAs could be amplified by the action of RNA-dependent RNA polymerase 6 (Borges and Martienssen, 2015)

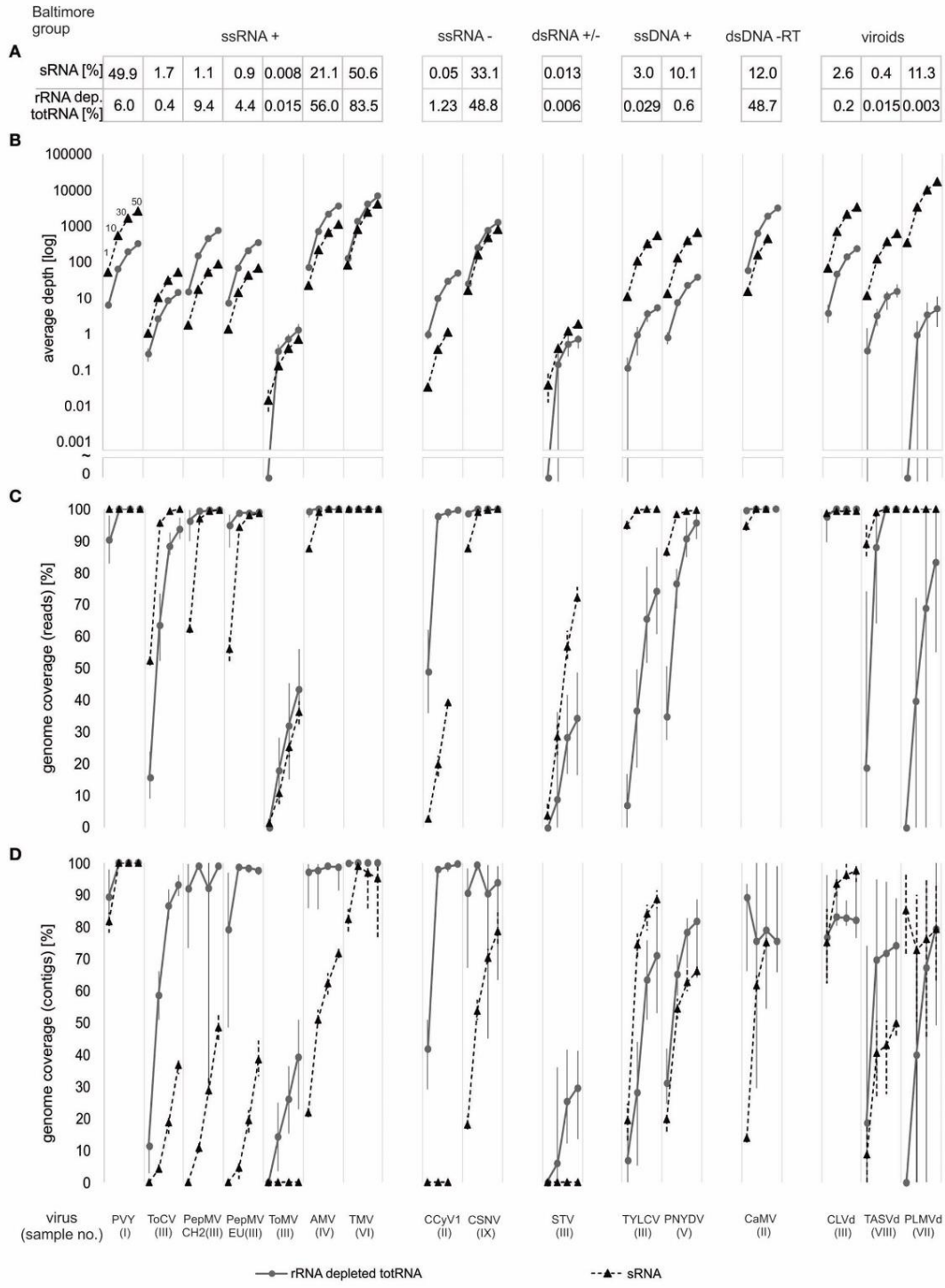

FIGURE 2 | Comparison of sRNA and rRNA depleted totRNA approaches using data size-normalized subsamples. Results for each virus included in the analysis are shown along the x-axis and are grouped according to Baltimore classification **(A)** Fraction (%) of virus nucleotides in trimmed and filtered complete NGS datasets. **(B)** Average depth (number of reads covering a position in a viral genome, averaged over the complete genome sequence) at different subsample sizes. Symbol ~

(Continued)

FIGURE 2 | Continued
indicate interruption of log scale, below, 0 values are plotted. **(C)** Fraction of viral genome (in %) covered by reads [genome coverage (reads)] at different subsample sizes. **(D)** Fraction of viral genome (in %) covered by contigs [genome coverage (contigs)] at different subsample sizes. For **(B–D)** Dots/triangles represent the mean, whereas vertical bars connect minimum and maximum results of 10 repeated analyses. Four different subsample sizes were used (1, 10, 30, and 50 million nts) and are designated in the first column, other columns follow the same logic. Triangles and dashed lines represent results for sRNA approach, dots and solid lines represent results for rRNA depleted totRNA. In some cases data points are missing, since the size of the complete dataset was smaller than the largest subsample.

during the production of secondary sRNAs. The exception among the DNA viruses in this study was CaMV (DNA-RT), for which a higher fraction of virus nucleotides was recovered by sequencing rRNA depleted totRNA. The CaMV dsDNA genome is replicated through an RNA intermediary, in addition to producing messenger RNAs through transcription (Hull, 2014,) which could explain a larger proportion of viral nucleotides in this pool.

All linear viruses in our infected plant samples had a ssRNA genome organization and synthesize different types of RNA throughout their replication cycle. For most of these viruses, sequencing rRNA depleted totRNA resulted in a larger proportion of reads mapping to the viral genomes (**Figure 1**) compared with sRNA. However, a few exceptions were observed, PVY being the most notable with many more viral reads being present in the sRNA dataset. The high abundance of virus derived sRNA has already been reported for PVY (Kutnjak et al., 2015) and other potyviruses (Kreuze et al., 2009) even though they encode strong RNA silencing suppressors (Yelina et al., 2002; Ivanov et al., 2016.)

In general, when read mapping was performed, 10 million nucleotides was sufficient to cover complete viral genomes using any of the two approaches (**Figure 2C**). However, in some cases (STV and ToMV in sample III) very low numbers of viral reads were recovered (by both approaches), which negatively affected all the evaluated parameters. For those two cases, the percentage of virus reads (for both approaches) was lower than 0.1%, and the average read depth remained lower than 10×, and none of the viral genomes were completely covered by the reads even at the highest subsample size (50 million) (**Figure 2C**.)

When comparing *de novo* assembly of sequencing reads, the rRNA depleted totRNA approach was generally more efficient than sRNA approach; this was demonstrated in higher proportion of viral genomes covered by *de novo* generated contigs from rRNA depleted totRNA datasets. The contigs assembled from rRNA depleted totRNA data covered at least a fraction of the consensus genome even in cases where the percentage of virus/viroid reads was lower than 0.1% and average depth lower than 10 (i.e., ToMV and STV) (**Figure 2D**). In those cases, no viral contigs were assembled using sRNA datasets, probably due to a combination of low amount and small sizes of viral reads. Poorer coverage of viral genomes by sRNA derived *de novo* contigs is likely related to the more difficult assembly of very short sRNA reads into longer contigs, which has been observed previously (Kutnjak et al., 2015; Visser et al., 2016.)

In some cases (PepMV, TMV, CSNV CaMV, CLVd, and PLMVd) smaller genome fractions are covered by contigs, when larger data sets are used for the assembly (corresponding to average depths > 100). This has been observed previously and is an artifact of the assembly algorithms (see CLC Analyses-related questions, 2017), which are not optimized for very high sequencing depths. After mapping reads or contigs to evaluate average depth and genome coverage (reads/contigs) we observed also the trend in generating higher or lower variability within 10 repeats. Unrepeatable random subsampling occurred when analysing smaller datasets and/or lower viral/viroid nucleotide proportion within the datasets, since all samples with this two features had greater variability.

The study has highlighted some points of difference between the compared approaches that may help to inform the choice of approach based on the purpose of the sequencing. This could be (i) screening against a list of known target organisms (e.g., at the import/export) and (ii) identification of the (possibly yet unknown) causal agent of the disease. Considering (i) screening against a list of known targets, this would be most cost effectively achieved using a method that maximizes the amount of viral sequences compared with host sequences. This study showed (**Figures 1**, **2A**) that the performance of the two compared approaches is very virus dependent. Broadly, sRNA performed better for circular ssDNA viruses and viroids, whilst rRNA depleted total RNA performed better for most of the tested linear RNA viruses with a notable exception (PVY). If considering (ii) sequencing for novel virus discovery, long contigs would provide the greatest chance of detecting very dissimilar sequences by comparing predicted amino-acid sequence from virus ORFs (e.g., with the use of BLASTx analysis or hidden Markov model based protein domain searches). The data shows that rRNA depleted total RNA generated longer contigs (which covered greater fractions of viral genomes) for most of the investigated viruses (**Figure 2D**). As the most prominent example, an important difference between the compared approaches was observed on a case of a previously un-described *Cytorhabdovirus*, which was identified from the rRNA depleted total RNA following *de novo* assembly and BLASTx analysis, whilst the virus reads could only be found in the sRNA sequence data *post-hoc* (de novo assembly of sRNA reads did not generate any matching contigs).

The results of the comparison between the two NGS approaches highlight some trends that may guide diagnostic laboratories in the selection of a method appropriate for a specific application. However, whichever method is selected it is important to be aware of the limitations, some of which are detailed in this study, and follow up putative identification using an appropriate method. The recently published framework for handling novel plant viruses detected using NGS provides guidelines for achieving this (Massart et al., 2017.)

In order to examine the potential costs of each method on commonly used Illumina sequencing platforms (HiSeq/sRNA and MiSeq/rRNA depleted totRNA) staff time used and reagent

costs (in GBP) were calculated using list prices (Illumina) obtained on 1st March 2017. In general, both approaches generate more than sufficient amount of data than required to identify all of the viruses if mapping is used (50 million nts; **Figure 2**). HiSeq/sRNA sample will cost £138 and MiSeq/rRNA depleted totRNA sample will cost £159 if 24 samples (reasonable diagnostic throughput) are run per lane / flow cell, which is comparable price for output of 24 samples. Detail information about calculations is described in Supplementary data 1.2 and in Supplementary Table 6.

The outcomes presented in this study showed that all included known viruses/viroids could be identified by both NGS approaches. Both approaches successfully identified also two divergent strains of PepMV, which was, despite short fragments of sRNA already shown previously (Kutnjak et al., 2014). However, a putative novel *Cytorhabdovirus* was only detected by analysing the data generated from ribosomal RNA depleted total RNA. Additionally, the results revealed the strength of NGS technology for the simultaneous detection and identification of several different known/unknown plant viruses from a different sample material, with a different amount of viral/viroid nucleotides and in a different host plants. Similar conclusions were derived from studies using other virus enrichment approaches on single or few viral species (Adams et al., 2009; Al Rwahnih et al., 2009; Kreuze et al., 2009; Kutnjak et al., 2014; Visser et al., 2016), e.g., both, sequencing of virion-associatad nucleic acids and sRNAs enabled a discovery of a new virus, previously overlooked by other detection techniques (Candresse et al., 2014). Our study further indicates the advantages of NGS in such cases and strengthens its use as a tool in plant virus/viroid diagnostics.

AUTHOR CONTRIBUTIONS

MR, DK, and NB conceived the idea, AP, MR, DK, and NB designed the experiments. AF provided samples. AP performed laboratory part of the experiment and analyzed the data with the assistance of IA and DK. AP wrote the draft of the manuscript. All authors significantly contributed with reviewing and editing the manuscript.

FUNDING

The work was supported by COST Action FA1407 (DIVAS), thought STSM (short term scientific mission), Euphresco NGS-Detect project and Slovenian Research Agency, AP is a recipient of a Ph.D. research grant from the Slovenian Research Agency.

ACKNOWLEDGMENTS

We thank Dr. Heiko Ziebell for providing the sample material, in this paper labeled as sample V, *Pisum sativum* infected with PNYDV, Dr. Ummey Hany for help with the library preparation and sequencing and Dr. Nataša Mehle for providing the sample material in this paper labeled as sample IX, *Nicotiana benthamiana* infected with CSNV.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.01998/full#supplementary-material

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, I., and Fox, A. (2016). "Diagnosis of plant viruses using next-generation sequencing and metagenomic analysis," in *Current Research Topics in Plant Virology*, eds A. Wang and X. Zhou (Cham: Springer), 323–335. doi: 10.1007/978-3-319-32919-2_14

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., Dave, A., Anderson, M. M., Rowhani, A., Uyemoto, J. K., and Sudarshana, M. R. (2013). Association of a DNA virus with grapevines affected by red blotch disease in California. *Phytopathology* 103, 1069–1076. doi: 10.1094/PHYTO-10-12-0253-R

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* 35, 235–241.

Boben, J., Kramberger, P., Petrovic, N., Cankar, K., Peterka, M., Štrancar, A., and Ravnikar, M. (2007). Detection and quantification of tomato mosaic virus in irrigation waters. *Eur. J. Plant Pathol.* 118, 59–71. doi: 10.1007/s10658-007-9112-1

Boonham, N., Kreuze, J., Winter, S., van der Vlugt, R., Bergervoet, J., Tomlinson, J., et al. (2014). Methods in virus diagnostics: from ELISA to next generation sequencing. *Virus Res.* 186, 20–31. doi: 10.1016/j.virusres.2013.12.007

Borges, F., and Martienssen, R. A. (2015). The expanding world of Small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 1–12. doi: 10.1038/nrm4085

CLC Analyses-related questions, (2017). "*Analyses-Related Questions: De Novo Assembly*." QIAGEN. Available online at https://secure.clcbio.com/helpspot/index.php?pg=kb.page&id=185

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

Dovas, C. I., Katis, N. I., and Avgelis, A. D. (2002). Multiplex detection of criniviruses associated with epidemics of a yellowing disease of tomato in greece. *Plant Disease* 86, 1345–1349. doi: 10.1094/PDIS.2002.86.12.1345

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

Gaafar, Y., and Ziebell, H. (2016). *Vicia Faba*, *V. Sativa* and Lens Culinaris as new hosts for pea necrotic yellow dwarf virus in Germany and Austria. *New Dis. Rep.* 34:28. doi: 10.5197/j.2044-0588.2016.034.028

Gutiérrez-Aguirre, I., Mehle, N., Delić, D., Gruden, K., Mumford, R., and Ravnikar, M. (2009). Real-time quantitative PCR based sensitive detection and genotype discrimination of Pepino Mosaic virus. *J. Virol. Methods* 162, 46–55. doi: 10.1016/j.jviromet.2009.07.008

Hull, R. (ed.). (2014). "Replication of plant viruses," in *Plant Virology, 5th Edn.* (Norwich: Academic Press), 341–421.

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

Ivanov, K. I., Eskelin, K., Bašic, M., De, S., Lõhmus, A., Varjosalo, M., et al. (2016). Molecular insights into the function of the viral RNA silencing suppressor HCPro. *Plant J.* 85, 30–45. doi: 10.1111/tpj.13088

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, E1–E7. doi: 10.1016/j.virol.2009.03.024

Kreuze, J. (2014). "siRNA deep sequencing and assembly: piecing together viral infections,"in *Detection and Diagnostics of Plant Pathogens*, eds M. L. Gullino and P. J. M. Bonants (Dordrecht: Springer), 21–38.

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. doi: 10.1111/j.1472-765X.2011.03117.x

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. doi: 10.1128/JVI.00690-17

Kutnjak, D., Rupar, M., Gutierrez-Aguirre, I., Curk, T., Kreuze, J. F., and Ravnikar, M. (2015). Deep sequencing of virus derived small interfering RNAs and RNA from viral particles shows highly similar mutational landscape of a plant virus population. *J. Virol.* 89, 4760–4769. doi: 10.1128/JVI.03685-14

Kutnjak, D., Silvestre, R., Cuellar, W., Perez, W., Müller, G., Ravnikar, M., et al. (2014). Complete genome sequences of new divergent potato virus X isolates and discrimination between strains in a mixed infection using small RNAs sequencing approach. *Virus Res.* 191, 45–50. doi: 10.1016/j.virusres.2014.07.012

Loebenstein, G., (2008). "Plant virus diseases: economic aspects" in *Desk Encyclopedia of Plant and Fungal Virology,* eds M. H. V. van Regenmortel and W. J. Mahy Brian (Oxford: Academic Press), 426–430.

Loreti, S., Faggioli, F., Cardoni, M., Mordenti, G., Babini, A. R., Poggi Pollini, C., et al. (1999). Comparison of different diagnostic methods for detection of peach latent mosaic viroid. *EPPO Bull.* 4, 433–438. doi: 10.1111/j.1365-2338.1999.tb01414.x

Martínez de Alba, A. E., Flores, R., and Hernández, C. (2002). Two chloroplastic viroids induce the accumulation of the small RNAs associated with post-transcriptional gene silencing. *J. Virol.* 76, 13094–13096. doi: 10.1128/JVI.76.24.13094-13096.2002

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

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

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

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

Soliman, T., Mourits, M. C. M., Oude Lansink, A. G. J. M., and van der Werf, W. (2012). Quantitative economic impact assessment of an invasive plant disease under uncertainty - a case study for potato spindle tuber viroid (PSTVd) invasion into the european union. *Crop Prot.* 40, 28–35. doi: 10.1016/j.cropro.2012.04.019

Verhoeven, J., Th,. J., Jansen, C. C. C., Willemen, T. M., Kox, L. F. F., Owens, R. A., et al. (2004). Natural infections of tomato by citrus exocortis viroid, columnea latent viroid, potato spindle tuber viroid and tomato chlorotic dwarf viroid. *Eur. J. Plant Pathol.* 110, 823–831. doi: 10.1007/s10658-004-2493-5

Visser, M., Bester, R., Burger, J. T., and Maree, H. J. (2016). Next-generation sequencing for virus detection: covering all the bases. *Virol. J.* 13:85. doi: 10.1186/s12985-016-0539-x

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. doi: 10.1146/annurev-phyto-080614-120030

Yelina, N. E., Savenkov, E. I., Solovyev, A. G., Morozov, S. Y., and Valkonen, J. P. (2002). Long-distance movement, virulence, and RNA silencing suppression controlled by a single protein in Hordei- and Potyviruses: complementary functions between virus families. *J. Virol.* 76, 12981–12991. doi: 10.1128/JVI.76.24.12981-12991.2002

Conflict of Interest Statement: 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.

2.2 High-Throughput Sequencing Facilitates Characterization of a "Forgotten" Plant Virus: The Case of a Henbane Mosaic Virus Infecting Tomato

Anja Pecman, Denis Kutnjak, Nataša Mehle, Magda Tušek Žnidarič, Ion Gutiérrez-Aguirre, Patricija Pirnat, Ian Adams, Neil Boonham and Maja Ravnikar

Frontiers in Microbiology, 2018, 9:2739. doi: 10.3389/fmicb.2018.02739

In this research, we report the detection of henbane mosaic virus (HMV) for the first time in a new natural host and new geographical location (a tomato plant in Slovenia) using HTS. The tomato plant was brought from the field with severe necrotic symptoms and was further diagnostically investigated. In the sampled leaf tissue, potato virus M (PVM) was initially detected using ELISA screening and was subsequently confirmed with electron microscopy. However, inoculated test plants showed symptoms atypical for PVM, thus an additional analysis with HTS using sRNA and rRNA-depleted totRNA enrichment protocols was performed. The obtained sequencing reads confirmed detection of PVM, southern tomato virus (STV) and enabled assembly for the first time of the complete genomic RNA of HMV. HTS technology using sRNA protocol was also applied to sequence the genome of four other known strains of HMV, followed by phylogenetic analysis. The latter taxonomically classified new sequences of HMV within *Potyvirus* genus cluster and showed the relationship of the new HMV tomato isolate with the historic ones, suggesting the existence of at least four putative strains of the virus. With the re-inoculations of different test plants, we achieved isolation of henbane mosaic virus from PVM and STV in the original sample, and performed a host range analysis of HMV isolate found in Slovenia via the inoculation of 20 species belonging to four plant families. HMV caused strong local or systemic disease symptoms in several species from Solanaceae family. Among others, tomato showed severe symptoms after single infection with HMV. The use of rRNA-depleted totRNA and/or sRNA deep sequencing for complete plant viral genome assembly resulted in first report of five complete HMV genomes: the one of the new isolate found in Slovenia and of four other isolates obtained from different virus collections. The countries of origin of those four other isolates were Italy, United Kingdom and United States (California).

frontiers in Microbiology

ORIGINAL RESEARCH
published: 19 November 2018
doi: 10.3389/fmicb.2018.02739

High-Throughput Sequencing Facilitates Characterization of a Forgotten Plant Virus: The Case of a Henbane Mosaic Virus Infecting Tomato

Anja Pecman[1,2], Denis Kutnjak[1], Nataa Mehle [1], Magda Tuek nidari c[1], Ion Guti rrez-Aguirre [1], Patricija Pirnat[3], Ian Adams[4], Neil Boonham[5] and Maja Ravnikar[1,6]**

[1] Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia, [2] Joef Stefan International Postgraduate School, Ljubljana, Slovenia, [3] KZ Agraria Koper z.o.o., Koper, Slovenia, [4] Fera Science Ltd., York, United Kingdom, [5] Institute for Agri-Food Research and Innovation, Newcastle University, Newcastle upon Tyne, United Kingdom, [6] Wine Research Centre, University of Nova Gorica, Nova Gorica, Slovenia

OPEN ACCESS

Edited by:
John P. Carr,
University of Cambridge,
United Kingdom

Reviewed by:
Jes s Navas-Castillo,
Instituto de Hortofruticultura Subtropical y Mediterr nea La Mayora (IHSM), Spain
F Murilo Zerbini,
Universidade Federal de Vi osa, Brazil

***Correspondence:**
Anja Pecman
anja.pecman@nib.si
Maja Ravnikar
maja.ravnikar@nib.si

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

Received: 29 August 2018
Accepted: 26 October 2018
Published: 19 November 2018

Citation:
Pecman A, Kutnjak D, Mehle N, nidari c MT, Guti rrez-Aguirre I, Pirnat P, Adams I, Boonham N and Ravnikar M (2018) High-Throughput Sequencing Facilitates Characterization of a Forgotten Plant Virus: The Case of a Henbane Mosaic Virus Infecting Tomato. Front. Microbiol. 9:2739. doi: 10.3389/fmicb.2018.02739

High-throughput sequencing has dramatically broadened the possibilities for plant virus research and diagnostics, enabling discovery of new or obscure viruses, and virus strains and rapid sequencing of their genomes. In this research, we employed high-throughput sequencing to discover a new virus infecting tomato, *Henbane mosaic virus (Potyvirus*, *Potyviridae*), which was rst discovered at the beginning of 20th century in the United Kingdom in cultivated henbane. A eld tomato plant with severe necrotic symptoms of unknown etiology was sampled in Slovenia and high-throughput sequencing analysis using small RNA and ribosomal RNA depleted total RNA approaches revealed a mixed infection with *Potato virus M* (*Carlavirus*, *Betaexiviridae*), *Southern tomato virus* (*Amalgavirus*, *Amalgamaviridae*) and henbane mosaic virus in the sample. The complete genomic sequence of henbane mosaic virus was assembled from the sequencing reads. By re-inoculation of the infected material on selected test plants, henbane mosaic virus was isolated and a host range analysis was performed, demonstrating the virus was pathogenic on several plant species. Due to limited metadata in public repositories, the taxonomic identication of the virus isolate was initially putative. Thus, in the next step, we used small RNA sequencing to determine genomic sequences of four historic isolates of the virus, obtained from different virus collections. Phylogenetic analyses performed using this new sequence information enabled us to taxonomically position *Henbane mosaic virus* as a member of the *Potyvirus* genus within the chili veinal mottle virus phylogenetic cluster and dene the relationship of the new tomato isolate with the historic ones, indicating the existence of at least four putative strains of the virus. The rst detection of henbane mosaic virus in tomato and demonstration of its pathogenicity on this host is important for plant protection and commercial tomato production. Since the virus was initially present in a mixed infection, and its whole genome was not sequenced, it has probably been

overlooked in routine diagnostics. This study conrms the applicability of a combination of high-throughput sequencing and classic plant virus characterization methods for identication and phylogenetic classication of obscure viruses and historical viral isolates, for which no or limited genome sequence data is available.

Keywords: potyvirus, henbane mosaic virus, tomato, high-throughput sequencing, host range analysis, phylogeny

INTRODUCTION

The immense sequence data generating potential of high-throughput sequencing (HTS) has enabled an accelerated discovery of new virus species in recent years (Barba et al., 2013; Massart et al., 2014; Adams and Fox, 2016) and increased resolution of viral population and evolution studies (Kutnjak et al., 2017). However, currently there is a gap between sequence discovery and biological characterization of new viral species, the latter requiring time-consuming research eorts (Massart et al., 2017). In the present research we address another angle of this problem; we used HTS to detect and sequence the genome of a known plant virus, namely henbane mosaic virus (HMV, genus *Potyvirus*, family *Potyviridae*), which had a poorly characterized genome, despite being the subject of several biological studies in the past.

HMV was discovered in cultivated henbane (*Hyoscyamus niger*) in 1932 in the United Kingdom (Rothamsted Experimental Station) (Hamilton, 1932) and later reported as HMV-R (Rothamsted) (Kitajima and Lovisolo, 1972). However, due to diculties characterizing viruses in the pre-molecular era, new isolates might sometimes be named dierently, as a new virus, based on the induced symptoms and hosts, thus the following historical overview could be incomplete. The virus was later reported in the United Kingdom in *Atropa belladonna* (Smith, 1945) and in *Datura stramonium* (Bradley, 1952). HMV was then reported in 1970 in Italy, infecting *Datura inermis*, *D. stramonium,* and *Physalis alkekengi*. *P. alkekengi* was considered the main host and consequently the new strain was labeled as HMV-A (Lovisolo and Bartels, 1970). The virus was found also in Hungary (Horv□th et al., 1988) in *D. stramonium* and designated as HeMV-W/H. Additionally, there were reports of the virus from Germany and India, where henbane is grown for medical purposes (stated in Lovisolo, 1992). The suggested main host plants for HMV are dierent species from the Solanaceae family (Govier and Plumb, 1972). In nature HMV can be transmitted by aphids (Hamilton, 1932; Bradley, 1952; Lovisolo and Bartels, 1970; Govier and Plumb, 1972) and by mechanical inoculation in greenhouse plants (Govier and Plumb, 1972). Its particles are lamentous in shape (Govier and Plumb, 1972) with a length, estimated by electron microscopy to be approximately 850 nm (Lovisolo and Bartels, 1970).

In all the previous reported cases HMV was found in dierent solanaceous plants, which were not crop or vegetable plants, important from the agricultural perspective. However, in this study, we report for the rst time natural HMV infection in eld-grown tomato, found in Slovenia in 2015 (Ankaran) showing severe necrotic symptoms. Tomato (*Solanum lycopersicum* L.) is one of the most widely grown vegetable crops worldwide (FAOSTAT, 2001). Tomato is a natural host for many plant viruses (Brunt et al., 1996; Mihara et al., 2016) which cause signicant economic losses by reducing crop quality and quantity, thus timely detection of emerging viral diseases in tomato is crucially important (Hanssen et al., 2010). HMV was detected in a symptomatic tomato plant, using HTS, in a mixed infection with *Potato virus M* (*Carlavirus, Betaexiviridae*) and *Southern tomato virus* (*Amalgavirus*, *Amalgamaviridae*). Prior to this study only a short fragment of the HMV genome sequence was available in databases with very limited metadata, thus, the initial species identication was only putative. To address this gap, we have obtained four other historic isolates of the virus from two dierent virus collections and used small RNA high-throughput sequencing to obtain complete genome sequences. Besides the rst detection of the HMV in an important vegetable crop (tomato), we also report the rst detection of the virus in Slovenia and the rst complete genome sequence. Moreover, we demonstrate the utility of HTS for the rapid characterization of known plant virus species with no or little sequence information. We discuss implications of the approach for other viral species and discuss the power of the method for the revision of historical virus isolates and viral collections in general.

MATERIALS AND METHODS

Description of Isolates Included in the Analysis

A tomato (*S. lycopersicum* L.) sample (with laboratory diagnostic identication number D159/15) with necrotic disease symptoms was collected in Slovenia (Ankaran) in the summer of 2015 and stored at −80°C for further analyses. The sample contained a mixed infection with HMV, PVM, and STV. The HMV virus isolate found in this sample will be designated throughout the manuscript as isolate HMV-SI/L representing henbane mosaic virus (HMV) from Slovenia (SI) from *S. lycopersicum* (L).

Four other isolates were sourced from virus collections as lyophilized leaf material and stored at −20°C for further analyses. Isolates HMV-146 and HMV-R were obtained from the Institute for Sustainable Plant Protection (IPSP), Italy. The original host of isolate HMV-146 is *D. inermis* from Torino, Italy, and the origin of isolate HMV-R is Rothamsted, United Kingdom. Isolates henbane mosaic virus ATCC® PV-76™ and henbane mosaic virus ATCC® PV-79™ were purchased from the American Type Culture Collection (ATCC) and will be from this point on labeled as HMV-PV-76 and HMV-PV-79, respectively. The original host

of isolates HMV-PV-76 and HMV-PV-79 is *H. niger* (henbane), and their origin is United States (California) and England, respectively. Detailed information for the isolates is shown in **Supplementary Table 1**.

First Test on Sample HMV-SI/L

Tomato sample from Slovenia was inspected using transmission electron microscopy and tested by ELISA for Impatiens necrotic spot virus (INSV, genus *Orthotospovirus*, family *Tospoviridae*), tomato spotted wilt virus (TSWV, genus *Orthotospovirus*, family *Tospoviridae*), potato virus S (PVS, genus *Carlavirus,* family *Betaexiviridae*) and potato virus M (PVM, genus *Carlavirus,* family *Betaexiviridae*). ELISA was performed using kits containing virus specic antibodies as follows: INSV (Loewe Biochemica GmbH, Germany), TSWV (Adgen, United Kingdom), PVS (Bioreba AG, Switzerland), and PVM (Bioreba AG, Switzerland). Additionally, selected test plants (*S. lycopersicum* cv. Moneymaker, *Nicotiana rustica*, *Nicotiana tabacum* cv. White Burley, *Nicotiana benthamiana*, *Nicotiana clevelandii*, *Nicotiana glutinosa*, *Chenopodium quinoa*, *D. stramonium* and *Capsicum annuum*) were mechanically inoculated with a (1:10) dilution of the original sample in phosphate buer (0.02 M with 2% PVP) and applied to the rst two to three completely expanded leaves dusted with carborundum. Test plants were visually inspected for symptoms and tested by electron microscopy and ELISA for PVM 4 weeks after mechanical inoculation. The original plant sample and symptomatic test plants were also tested for STV using RT-PCR (Sabanadzovic et al., 2009).

RNA Isolation and High-Throughput Sequencing

For all of the reverse-transcription PCRs (RT-PCRs) in the testing and conrmation steps, RNA was isolated from leaf samples using RNeasy Plant Mini Kit (Qiagen, Netherlands) following the manufacturers protocol, with some minor modications as follows. RLT buer without β-mercaptoethanol was added to plant material and RNA was eluted from the RNeasy Mini Spin columns using 50 µl of RNase-free warm water (65°C). All of the samples were stored at −80°C between sampling and extraction. Isolated RNA was stored at −80°C, when not in use.

Ribosomal RNA depleted total RNA sequencing was performed for the sample HMV-SI/L. The RNeasy Plant Mini Kit (Qiagen, Netherlands) was again used for RNA isolation, including the optional DNase treatment (RNA Cleanup protocol; RNeasy Mini Kit; Qiagen, Netherlands). Ribosomal RNA was then depleted from the total RNA and libraries for sequencing were prepared using the ScriptSeqTM Complete Kit (plant leaf) (Illumina, United States). The libraries were sequenced using MiSeq (Illumina, United States) in 2 × 300 bp (V3) mode.

Small (s)RNA sequencing was performed for samples HMV-SI/L; HMV-146; HMV-R; HMV-PV-76; HMV-PV-79. In this case, total RNA was isolated using TRIzol reagent (Invitrogen, United States) following the manufacturers protocols. Total RNA from the four samples was sent to Seqmatic LLC (United States) for sRNA library preparation and sequencing using a HiSeq 2000 (Illumina, United States) in 1 × 50 bp mode.

Analysis of HTS Data for Virus Detection, Reconstruction of HMV Genomes and Their Annotation

In the rst stage, HTS of sample HMV-SI/L was performed for both, rRNA depleted total RNA and sRNA. Two dierent detection pipelines were used to detect viral sequences in the HTS data for both approaches as previously described (Pecman et al., 2017). In both datasets, several contig sequences matched dierent potyviruses with relatively low similarity, indicating the presence of an unknown potyvirus (later identied as HMV). Since contigs assembled from sRNA reads using the above-cited pipeline were relatively short, we performed an additional assembly of sRNA reads using SPAdes (Bankevich et al., 2012; Nurk et al., 2013). The parameter careful and combined k-mer sets of 15, 17, 19, and 21 were applied to produce long assembled sequences (Barrero et al., 2017). For each of the two sequencing approaches, all of the contigs matching potyvirus sequences were further assembled to obtain a complete or near complete viral genome sequence using CLC Genomic Workbench 10 (Qiagen). The nal consensus genomic sequences assembled from rRNA depleted total RNA and sRNA data were compared to conrm they were identical. Finally, the trimmed reads from both datasets were mapped to the nal complete consensus viral genomic sequence. The mapping results were visually inspected for any errors. For genome assembly conrmation six pair of primers were designed (**Supplementary Table 2**). RT-PCR reactions were carried-out using those primers and OneStep RT-PCR Kit (Qiagen, Netherlands). The reaction conditions were 50°C for 30 min, 95°C for 15 min, followed by 35 cycles of 94°C for 30 s, 51°C for 60 s and 72°C for 60 s. The amplicons obtained by RT-PCR were puried using MinElute PCR Purication Kit (Qiagen, Netherlands) and submitted to Sanger sequencing (GATC Biotech AG, Germany). The sequences obtained were aligned against the HMV-SI/L genome to conrm that they were identical.

For samples HMV-146, HMV-R, HMV-PV-76 and HMV-PV-79, sRNA sequencing was performed and the resulting datasets were analyzed to identify viral sequences using the pipeline described previously (Pecman et al., 2017). Complete consensus genomic sequences of HMV in these samples were reconstructed by assembly of sRNA reads as described above. Finally, the sRNA reads derived from each isolate were mapped to the corresponding reconstructed whole genome consensus sequences and the mapping results were visually inspected for any errors.

Furthermore, in order to validate the assemblies within the repeated region (TATATA) around position 9980 nt, we designed universal PCR primers for all HMV isolates: HMV-UNI-F: 5′-TTAGCCCGATATGCTTTC-3′ and HMV-UNI-R: 5′-CTATCTTCCACTTCAGGT-3′. The RT-PCR reaction was performed using OneStep RT-PCR Kit (Qiagen, Netherlands). The reaction conditions were 50°C for 30 min, 95°C for 15 min, followed by 35 cycles of 94°C for 30 s, 48°C for 60 s, and 72°C for 60 s. The amplicons obtained by RT-PCR were puried

using MinElute PCR Purication Kit (Qiagen, Netherlands) prior to sequencing (GATC Biotech AG, Germany). The sequences obtained were aligned against the HMV genomes and the repeated region was validated.

The polyproteins of each of the HMV isolates were annotated using Sequin[1] and the individual putative cleavage sites for each gene product were manually determined based on known cleavage sites for other potyviruses reported in the literature (King et al., 2012). The predictions of molecular weight for putative viral protein products were calculated using the Protein Molecular Weight Calculator[2]. Genomic sequences of HMV-SI/L; HMV-146; HMV-R; HMV-PV-76; HMV-PV-79 were deposited in GenBank under accessions numbers MH779472, MH779473, MH779474, MH779475, and MH779476, respectively (**Supplementary Table 1**).

RT-PCR Assay for HMV-SI/L Detection

In order to conrm HMV-SI/L infection in test plants, one (HMV-NIb-F: 5′-GTCAAGAAGTTCAAAGGG-3′ and HMV-CP-R: 5′-TACACCACACCATCAATC-3′) out of the six primer pairs designed for genome assembly was used. Negative controls for RNA isolation and no template controls were also tested. RT-PCR was done as described above using the OneStep RT-PCR Kit (Qiagen, Netherlands) in a 10 μl reaction volume.

Pairwise Comparisons of Viral Genome Sequences and Phylogenetic Analyses

Pairwise comparison between nucleotide sequences of HMV isolates was performed in CLC Genomic Workbench 11. The comparison was done rstly by comparing whole genome sequences for the ve isolates sequenced (HMV-SI/L; HMV-146; HMV-R; HMV-PV-76; HMV-PV-79). The comparison was also performed using partial genome sequences (1600 nt) to enable comparison with the published sequence of isolate HMV PHYS/H-Hungary (accession number AM184113). The results were visualized as heatmaps. In addition, nucleotide identities of the HMV-SI/L polyprotein sequence were compared with four other isolates (HMV-146, HMV-R, HMV-PV-76, and HMV-PV-79) and visualized using SimPlot 3.5.1 (Lole et al., 1999). A plot of nucleotide identity was obtained using the polyprotein sequence of isolate HMV-SI/L as a query sequence; HMV-R, HMV-PV-76 and HMV-PV-79 were compared to the query sequence as a group, since they were highly similar to each other (above 99% of nucleotide identity). A sliding window of 400 nt was used across the alignment in steps of 40 nt.

To reveal phylogenetic relationships between henbane mosaic virus isolates and other members of *Potyviridae* family we performed a phylogenetic analysis including complete viral polyprotein sequences of the ve HMV isolates from this study and other members of the *Potyviridae* family. Alignment of the complete polyproteins of the known members of *Potyviridae* family was obtained from the International Committee on Taxonomy of Viruses (ICTV) resources[3], including 130 viral species. Sequences of the ve HMV isolates sequenced in this study were added to the alignment. Additionally, we performed blastn searches using HMV-SI/L against the NCBI *nt* database and added also two viral species with sequenced complete genomes (KY623506 and MF997470) with relatively high similarity to HMV (detected within rst 100 blastn hits), which were not present in the initial ICTV alignment. All sequences were codon aligned applying MUSCLE in MEGA7 (Tamura et al., 2013). Phylogenetic trees were constructed from the alignment obtained in MEGA7 using a maximum likelihood algorithm and applying the GTR + G + I substitution model, which was determined to best t the data. Bootstrap replication (100 pseudoreplicates) was used to assess the statistical support of the groups on the tree.

We aligned also partial genome sequences (∼1600 nt) of the six HMV isolates (HMV-SI/L, HMV-146, HMV-R, HMV-PV-76, HMV-PV-79, and Hungarian HMV PHYS/H) and other *Potyvirus* species, which clustered in the same group as HMV isolates in the above described complete polyprotein analysis. Additionally, we performed blastn searches of the partial genome sequence of HMV-SI/L against the NCBI *nt* database and added one viral species with partial genome sequence (FJ543110) and with relatively high similarity to HMV (detected within rst 100 blastn hits), which was not present in the initial ICTV alignment. Several *Rymovirus* species (Y09854, AY623626, AY623627) were used as an outgroup. As described above, the partial genome sequences were codon aligned applying MUSCLE in MEGA7 and phylogenetic tree was constructed in MEGA7 on the basis of this alignment using maximum likelihood algorithm with GTR + G + I substitution model. Bootstrap replication (1000 pseudoreplicates) were used to assess the statistical support of the groups within the tree.

Isolation of HMV-SI/L From Mixed Infection and Analysis of Its Host Range

HMV was detected in the eld-grown tomato sample in a mixed infection with two other viruses. The isolation of HMV from the mixed infection was achieved by several re-inoculation steps. A tomato sample, positive for HMV, PVM, and STV, was used for inoculation of several test plants (see section First Test on Sample HMV-SI/L). The *N. clevelandii* plant material, to which PVM and HMV were successfully transferred, was then used to inoculate several test plants of *S. lycopersicum* cv. Moneymaker. The leaf material of *S. lycopersicum* with a single infection of HMV (conrmed by ELISA and RT-PCR) was collected and further used as an inoculum for the host range analysis.

A total of 20 dierent plant species/varieties from several botanical families were selected (**Table 1**) and included in the host range analysis. Since the virus was found in tomato, species from the Solanaceae family were mainly included in the analysis, however, to test for potential new hosts, plants

[1]https://www.ncbi.nlm.nih.gov/Sequin/index.html

[2]https://www.sciencegateway.org/tools/proteinmw.htm

[3]https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/potyviridae/565/resources-potyviridae

TABLE 1 | Symptoms of HMV-SI/L isolate developed on different test plants and their conrmation using RT-PCR.

Test plants		Number of symptomatic/ asymptomatic plants	Local symptoms on inoculated leaves	RT-PCR result inoculated symptomatic/ asymptomatic leaves	Symptoms on non-inoculated leaves	RT-PCR result non-inoculated symptomatic/ asymptomatic leaves
Botanical family	**Species**					
Solanaceae	*Solanum lycopersicum* Moneymaker	7/1	Leaf deformation, necrosis, chlorosis,	NA	Leaf deformation, blistering, necrosis, chlorosis,	+/−
	Solanum lycopersicum Rio grande	8/0	Leaf deformation, necrosis, chlorosis, mosaic	NA	Leaf deformation, blistering, necrosis, chlorosis, mosaic	+
	Solanum lycopersicum Roma	6/2	Leaf deformation, necrosis, chlorosis, mosaic	NA	Leaf deformation, blistering, necrosis, chlorosis, mosaic	+/−
	Solanum melongena Black beauty	8/0	Local lesions	+	None	−
	Capsicum annuum Soroksari	0/8	None	NA	None	−
	Solanum tuberosum Pentland	0/8	None	−	None	−
	Solanum tuberosum Igor	0/8	None	NA	None	−
	Solanum tuberosum Desiree	0/8	None	−	None	−
	Nicotiana glutinosa	8/0	Necrosis, mosaic	NA	Mosaic	+
	Nicotiana tabacum Samsun	8/0	Necrosis, mosaic	NA	Mosaic	+
	Nicotiana benthamiana	8/0	Leaf deformation	NA	Leaf deformatons dark coloration, chlorosis	+
	Hyoscyamus niger	8/0	Necrosis, mosaic	NA	Mosaic, blistering	+
	Datura stramonium	4/4	Wilting, yellowing, mosaic	NA	Wilting, yellowing, mosaic	+/−
	Physalis oridana	2/6	Leaf deformation, blistreing	NA	Leaf deformation, blistering, dwarf growth	+/−
Brassicaceae	*Brassica oleracea*	0/8	None	NA	None	−
	Brassica rapa	0/8	None	NA	None	−
Cucurbitaceae	*Cucumis melo*	0/8	None	NA	None	−
Amaranthaceae	*Chenopodium quinoa*	0/8	None	NA	None	−
	Chenopodium amaranticolor	0/8	None	NA	None	−
	Amaranthus sp.	0/8	None	NA	None	−

NA, not applicable; +, positive RT-PCR result; −, negative RT-PCR result.

from three other families were also included. Mechanical inoculation was performed as explained in section First Test on Sample HMV-SI/L. For each plant species, we mechanically inoculated 8 plants with infected material, and 4 plants were mock inoculated with buer only, as a negative control. Inoculated test plants were grown in the quarantine greenhouse (20 ± 5°C; 16 h photoperiod) and the appearance of disease symptoms were recorded weekly. Three weeks post inoculation, locally and/or systemically infected leaves of symptomatic and asymptomatic plants were collected as pooled samples (separate pools were collected for each category: local leaves and systemic leaves, separately for symptomatic plants and asymptomatic plants). Collected pooled samples were tested for HMV with RT-PCR using the primers HMV-NIb-F and HMV-CP-R (**Supplementary Table 2**). In parallel, to conrm the absence of PVM, all samples were analyzed using reverse-transcription real-time PCR for PVM (Yang et al., 2014). Mock inoculated plants were analyzed in the same manner. The absence of STV was already conrmed in all of the diagnostic test plants used in rst stage of the experiments (which were the source for further experiments).

Out of four isolates obtained from the virus collections, HMV-PV-76 was the only isolate present in a single infection. Thus, additionally we also checked the pathogenicity of the isolate HMV-PV-76 and compared it to the HMV-SI/L isolate by inoculating *N. benthamiana* and *S. lycopersicum* cv. Moneymaker using the same method as described in section First Test on Sample HMV-SI/L.

Transmission Electron Microscopy (TEM)

The original sample, selected test plants (see First Test on Sample HMV-SI/L) and leaves from *S. lycopersicum* and *N. benthamiana* (**Figure 1B**) infected with the HMV-SI/L isolate were examined using TEM. The sample (20 µl) was applied to Formvar-coated, carbon-stabilized copper grids and negatively stained using a 1% aqueous solution of uranyl acetate (SPI Supplies), followed by visualization using Philips CM 100 transmission electron microscope (FEI, Eindhoven, Netherlands). Images were captured using an ORIUS SC 200 CCD camera (Gatan Inc., Pleasanton, United States).

RESULTS

Field-Grown Tomato Sample Contained HMV in a Mixed Viral Infection

A sample from eld-grown tomato plant (*S. lycopersicum* L.) with severe necrotic symptoms (**Figure 1A**) was brought to the laboratory for a diagnostic investigation. The sample was analyzed for several dierent plant viruses and gave a positive result for PVM using ELISA. The presence of virus particles matching carlavirus morphology was conrmed by TEM. After the initial mechanical inoculation of nine dierent species of test plants, ve of them (*N. benthamiana, N. glutinosa, C. quinoa, D. stramonium, C. annuum*) did not show any symptoms, the ELISA assay was negative for PVM and no virus particles were observed in those plant samples. Four of the test plants (*S. lycopersicum* cv. Moneymaker, *N. rustica, N. tabacum* cv. White Burley, *N. clevelandii*) were PVM positive by ELISA (the results were also conrmed by TEM). Since the disease symptoms observed on the original sample (**Figure 1A**) and on the tomato test plants (**Supplementary Figure 1**) were not typical for infection with PVM alone, the original plant sample was further analyzed. Total RNA was isolated from the original sample and sequenced using sRNA and rRNA depleted total RNA approaches. Henbane mosaic virus (HMV) was detected in both HTS data sets in mixed infection with southern tomato virus (STV) and with the previously detected PVM. The presence of HMV in the original sample and in the symptomatic test plants was conrmed using a RT-PCR assay designed in this study. The presence of STV was conrmed using RT-PCR (Sabanadzovic et al., 2009) only in the original plant sample and was not detected in the test plants, which was an expected outcome, since STV is not known to be mechanically transmissible (Sabanadzovic et al., 2009).

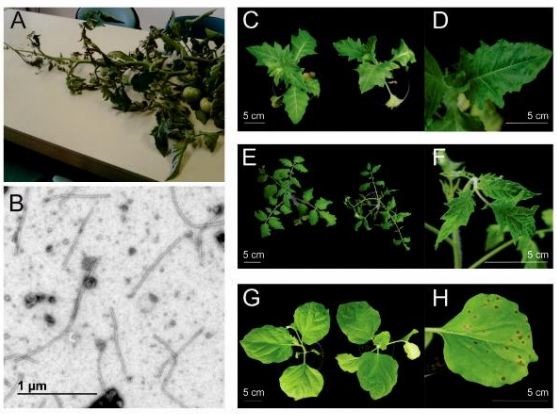

FIGURE 1 | Disease symptoms on the original eld-grown tomato sample and test plants infected for host range analysis with HMV-SI/L isolate. **(A)** Infected eld-grown tomato plant with severe necrotic symptoms brought to laboratory for diagnostic investigation. **(B)** Viral particles of HMV-SI/L from mechanically inoculated *Nicotiana benthamiana* sample, visualized by transmission electron microscopy. **(CH)** Disease symptoms caused by HMV-SI/L isolate on selected plant species included in the host range analysis. **(C)** Left: mock-inoculated *H. niger*, right: HMV-SI/L inoculated *H. niger*. **(D)** *H. niger* infected leaf showing blistering. **(E)** Left: mock-inoculated *S. lycopersicum* cv. Moneymaker, right: HMV-SI/L inoculated *S. lycopersicum* cv. Moneymaker showing leaf deformation. **(F)** S. *lycopersicum* cv. Moneymaker infected leaf showing blistering. **(G)** Left: mock-inoculated *S. melongena*, right: HMV-SI/L inoculated *S. melongena*. **(H)** *S. melongena* inoculated leaf showing necrotic lesions (local symptoms).

HMV Was Mostly Present in Isolates From Viral Collections in Mixed Infections

Before this study only a partial genome sequence of a single isolate of HMV was present in the public database, with no aliated publication and little metadata (NCBI GenBank Acc. No. AM184113). The overlapping part of HMV-SI/L isolate sequenced in this study was 88% identical to this sequence, leading to a putative identication of the virus as HMV. To validate this putative identication, we obtained four other isolates designated as HMV from virus collections (HMV-R, HMV-146, HMV-PV-76, and HMV-PV-79) and performed sRNA sequencing to obtain their complete genome sequences. Using the sRNA virus detection pipeline (Pecman et al., 2017) we detected potyvirus sequences in all four samples. They had high sequence identities (**Figure 2C**) to the putative HMV isolate from tomato (HMV-SI/L). In all four samples HMV was detected and conrmed using the HMV-UNI RT-PCR assay. However, HTS analysis revealed that three out of four isolates contained sequences of other virus species, specically: HMV-R and HMV-146 contained also *Potato aucuba mosaic virus* (*Potexvirus*, *Alphaexiviridae*) and HMV-PV-79 contained *Potato virus Y* (*Potyvirus*, *Potyviridae*).

Genome Characteristics of HMV

Complete consensus genome sequences were reconstructed for all ve HMV isolates included in the study and deposited into NCBI GenBank (**Supplementary Table 1**). They had a genome organization typical for the genus *Potyvirus* with a single open reading frame (ORF) encoding a polyprotein of 32483249 amino acids and an estimated molecular weight of 367 kDa. The ORF encoded all typical potyviral protein products, with an unusually long P1 protein (476 amino acid residues) (**Figure 2A**). Additionally, one of the isolates (HMV-PV-76) contained an uncommon Nib-Pro/CP cleavage site (Q/V) which has previously been reported for an isolate of lettuce

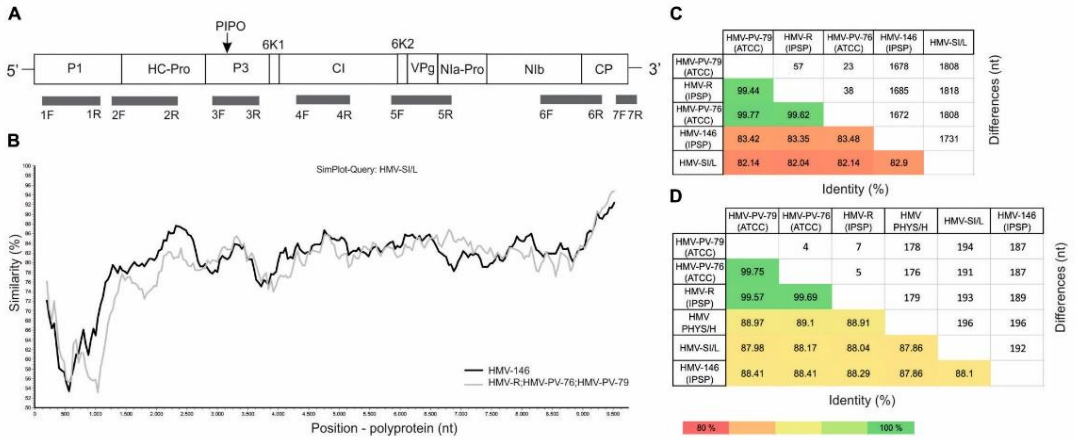

FIGURE 2 | HMV genome representation, position of RT-PCR primers used in this study and pairwise comparisons of sequenced isolates. **(A)** Scheme of HMV genome. The positions and lengths of the protein products are indicated with corresponding box lengths. Below, the locations of all primers described in **Supplementary Table 2** are marked. **(B)** The SimPlot graph showing the nucleotide similarity (%) across the HMV polyprotein between HMV-SI/L (query), HMV-146 (black line), and HMV-R, HMV-PV-76 and HMV-PV-79 (as a group, gray line). **(C)** Heatmap showing the results of pairwise comparisons between nucleotide sequences of whole genomes of HMV-SI/L, HMV-R, HMV-146, HMV-PV-76, and HMV-PV-79. Average nucleotide identities (%) are shown below the diagonal, number of different nucleotides are shown above the diagonal. **(D)** Heatmap showing the results of pairwise comparisons between partial genome sequence (1600 nt) of HMV-SI/L, HMV-R, HMV-146, HMV-PV-76, HMV-PV-79, and HMV-PHY/H (AM184113). Average nucleotide identities (%) are shown below the diagonal, number of different nucleotides are shown above the diagonal.

mosaic virus (Dinant et al., 1991) and was later found also in one isolate of potato virus Y (Blanco-Urgoiti et al., 1998) and several other potyviruses (Adams et al., 2005a). All HMV isolates had the PIPO frame shift protein within the P3 cistron and HC-Pro motifs involved in aphid transmission: $RITC_{51-54}$ and $PTK_{309-311}$. The detailed data about the polyprotein cleavage sites, molecular weights and lengths of products are presented in the **Supplementary Table 3**.

Investigated HMV Isolates Cluster Into a Monophyletic Group and Contain High Within-Group Diversity

In the phylogenetic analysis based on the alignment of the complete nucleotide polyprotein sequences of HMV isolates and other species from the *Potyviridae* family (**Figure 3A**), the HMV isolates were positioned within the *Potyvirus* genus. They represented a monophyletic group, which clustered within the subgroup of potyvirus species, together with: *Habenaria mosaic virus* (AB818538), *Gloriosa stripe mosaic virus* (EF427894), *Yam mild mosaic virus* (JX470965), *Tobacco vein banding mosaic virus* (EF219408), *Chili ringspot virus* (JN008909), *Wild tomato mosaic virus* (DQ851495), *Chili veinal mottle virus* (AJ237843), *African eggplant mosaic virus* (MF997470), and *Pepper veinal mottle virus* (DQ645484). Phylogenetic analysis based on the alignment of the smaller subset of partial genome sequences (**Figure 3B**) enabled us to also include the Hungarian isolate of HMV (HMV PHYS/H, AM184113) into the analysis, positioning it within the HMV cluster.

In both phylogenetic trees, three HMV isolates (HMV-R, HMV-PV-76, and HMV-PV-79) clustered together with high bootstrap support. HMV-146 and HMV-SI/L were signicantly dierent from the HMV-R, HMV-PV-76, HMV-PV-79 group, but also from each other, thus they represented separate groups on the tree (**Figure 3A**). HMV PHYS/H was also dierent from all other isolates, not clustering with any of them on the tree (**Figure 3B**). To evaluate similarities and dierences between HMV isolates in a greater detail, SimPlot analysis and pairwise comparisons of nucleotide sequences were performed. The pairwise comparisons was used to calculate the average percentage of identity and the number of dierences in nucleotides (i) using whole genomes of all ve HMV isolates (**Figure 2C**) and (ii) using partial genome sequence enabling the inclusion of the Hungarian isolate (AM184113) (**Figure 2D**). The highest percent of identity (above 99%) and thus the lowest number of dierences were between HMV-R, HMV-PV-76, HMV-PV-79 isolates. The similarity of HMV-146 and HMV-SI/L to the above mentioned isolates and to each other was about 82% (**Figure 2C**). HMV PHYS/H identity to all ve above mentioned HMV isolates was in each case around 88% (**Figure 2D**). Taken together, the results show that six HMV isolates sequenced to date represent four divergent phylogenetic groups, which may correspond to four distinct viral strains.

SimPlot analysis (**Figure 2B**) demonstrated the similarity between the three putative strain groups across the polyprotein sequence, specically: between HMV-SI/L (used as query), HMV-R, HMV-PV-76, HMV-PV-79 isolates (grouped as one, since their similarity was above 99%) and HMV-146. A signicant drop in similarity between the groups is observed in the P1 protein region; subsequent blastx (against NCBI *nr*) and blastn (against NCBI *nt*) similarity searches did not show signicant

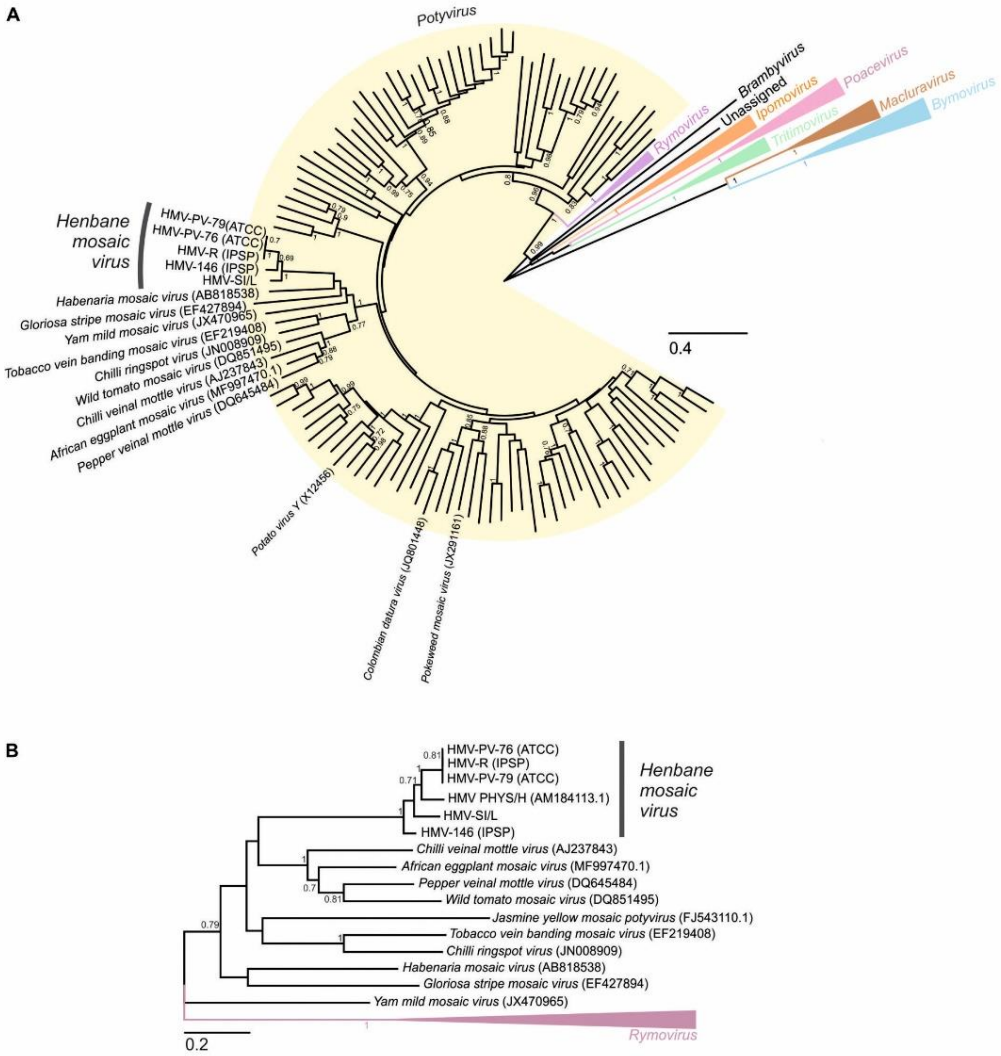

FIGURE 3 | Maximum likelihood phylogenetic trees showing the relationship between *Henbane mosaic virus* and other member of *Potyviridae* family. Scale bars represent maximum-likelihood estimates of the number of substitution per site. **(A)** Phylogenetic tree based on the codon-aligned nucleotide sequences of the polyproteins of fully sequenced members of the family *Potyviridae*. Sequences were codon aligned using MUSCLE in MEGA. **(A)** Maximum Likelihood tree was constructed using the General Time Reversible (+G + I) substitution model and 100 bootstrap replicates; only bootstrap support values of 0.7 or above are shown for clarity reasons. The *Potyvirus, Rymovirus, Brambyvirus, Tritimovirus, Poacevirus, Ipomovirus, Macluravirus*, and *Bymovirus* genera are highlighted in yellow, purple, black, green, pink, orange, brown and blue, respectively, and labeled accordingly. **(B)** Maximum likelihood phylogenetic tree constructed from the alignment of partial (1600 nt) genome nucleotide sequences of the viral species clustering in the same group as HMV (see **Figure 3A**) and *Rymovirus* representatives used as an outgroup. Sequences were codon aligned using MUSCLE in MEGA. A Maximum Likelihood tree was then constructed using the General Time Reversible (+G + I) substitution model and 1000 bootstrap replicates; only bootstrap values of 0.7 or above are shown for clarity reasons.

similarity (i.e., identity > 40%) of HMV P1 product to any known sequence.

HMV-SI/L Causes Symptomatic Infection of Several Solanaceous Species

Host range analysis for the HMV-SI/L isolate (detailed in **Table 1**) showed that eight of ten tested solanaceous species are hosts for HMV, which was conrmed by observation of disease symptoms and positive RT-PCR tests for HMV. For *H. niger, S. lycopersicum* (**Figures 1CF**) and plants from the genus *Nicotiana* all or almost all (68) of the eight inoculated plants expressed disease symptoms 3 weeks after mechanical inoculation. Smaller number of *D. stramonium* (4/8) and *P. oridana* (2/8) plants expressed disease symptoms. Most severe disease symptoms were observed in the following two cases: in *D. stramonium*, where 4 of 8

plants died 6 weeks post inoculation and in *P. oridana* , where plants showed very strong dwarf growth. In *Solanum melongena*, the HMV did not move from the inoculated leaves and local lesions developed (**Figures 1G,H**) without developing systemic infection. This was conrmed by positive RT-PCR test results for inoculated leaves and negative RT-PCR test results for non-inoculated leaves, sampled three and four (data not shown) weeks post inoculation. Pooled samples of plants with no disease symptoms were RT-PCR negative and none of these plants showed disease symptoms not even 7 or 8 weeks after the inoculation.

Additional comparison of the pathogenicity between isolates HMV-SI/L and HMV-PV-76 (the latter is 99% identical to HMV-R on genome level) in *N. benthamiana* and *S. lycopersicum* cv. Moneymaker revealed more severe symptoms following infection with isolate HMV-PV-76. Three weeks after inoculation, plants of *N. benthamiana* infected with HMV-PV-76 showed more severe symptoms (curling, blistering, yellowing) than plants infected with HMV-SI/L, and died 4 weeks after inoculation. Similarly, in the case of *S. lycopersicum* cv. Moneymaker, the symptoms induced by HMV-PV-76 were more severe (blistering, curling, dwarf growth, parsley leaves), in comparison to HMV-SI/L (**Supplementary Figure 2**).

DISCUSSION

High-throughput sequencing has in the recent decade facilitated discovery of many new viruses and viral strains and largely improved the possibilities for broad range virus detection and screening. The increased rate of virus discovery, also brings new challenges, especially considering the biological characterization of the newly detected viral sequences. The latter is crucial to establish their importance, i.e., their impact on agriculture and trade (Massart et al., 2017) or their natural ecosystems. Since this process comprises laborious classical virology techniques, in many cases, the biological characterization of rapidly discovered new viral species is lagging behind their discovery, often simply due to time constrains.

However, for henbane mosaic virus described here, the perspective was reversed; very limited genomic information was available for a virus rst described in 1932 (Hamilton, 1932) and subsequently characterized biologically. Here, rstly, HTS was used as a reliable generic detection technique for identifying the presence of known and unknown viruses in eld samples, which resulted in the detection of HMV for the rst time in tomato and the rst time in any plant in Slovenia. Since the virus was present in a mixed infection, and little genomic information was available at the time, it was overlooked by routinely used diagnostic methods. Secondly, HTS enabled us to rapidly generate complete genomic sequences of several isolates of the HMV, which was followed by their genome annotation and phylogenetic analyses.

Although HMV detection and characterization was reported 86 years ago, the correct taxonomic determination of the virus detected in eld-grown tomato was uncertain due to the lack of genomic information in sequence databases. To overcome this problem, we obtained, sequenced and analyzed four other HMV isolates from dierent virus collections using the sRNA sequencing approach. This enabled us, rst, to characterize the HMV genome, which has typical potyvirus organization and contains an unusually long P1 protein, known for its great variability within the genus (Adams et al., 2005b). Consequently, HMV has one of the longest genomes in the genus, which is in agreement (Cui et al., 2014) with previous observations of longer virus particles (Lovisolo and Bartels, 1970). Secondly, the genomic information obtained enabled us to perform phylogenetic analyses and taxonomically classify the HMV isolates within the *Potyvirus* genus. Interestingly, HMV is placed within the chili veinal mottle virus cluster and is not closely related to any of the virus species (*Potato virus Y*, *Colombian datura virus,* or *Pokeweed mosaic virus*), for which a distant serological relationship with HMV has been previously speculated (Lovisolo and Bartels, 1970; Govier and Plumb, 1972; **Figure 3A**). Further, pairwise comparison of HMV isolates genomic sequences revealed that three dierently named isolates (HMV-PV-76, HMV-PV-79, and HMV-R) were very similar (99% identity), while the other two (HMV-SI/L and HMV-146) form distinct phylogenetic clusters (**Figures 2**, **3**). Sequence variability between HMV-SI/L and HMV-PV-76 was reected also in a dierent pathogenicity on *N. benthamiana* and *S. lycopersicum* cv. Moneymaker (**Supplementary Figure 2**). According to the sequence similarities within other potyvirus species (Adams et al., 2005b) these three units might belong to 3 dierent strain groups within the same species. The partially sequenced HMV-PHYS/H isolate detected in Hungary, might belong to the 4th group. The observed sequence diversity among the low number of analyzed HMV isolates is surprisingly high. This could be explained by several scenarios. They were isolated from dierent countries, dierent host plants and put into the collections at dierent decades during the last 80 years (**Supplementary Table 1**). To estimate the current diversity among dierent HMV isolates in nature, additional studies, including several samples collected from the natural environment, from dierent host plants and from a broader geographic range, should be completed.

Previously, the relationships between dierent HMV isolates were examined mainly based on their biological properties, e.g., host range and symptomatology. Although sequenced isolates described in this paper are probably not identical as the isolates described in previously published biological studies, the comparison between them, considering the host range and symptomatology, could give us an additional insight into their diversity. Susceptible hosts of a HMV-SI/L isolate with pronounced disease symptoms were closely related plant species, e.g., *H. niger*, *Nicotiana* spp. and *S. lycopersicum* (dierent varieties). Similar results were obtained by Lovisolo and Bartels (1970) for HMV-R and HMV-A. In our study, HMV-SI/L caused a hypersensitive reaction (showing local necrotic lesion and no systemic infection) in *S. melongena*. On the other hand, HMV-R showed no disease symptoms on *S. melongena* (Hamilton, 1932), thus it could be used as a test plant to discriminate between HMV-SI/L from HMV-R. HMV-SI/L did not infect *C. annuum*; however, taking dierent reports into account, there are contradicting results whether HMV-R can infect *C. annuum* (Lovisolo and Bartels, 1970) or not (Salamon, 2018). HMV-SI/L

did not infect *Chenopodium amaranticolor* or *C. quinoa* in our study. However, HMV-R and HMV-A isolates induced disease symptoms on *C. amaranticolor* (Lovisolo and Bartels, 1970) and, in another study, HMV-R did not induce symptoms on *C. quinoa* (Salamon, 2018).

Nevertheless, all the previously reported host range analysis should be interpreted with caution. The composition of those inoculums was not conrmed with HTS, so they could have contained mixed infections of viral species, which could explain some of the contradictory results. Indeed, HTS analysis done here on HMV isolates obtained from virus collections showed that three out of four HMV isolates were present in mixed infections. The host range analysis done in our study is the rst one in conducted where single infection of the inoculum was conrmed. Furthermore, when comparing the observed symptoms of mixed infection from the eld tomato sample (**Figure 1A**, severe necrotic symptoms) with symptoms of HMV-SI/L single infection of tomato plant in the greenhouse (**Figures 1E,F** and **Table 1**), the symptomatology is dierent, possibly due to the fact that the virus was present in mixed infection in the originally sampled tomato plant. We demonstrated that HTS can help improving complex plant disease etiology investigations, as a fast screening tool, implemented prior to biological characterization studies. It helps to assess the possibility of mixed infections. In addition, HTS can be used as an eective tool for screening and revision of the already established virus collections, to obtain more accurate status of the deposited isolates.

The study of HMV presented in this paper demonstrated the benet of using HTS for characterization of known viral species with limited or absent genomic information. The *Potyvirus* genus (*Potyviridae* family) is the largest genus of plant viruses, including 168 virus species, which cause diseases in a wide range of plant species. Among them, according to the latest ICTV report, 114 have complete genome sequences available, for 48 there are just partial genomic sequences available, and for 6 of them there is no sequence data in public repositories (ICTV, 2017); the latter two categories likely representing a similar scenario as that described in this study. It is likely that these viruses are currently not a major concern regarding agricultural production and trade and thus less research eort has been focused on them. However, such viruses can become emergent problems in the future, since they constantly evolve and adapt to new environments and can switch hosts (Geoghegan et al., 2017). As a result, they might become emerging problems if cultivated plants become natural hosts for some of them in the future. We demonstrated that HTS can be used to characterize full genomes of such viruses. This improves information in the databases by adding missing sequences, enabling more rapid development of diagnostic assays, leading to quicker response times for emerging disease problems.

AUTHOR CONTRIBUTIONS

AP, MR, DK, and NM designed the experiments. PP discovered new type of disease symptoms on tomato, started further investigation, and provided the eld tomato sample infected with HMV-SI/L. NM performed rst screening analysis on the eld tomato sample. AP performed laboratory part of the experiments and analyzed the data with the assistance of IA and DK. M recorded TEM photos. AP wrote the draft of the manuscript. All authors signicantly contributed with reviewing and editing the manuscript.

FUNDING

The study was nancially supported by the Ministry of Agriculture, Forestry and Food of the Republic of Slovenia through ERA-NET ArimNet2 project (EMERAMB) and Slovenian Research Agency (research core funding no. P4-0165 and program for young researchers Ph.D. research grant of AP). The work was supported also by COST Action FA1407 (DIVAS), through STSM (short term scientic mission) grant.

ACKNOWLEDGMENTS

We thank IPSP [Institute for Sustainable Plant Protection (Italy), Marina Ciuo] for providing the sample material in this study labeled as HMV-R and HMV-146, Ju Smole, Katarina Ba cnik, and Lidija Maticic for technical assistance and University of Ljubljana Botanical Garden for providing seeds of *H. niger*.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02739/full#supplementary-material

REFERENCES

Adams, I., and Fox, A. (2016). Diagnosis of plant viruses using next-generation sequencing and metagenomic analysis, in *Current Research Topics in Plant Virology*, eds A. Wang and X. Zhou (Cham: Springer), 323335. doi: 10.1007/978-3-319-32919-2_14

Adams, M. J., Antoniw, J. F., and Beaudoin, F. (2005a). Overview and analysis of the polyprotein cleavage sites in the family potyviridae. *Mol. Plant Pathol.* 6, 471487. doi: 10.1111/j.1364-3703.2005.00296.x

Adams, M. J., Antoniw, J. F., and Fauquet, C. M. (2005b). Molecular criteria for genus and species discrimination within the family potyviridae. *Arch. Virol.* 150, 459479. doi: 10.1007/s00705-004-0440-6

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455477. doi: 10.1089/cmb.2012.0021

Barba, M., Czosnek, H., and Hadidi, A. (2013). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106136. doi: 10.3390/v6010106

Barrero, R. A., Napier, K. R., Cunnington, J., Liefting, L., Keenan, S., Frampton, R. A., et al. (2017). An internet-based bioinformatics toolkit for plant biosecurity diagnosis and surveillance of viruses and viroids. *BMC Bioinformatics* 18:26. doi: 10.1186/s12859-016-1428-4

Blanco-Urgoiti, B., S□nchez, F., P□rez De San Rom□n, C., Dopazo, J., and Ponz, F. (1998). Potato virus y group c isolates are a homogeneous pathotype but two dierent genetic strains. *J. Gen. Virol.* 79, 20372042. doi: 10.1099/0022-1317-79-8-2037

Bradley, R. H. E. (1952). Studies on the aphid transmission of a strain of henbane mosaic virus. *Ann. Appl. Biol.* 39, 7897. doi: 10.1111/j.1744-7348.1952.tb01000.x

Brunt, A. A., *Mosaic Virus. Description of Plant Viruses* Crabtree, K., Dallwitz, M. J., Gibbs, A. J., Watson, L., and Zurcher, E. J. (1996). *Plant Viruses Online: Descriptions and Lists from the VIDE Database, Version: 16th January 1997.* Available at: http://bio-mirror.im.ac.cn/mirrors/pvo/vide/refs.htm

Cui, J., Schlub, T. E., and Holmes, E. C. (2014). An allometric relationship between the genome length and virion volume of viruses. *J. Virol.* 88, 64036410. doi: 10.1128/JVI.00362-14

Dinant, S., Lot, H., Albouy, J., Kuziak, C., Meyer, M., and Astier-Manifacier, S. (1991). Nucleotide sequence of the 3 terminal region of lettuce mosaic potyvirus rna shows a gln/val dipeptide at the cleavage site between the polymerase and the coat protein. *Arch. Virol.* 116, 235252. doi: 10.1007/BF01319245

FAOSTAT (2001). *FAO Statistical Databases*. Rome: Food and Agriculture Organization of the United Nations.

Geoghegan, J. L., Duch□ne, S., and Holmes, E. C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog.* 13:e1006215. doi: 10.1371/journal.ppat.1006215

Govier, D. A., and Plumb, R. T. (1972). *Henbane mosaic virus*. Available at: www.dpvweb.net/dpv/showdpv.php?dpvno=095

Hamilton, A. M. (1932). On three new virus diseases of hyoscyamus niger. *Ann. Appl. Biol.* 19, 550567. doi: 10.1111/j.1744-7348.1932.tb04341.x

Hanssen, I. M., Lapidot, M., and Thomma, B. P. (2010). Emerging viral diseases of tomato crops. *Mol. Plant Microbe Interact.* 23, 539548. doi: 10.1094/MPMI-23-5-0539

Horv□th, J., Salamon, P., Wolf, I., and K□lber, M. (1988). Henbane mosaic potyvirus pathogenic to wild and cultivated potato. *Potato Research.* 31, 311320. doi: 10.1007/BF02365540

ICTV. (2017). *International Committee on Taxonomy of Viruses*. Available at: https://talk.ictvonline.org/.

King, A. M. Q., Adams, M. J., Carsten, E. B., and Lefkowitz, E. J. (2012). *Virus Taxonomy: Classication and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses.* Amsterdam: Elsevier Inc.

Kitajima, E. W., and Lovisolo, O. (1972). Mitochondrial aggregates in datura leaf cells infected with henbane mosaic virus. *J. Gen. Virol.* 16, 265271. doi: 10.1099/0022-1317-16-3-265

Kutnjak, D., Elena, F. S., 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:, e00690-17e690e617. doi: 10.1128/JVI.00690-17

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., et al. (1999). Full-length human immunodeciency virus type 1 genomes from subtype c-infected seroconverters in india, with evidence of intersubtype recombination. *J Virol.* 73, 152160.

Lovisolo, O. (1992). Ecology and taxonomy of some european potyviruses. *Arch. Virol. Suppl.* 5, 311316. doi: 10.1007/978-3-7091-6920-9_32

Lovisolo, O., and Bartels, R. (1970). On a new strain of henbane mosaic virus from physalis alkekengi. *J. Phytopathol.* 69, 189201. doi: 10.1111/j.1439-0434.1970.tb03102.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 scientic impacts of plant viruses and viroids identied by NGS technologies. *Front. Microbiol.* 8:45. doi: 10.3389/fmicb.2017.00045

Massart, S., Olmos, A., Jijakli, H., and Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Res.* 188, 9096. doi: 10.1016/j.virusres.2014.03.029

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., et al. (2016). Linking virus genomes with host taxonomy. *Viruses* 8, 1015. doi: 10.3390/v8030066

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric mda products. *J. Comput. Biol.* 20, 714737. doi: 10.1089/cmb.2013.0084

Pecman, A., Kutnjak, D., Guti□rrez-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

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, 130137. doi: 10.1016/j.virusres.2008.11.018

Salamon, P. (2018). Termesztett □s vadon □lo burgonyaf□□k v□rusos betegs□gei □s v□rusai magyarorsz□gon. 9. A bel□ndek mozaik v□rus (Henbane mosaic virus, HMV) rezervo□r gazdan□v□nyei □s patol□giai v□ltoz□konys□gaN□w□*nyv□delem* 2018, 161171.

Smith, K. M. (1945). A further note on the viruses aecting atropa belladonna and a description of a virus complex attacking hyoscyamus niger. *Parasitology* 36, 209210. doi: 10.1017/S0031182000012191

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

Yang, L., Nie, B., Liu, J., and Song, B. (2014). A reexamination of the eectiveness of ribavirin on eradication of viruses in potato plantlets *in vitro* using elisa and quantitative RT-PCR. *Am. J. Potato Res.* 91, 304311. doi: 10.1007/s12230-013-9350-z

Conict of Interest Statement: IA was employed by company Fera Science Ltd., Sand Hutton, York, United Kingdom and PP was employed by company KZ Agraria Koper z.o.o., Koper, Slovenia.

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

2.3 Systematic Comparison of Nanopore and Illumina Sequencing for the Detection of Plant Viruses and Viroids Using Total RNA Sequencing Approach

Anja Pecman, Ian Adams, Ion Gutiérrez-Aguirre, Adrian Fox, Neil Boonham, Maja Ravnikar and Denis Kutnjak

Frontiers in Microbiology, 2022, 13:883921, doi:10.3389/fmicb.2022.883921

In the last paper, we compared the most established HTS platform, i.e., Illumina (MiSeq, benchtop sequencer) with the rapidly and constantly evolving nanopore technology (MinION, portable sequencer, Oxford Nanopore Technologies) for the detection of plant viruses/viroids. The comparison was made on a broad range of samples, containing two plant viroids (CLVd, TASVd), which were sequenced using nanopore technology for the first time, and 11 plant viruses (TYLCV, ToCV, PepMV (two strains), ToMV, STV, CaMV, CCyV1, TSWV, PVeV 1,2,3), which differ in their genome organization and concentration in tested samples. For the comparison five samples of different plant species (S*olanum lycopersicum, Brassica oleracea, Nicotiana tabaccum, Phaseolus vulgaris*) that contained plant viruses/viroids in single or mixed infections were selected. Total RNA extracts of all the samples were sequenced using the Illumina platform (MiSeq) (rRNA-depleted totRNA) and with the nanopore technology (MinION) using direct RNA library preparation. In addition, we sequenced one sample, which contained five different plant viruses (TYLCV, ToCV, PepMV (two strains), ToMV, STV) and a viroid (CLVd), using three additional protocols of the nanopore technology: direct RNA sequencing of rRNA-depleted totRNA, cDNA-PCR sequencing of totRNA and cDNA-PCR sequencing of rRNA-depleted totRNA. We performed bioinformatics analysis on the data to compare the performance of different protocols for detection of plant viruses and viroids. However, whilst direct RNA sequencing of total RNA was the quickest protocol, it performed slightly worse than the other protocols tested; using this protocol, we were not able to detect one virus, ToMV (out of 11 viruses and 2 viroids included in this study), which was present in the sample in an extremely low titer. All other MinION sequencing protocols showed improved performance with outcomes similar to Illumina sequencing (similar fractions of reads corresponding to different viruses were detected for all of the protocols), with cDNA-PCR sequencing of rRNA-depleted totRNA showing the best performance amongst all tested nanopore MinION sequencing protocols. The results of this systematic comparison confirmed that nanopore sequencing is a promising technique for plant virus/viroid detection in research and diagnostic laboratories when appropriate sample and library preparation is selected.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 May 2022
doi: 10.3389/fmicb.2022.883921

Systematic Comparison of Nanopore and Illumina Sequencing for the Detection of Plant Viruses and Viroids Using Total RNA Sequencing Approach

Anja Pecman[1,2], Ian Adams[3], Ion Gutiérrez-Aguirre[1], Adrian Fox[3], Neil Boonham[4], Maja Ravnikar[1] and Denis Kutnjak[1]**

[1] Department of Biotechnology and System Biology, National Institute of Biology, Ljubljana, Slovenia, [2] Jožef Stefan International Postgraduate School, Ljubljana, Slovenia, [3] Fera Science Ltd., York, United Kingdom, [4] Institute for Agri-Food Research and Innovation, Newcastle University, Newcastle upon Tyne, United Kingdom

OPEN ACCESS

Edited by:
Elvira Fiallo-Olivé,
La Mayora Experimental Station (CSIC), Spain

Reviewed by:
Pasquale Saldarelli,
National Research Council (CNR), Italy
Miguel A. Aranda,
Spanish National Research Council (CSIC), Spain

***Correspondence:**
Anja Pecman
anja.pecman@nib.si
Denis Kutnjak
denis.kutnjak@nib.si

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

Received: 25 February 2022
Accepted: 31 March 2022
Published: 11 May 2022

Citation:
Pecman A, Adams I, Gutiérrez-Aguirre I, Fox A, Boonham N, Ravnikar M and Kutnjak D (2022) Systematic Comparison of Nanopore and Illumina Sequencing for the Detection of Plant Viruses and Viroids Using Total RNA Sequencing Approach. Front. Microbiol. 13:883921. doi: 10.3389/fmicb.2022.883921

High-throughput sequencing (HTS) has become an important tool for plant virus detection and discovery. Nanopore sequencing has been rapidly developing in the recent years and offers new possibilities for fast diagnostic applications of HTS. With this in mind, a study was completed, comparing the most established HTS platform (MiSeq benchtop sequencer—Illumina), with the MinION sequencer (Oxford Nanopore Technologies) for the detection of plant viruses and viroids. Method comparisons were performed on five selected samples, containing two viroids, which were sequenced using nanopore technology for the first time and 11 plant viruses with different genome organizations. For all samples, sequencing libraries for the MiSeq were prepared from ribosomal RNA-depleted total RNA (rRNA-depleted totRNA) and for MinION sequencing, direct RNA sequencing of totRNA was used. Moreover, for one of the samples, which contained five different plant viruses and a viroid, three additional variations of sample preparation for MinION sequencing were also used: direct RNA sequencing of rRNA-depleted totRNA, cDNA-PCR sequencing of totRNA, and cDNA-PCR sequencing of rRNA-depleted totRNA. Whilst direct RNA sequencing of total RNA was the quickest of the tested approaches, it was also the least sensitive: using this approach, we failed to detect only one virus that was present in a sample at an extremely low titer. All other MinION sequencing approaches showed improved performance with outcomes similar to Illumina sequencing, with cDNA-PCR sequencing of rRNA-depleted totRNA showing the best performance amongst tested nanopore MinION sequencing approaches. Moreover, when enough sequencing data were generated, high-quality consensus viral genome sequences could be reconstructed from MinION sequencing data, with high identity to the ones generated from Illumina data. The results of this study implicate that, when an appropriate sample and library preparation are selected, nanopore MinION sequencing could be used for the detection of plant viruses and viroids

with similar performance as Illumina sequencing. Taken as a balance of practicality and performance, this suggests that MinION sequencing may be an ideal tool for fast and affordable virus diagnostics.

Keywords: high-throughput sequencing, plant virus/viroid detection, comparison, nanopore MinION sequencing, Illumina MiSeq sequencing

INTRODUCTION

Globalization of agriculture and international trade facilitate the spread of plant viruses and viroids to new geographic regions with unexpected consequences for food production and natural ecosystems (Jones and Naidu, 2019). To decrease the negative impact of viral diseases on crop production and food safety, rapid and generic plant virus or viroid detection technologies (potentially applicable onsite) are needed. Since the first use of high-throughput sequencing (HTS) for generic detection of plant viruses (Adams et al., 2009; Al Rwahnih et al., 2009; Donaire et al., 2009; Kreuze et al., 2009), a range of HTS platforms were developed and became commonly used for plant virus or viroid detection and discovery. The low error rate and relatively "high throughput" of different instruments of the most widely used platforms, such as Illumina, offer many possibilities for plant virus research (Villamor et al., 2019). However, such "high throughput" may not always be necessary, e.g., when analyzing the single or small number of samples in routine diagnostic laboratories, it increases costs. Often, such samples are outsourced to commercial service providers; however, this increases the turnaround time from a couple of days to several weeks and limits the possibilities for quality control of the full process, which might be crucial in some situations. On the other hand, nanopore sequencing implemented by Oxford Nanopore Technologies offers scalable solutions from small flow cells (up to 2.8 Gb of data per run) accessed using a Flongle adapter, through to the MinION flow cells (up to 50 Gb per run) used here to parallel platforms such as the GridION (up to 250 Gb per run) and PromethION (up to 14 Tb per run). MinION sequencing has the potential benefit that the data can be analyzed in real time (Branton and Deamer, 2019). Compared to the established Illumina sequencing, nanopore sequencing enables long-read sequencing (Van Dijk et al., 2018), which can be an advantage for some applications. However, depending on the specific flow cell used, the error rate can reach up to 15% (Ip et al., 2015; Van Dijk et al., 2018), which can limit the potential applications.

One of the first large-scale applications of the MinION sequencer in virology was for real-time genomic surveillance in the Ebola epidemic in West Africa (Quick et al., 2016). The use of MinION for virus detection and investigation is steadily increasing. In human virology and animal virology, MinION has been used to detect dengue (Yamagishi et al., 2017), Zika (Quick et al., 2017) chikungunya, hepatitis C (Greninger et al., 2015), and porcine reproductive and respiratory syndrome virus (Tan et al., 2019) and is at the moment globally utilized for SARS-CoV-2 genomic surveillance (Meredith et al., 2020.)

In plant pathology, MinION has been successfully used for the detection of bacteria, fungi, and viruses using RNA or DNA sequencing (Chalupowicz et al., 2019), plum pox virus and *Candidatus* liberibacter asiaticus in plant tissue and insect samples (Bronzato et al., 2018), and viruses in wheat (Fellers et al., 2019) and cassava (Boykin et al., 2019). In several studies (Filloux et al., 2018; Beddoe et al., 2020; Vazquez-Iglesias et al., 2022,) both, MinION sequencing and Illumina sequencing, were used for virus detection; however, systematic comparison between established Illumina sequencing and nanopore sequencing for the detection of a wide array of viruses with different genome types is still lacking.

A previous study, comparing the sequencing of small (s)RNA and total (tot)RNA sequencing using the Illumina platform (Pecman et al., 2017), demonstrated that both approaches can be used for the detection of most of the plant viruses and viroids in a sample, and that totRNA sequencing was a better choice for sequencing novel viruses at low titres. In this report, the focus is on testing the performance of the MinION sequencer for a rapid detection of a wide array of plant viruses or viroids using totRNA sequencing. First, a systematic comparison was made of the fastest or simplest approach for this platform involving direct RNA sequencing of total RNA with an established approach based on sequencing ribosomal RNA-depleted total RNA (rRNA-depleted totRNA) using the MiSeq platform (Illumina). Using the methodology described previously (Pecman et al., 2017) several well-characterized samples containing a broad range of plant viruses and viroids with different genome organizations were included in the comparison.

The main advantage of the direct RNA sequencing approach is a simple and fast library preparation protocol (SQK-RNA002) resulting in long reads without PCR bias (Garalde et al., 2018,) but unfortunately, a large amount (500 ng) of RNA is required as the input and the error rate is still relatively high (Wongsurawat et al., 2019). Thus, second, a study was completed using one of the samples, which contained five different plant viruses (including one viral species with two different strains) and one viroid which was analyzed using three other approaches: direct RNA sequencing of rRNA-depleted total RNA, cDNA-PCR sequencing of total RNA, and cDNA-PCR sequencing of rRNA-depleted total RNA.

The results obtained using different nanopore sequencing approaches and the Illumina sequencing were compared in terms of suitability for the detection of plant viruses, using complete datasets and rarefied subsets of data. Not all of the nanopore sequencing approaches performed equally well; however, the results demonstrate that some of them can be confidently used for generic detection of different genome types of plant viruses and viroids, since their performance was comparable to the established Illumina sequencing approach.

TABLE 1 | Samples included in the comparison with corresponding results from: HTS (+ virus/viroid detected using pipeline described in section Virus and Viroid Detection Workflow,–virus/viroid not detected using pipeline described in section Virus and Viroid Detection Workflow; NA, not applicable).

Sample number	Host	Virus/viroid present (Baltimore classification)		Initial detection with HTS sequencing using complete datasets				NCBI GenBank accession number	NCBI SRA accession number (MinION direct RNA sequencing of totRNA / MinION direct RNA sequencing of rRNA-depleted totRNA / MinION cDNA-PCR sequencing of totRNA / MinION cDNA-PCR sequencing of rRNA-depleted totRNA / Illumina rRNA-depleted totRNA sequencing)
			MinION direct RNA sequencing of totRNA	MinION direct RNA sequencing of rRNA-depleted totRNA	MinION cDNA-PCR sequencing of totRNA	MinION cDNA-PCR sequencing of rRNA-depleted totRNA	Illumina rRNA-depleted totRNA sequencing		
I	*Solanum lycopersicum*	TYLCV (ssDNA)	+	+	+		+	KY810789	SRR17660996/SRR17660995/ SRR17660994/SRR17660993/SRR17319908
		ToCV (ssRNA+)	+	+	+		+	KY810786, KY810787	
		PepMV (ssRNA+)	+	+	+		+	KF718832.1 (Pep-MV-EU), JX866666.1 (PepMV-CH)	
		ToMV (ssRNA+)	-	+	+	+	+	KY810788	
		STV (dsRNA)	+	+	+	+	+	KY810783	
		CLVd (viroid)	+	+	+	+	+	KY810771	
II	*Brassica oleracea*	CaMV (dsDNA-RT)	+	NA	NA	NA	+	KY810770	SRR17660992/NA/NA/NA/SRR17319907
		CCyV1 (ssRNA-)	+	NA	NA	NA	+	KY810772	
III	*Nicotiana tabaccum*	TSWV (ssRNA-)	+	NA	NA	NA	+	OM112200, OM112201, OM112202	SRR17660991/NA/NA/NA/SRR17319906
IV	*Solanum lycopersicum*	TASVd (viroid)	+	NA	NA	NA	+	KY810784	SRR17660990/NA/NA/NA/SRR17319905
V	*Phaseolus vulgaris*	PVeV1 (dsRNA)	+	NA	NA	NA	+	/	SRR17660989/NA/NA/NA/SRR17319904
		PVeV2 (dsRNA)	+	NA	NA	NA	+	OM112199	
		PVeV3 (dsRNA)	+	NA	NA	NA	+	/	

Virus/viroid names: tomato yellow leaf curl virus (TYLCV, Begomovirus, Geminiviridae), tomato chlorosis virus (ToCV, Crinivirus, Closteroviridae), pepino mosaic virus (PepMV, Potexvirus, Alphaflexiviridae), tomato mosaic virus (ToMV, Tobamovirus, Virgaviridae), southern tomato virus (STV, Amalgavirus, Amalgaviridae), columnea latent viroid (CLVd, Pospiviroid, Pospiviroidae), cauliflower mosaic virus (CaMV, Caulimovirus, Caulimoviridae), cabbage cytorhabdovirus 1 (CCyV1, Cytorhabdovirus, Rhabdoviridae), tomato spotted wilt orthotospovirus (TSWV, Orthotospovirus, Tospoviridae), tomato apical stunt viroid (TASVd, Pospiviroid, Pospiviroidae), and phaseolus vulgaris alphaendornavirus 1, 2, 3 (PVeV1, 2, 3, Alphaendornavirus, Endornaviridae).

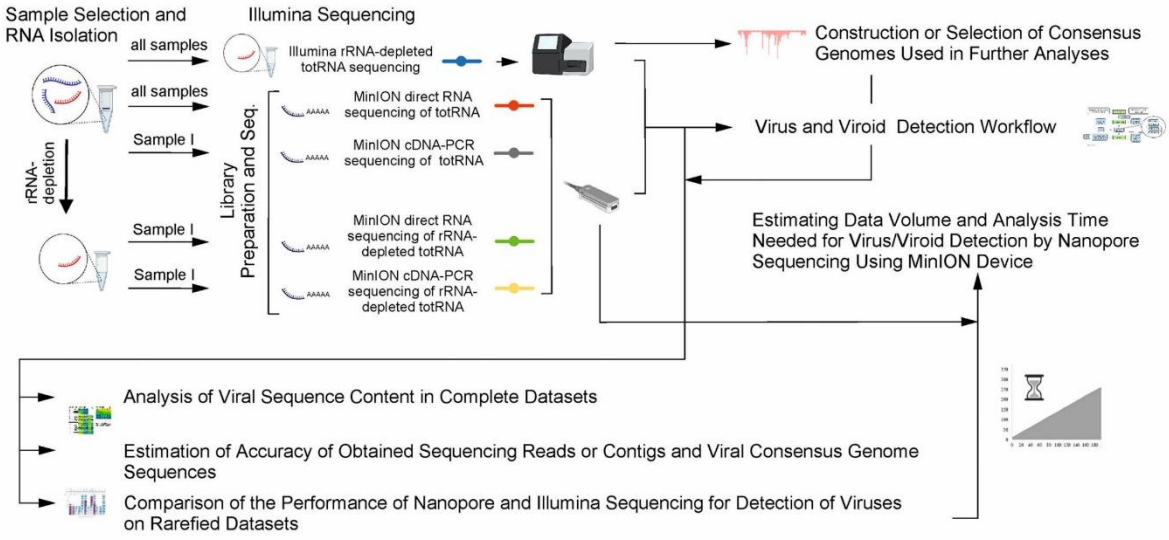

FIGURE 1 | An overview of the conducted study. Listed titles of chapter material and methods are connected with arrows according to the course of work (part of the figure was created with BioRender.com).

MATERIALS AND METHODS

Sample Selection and RNA Isolation

To perform an extensive comparison of the methods on a wide array of plant viruses with different genome types, five plant samples were selected, containing different viruses, most of which (samples I, II, and IV) have already been very well characterized for viruses using HTS and targeted detection methods (Pecman et al., 2017). Either infected plant leaf samples (samples I, II, III, and V) or an infected seed sample (sample IV) were used (**Table 1**; **Supplementary Tables S1**, **S2**). Additionally, leaves of healthy tobacco plants were used as a negative control. RNA was isolated (**Figure 1**) from all leaf samples using the RNeasy Plant Mini Kit (Qiagen) including a DNase step (RNase-Free DNase Set, Qiagen) according to the manufacturer's instructions. From the seed sample (sample IV), RNA was isolated using a combination of CTAB buffer and RNeasy Plant Mini Kit (Qiagen) as described in the study of Adams et al. (2009) with minor modification: incubation with 4 M LiCl was done at 4°C overnight.

A part of the RNA extracted from sample I was further processed: ribosomal RNA was depleted from the extract using the RiboMinusTM Plant Kit for RNA-seq (Invitrogen# A1083808), obtaining two versions of sample I: with and without ribosomal RNA depletion (**Figure 1**, for more details refer to **Supplementary Tables S1, S2**.)

Library Preparation and Sequencing

The direct RNA sequencing kit (SQK-RNA002) and cDNA-PCR sequencing kit (SQK-PCS108) required polyA+ RNA as input RNA. Therefore, for each sample, polyA tailing of RNA was performed using E. coli poly(A) polymerase (NEB# M0276). The mixture was incubated at 37°C for 15 min. The reaction was stopped by directly proceeding to the clean-up step with Agencourt® AMPure® XP beads (Beckman Coulter) using 1.8 (AMPure® XP beads): 1 (poly(A) tailing mixture) ratio.

The direct RNA sequencing kit (SQK-RNA002, version DRS_9080_v2_revB_22Nov2018, Oxford Nanopore Technologies) was used to prepare sequencing libraries for all the samples included in the study. The cDNA-PCR sequencing kit (SQK-PCS108, version PCS_9035_v108_revH_26Jun2017, Oxford Nanopore Technologies) was additionally used to prepare libraries from sample I and sample I with ribosomal-depleted RNA.

For all but one library, the recommended amount of RNA input was used (Oxford Nanopore Technologies protocols); however, when preparing the library for direct RNA sequencing of rRNA-depleted totRNA (sample I), 278 ng of RNA, instead of 500 ng of RNA, was used (due to the lower extraction yield). Each library was then sequenced on a separate flow cell (R9.5.1) for 46–48 h using MinION device and MinKNOW software (v18.12.6). The reads were base-called using Guppy v3.1.5 and command: rna_r9.4.1_70bps_hac.cfg / dna_r9.4.1_bps450_hac.cfg –device auto –u_substitution false.

Illumina Sequencing

Sequencing libraries for each sample were prepared using total RNA and the ScriptSeq™ Complete Plant Leaf Kit (production discontinued, Illumina, USA) which included a ribosomal RNA depletion step. The libraries were sequenced on a MiSeq (Illumina, USA) using a 2x300-bp (V3) cartridge.

Virus and Viroid Detection Workflow

In the first part of the data analysis, the aim was to investigate how well different sequencing approaches compare in terms of virus detection from complete datasets (**Figure 1**). To achieve

this, established in-house virus detection workflows were used and virus presence was reported as follows.

Illumina reads were analyzed using a pipeline in CLC Genomic Workbench (v12, v21) and additional Diamond BLASTX analysis, as described below. Quality control was performed, then, adapters were removed from all reads, and additionally, reads were trimmed by quality (quality limit = 0.05) and length (all reads shorter than 30 nts were discarded). Trimmed reads were mapped to viral RefSeq (NCBI database, updated 19.05. 2019) and *de novo* assembled. Contigs (longer than 100 nts) were mapped to viral RefSeq (NCBI database, updated 19.05.2019) and unmapped contigs were further analyzed by searching for conserved protein domains using Pfam analysis (v32) (refer to **Supplementary Tables S3–S6** for details about the used parameters). Additionally, *de novo* assembled contigs were analyzed with Diamond BLASTX (v0.9.22) (Buchfink et al., 2015) against the NCBI nr database (June 2018). Diamond outputs were taxonomically classified and visualized using Megan 6.19.2 (Huson et al., 2007.)

To analyse nanopore sequencing data, a similar pipeline was constructed using tools, which enable analysis of long reads. The statistics and quality of MinION reads were checked using the programs NanoQC v0.8.1, NanoStat v1.1.2, and NanoPlot v1.20.1 (De Coster et al., 2018). The read plots generated using the NanoQC v0.8.1 were inspected for each sample individually and were used to determine how to trim them (length of reads, head of reads, and tail of reads) (**Supplementary Table S1**) using program NanoFilt 2.5.0 (De Coster et al., 2018). The trimmed reads were again quality checked with NanoStat v1.1.2 and then mapped to the viral RefSeq (NCBI database, updated 19.05.2019) using minimap2 (v2.16-r922) and the commands: minimap2 -ax splice -uf -k14 for direct RNA reads and minimap2 -ax map-ont for cDNA-PCR reads. Reads were also analyzed using Diamond BLASTX (v0.9.22) (Buchfink et al., 2015) with the –frameshift 15–range-culling–sensitive command option. All reads were *de novo* assembled by combining different programs using a Pomoxis (https://github.com/nanoporetech/pomoxis) inspired approach: after fast mapping (minimap2) (Li, 2018) and *de novo* assembly (Miniasm) (Li, 2016), two rounds of the contig correction using racon (Vaser et al., 2017) were run. The script together with corresponding parameters is shown in **Supplementary Data 1**. The assembled contigs were analyzed using BLASTn against the NCBI nt database and visualized with Megan 6.19.2 (Huson et al., 2007). TASVd (sample V) was not detected by mapping direct RNA MinION reads to the viral RefSeq database, so in the next step, more closely related sequence from NCBI GenBank (KY810784) was used as the reference for reads and contig mapping.

Construction or Selection of Consensus Genomes Used in Further Analyses

To be able to perform comparisons of different sequencing approaches for different viruses or viroids, complete or near complete consensus genomic sequences of viruses in the samples were obtained. Since some of the samples were identical to the ones from a previous study (Pecman et al., 2017), these reference sequences were already available. For phaseolus vulgaris alphaendornavirus 2 (PVeV2, sample V) and tomato spotted wilt orthotospovirus TSWV (sample III), a reference consensus was generated based on the mapping of Illumina reads and contigs, as previously described (Pecman et al., 2017). For pepino mosaic virus (PepMV), two divergent strains (80% nucleotide identity) were detected in the sample I (PepMV-EU and PepMV-Ch2); thus, in this case, the complete genome sequences of KF718832.1 and JX866666.1 were used in subsequent comparisons as described previously (Pecman et al., 2017). For tomato yellow leaf curl virus (TYLCV) (KY810789), tomato chlorosis virus (ToCV) (KY810786), (KY810787), cauliflower mosaic virus (CaMV) (KY810770), and cabbage cytorhabdovirus 1 (CCyV1) (KY810772), the reference sequences described in the study of Pecman et al., (2017) were used in first step, but due to some mismatches after mapping Illumina reads and contigs to those reference sequences, few nucleotides were changed and "new" consensus genome sequences were used for the purpose of the following analysis only (**Supplementary Data 2**). In the case of sample V, only PVeV2 complete genome was covered by reads by both approaches, and thus, only this endornavirus was included in further analyses.

Analysis of Viral Sequence Content in Complete Datasets

To calculate the viral sequence content in the complete datasets, trimmed reads for each dataset were mapped to the corresponding reconstructed or selected viral or viroid reference sequences (Section Construction or Selection of Consensus Genomes Used in Further Analyses). For each sample–virus–sequencing type combination, three parameters were reported: the percentage of mapped reads, average depth (the number of times a nucleotide is covered by a sequencing read averaged across the complete reference genome sequence), and fraction of reference covered by reads.

Estimation of Accuracy of Obtained Sequencing Reads or Contigs and Viral Consensus Genome Sequences

The next step enabled the investigation of (I) average nanopore reads and contig identities (compared to the corresponding reference sequences) and (II) identities of consensus genome sequences generated after mapping the reads to the reference sequence, from now on named "consensus sequence identity". (I) To determine the average identities of nanopore sequencing reads (proxy of error rate), and the identities of *de novo* assembled contigs (generated from those reads), compared to reference sequences, reads and contigs were mapped to corresponding viral genome reference sequences. Identities were calculated using Minimap2 PAF output (Pairwise mApping Format) (Li, 2018). In this way, the average BLAST-like alignment identity was calculated for each mapping of nanopore sequencing data, either with reads or with contigs.

(II) If sequencing errors in nanopore reads are relatively random and if there is a substantial number of reads covering the reference genome sequence after mapping, the resulting

consensus sequence should be "error corrected". To test this, extracted consensus sequences (or their fragments–if whole genomes were not obtained) derived from mapping nanopore or Illumina reads to reference sequences were aligned with original reference sequences and pairwise identities were calculated using CLC Workbench Genomics v12, v21. For this analysis, consensus sequences (or their fragments if whole genomes were not obtained) were used if they had at least 1x average coverage in the read mapping step.

Comparison of the Performance of Nanopore and Illumina Sequencing for Detection of Viruses on Rarefied Datasets

To be able to compare the datasets obtained by different sequencing approaches, complete datasets were rarefied to obtain subsamples with comparable numbers of nucleotides. Depending on the original size (**Supplementary Tables S1, S2**,) datasets from different sequencing approaches generated a different number of subsamples. The largest datasets, contained 1,500 million nucleotides, followed by 1,300, 1,100, 900, 700, 500, 300, 200, 100, 50, 30, and 10 millions of nucleotides (**Supplementary Table S7**). Each set of subsampled reads was randomly generated using CLC Genomic Workbench (v12, v21) for Illumina reads and Seqtk Sample for MinION reads. Subsampling was repeated five times for each of the subsample sizes for each sample.

The rarefied subsets were used for further comparative analysis of different sequencing approaches. The read subsamples were mapped to reconstructed or selected reference sequences. The reads were also *de novo* assembled and the resulting contigs were mapped to the selected reference sequences (for nanopore approaches as described above, for Illumina approach refer to **Supplementary Table S3–S5** for details about the used parameters). Finally, the fractions of reference sequences covered by reads or contigs for different subsamples were calculated and visualized as line or bar plots.

Longer reads or contigs mapping to small, circular genomes, could influence mapping efficiency (Visser et al., 2016). All mappings were performed to an artificially constructed viroid sequence, which was made by joining 10 repeated viroid genome sequences. Parameters for mapping contigs from Illumina rRNA-depleted totRNA dataset to viroids were adjusted (for samples I and IV). The length fraction parameter [CLC Genomic Workbench (v12, v21)] was set to 0.5 (50 %) instead of the 0.90 (90 %) used for viruses. Every mapping was individually inspected, and in cases where contigs were longer than the reference sequence, the fraction of reference covered by contig was reported as 100%.

In the case of sample I, in some subsamples, uneven coverage of the two different PepMV strain genome sequences was observed; thus, we performed additional analyses to test whether those observations are the consequence of the presence of two strains of the same virus present in the dataset. For each sequencing approach, one subsample (in which unequal contig coverage was observed for the two viral strains) was chosen (**Figure 4C**, see *). For each of those chosen subsamples, further analyses were implemented: (i) mapping original subsampled reads to the strain better covered by contigs in the original analysis, (ii) *de novo* assembly of the unmapped reads, and (iii) mapping newly assembled contigs to the corresponding reference genome sequence of the other present PepMV strain.

Estimating Data Volume and Analysis Time Needed for Virus/Viroid Detection by Nanopore Sequencing Using MinION Device

To obtain the proxy of the speed of virus or viroid detection and compare different employed MinION sequencing approaches, the sequencing time needed to obtain 50% of viral or viroid genome covered by generated reads was calculated. For this, a script (get_cumulative_yield_table.py, **Supplementary Data 3**) was used to calculate the cumulative yields of reads in gigabases for every 10 min of sequencing from MinION basecalling output files (summary.txt, refer to **Supplementary Table S8**.) According to the rarefaction analysis from section Comparison of the Performance of Nanopore and Illumina Sequencing For Detection of Viruses on Rarefied Datasets (the number of nucleotides estimated to cover more than 50% of a reference viral genome) for each virus or sequencing type combination, the time point at which this was reached was calculated according to the cumulative yields of the reads during sequencing.

RESULTS

Nanopore Sequencing Using MinION Device Resulted in Comparable Detection of Plant Viruses/Viroids as Illumina Sequencing

Using the pipeline described above for detection of viral sequences, all viruses except one were detected in the samples employing the MinION direct RNA sequencing approach (**Table 1**). Using this approach, ToMV, which was present in sample I in an extremely low titer (**Figure 2**, **Figure 3A**), was not detected. TASVd (sample IV) was at first not detected by mapping reads to the viral RefSeq database; however, when using a more closely related sequence from NCBI GenBank (KY810784) as the reference, a few reads of this viroid were detected.

When using the three additional MinION sequencing approaches: direct RNA sequencing of rRNA-depleted totRNA, cDNA-PCR sequencing of totRNA, and cDNA-PCR sequencing of rRNA-depleted totRNA, all of the viruses or viroids present in the samples were detected (**Figure 2**.)

Performance Comparison of MinION Direct RNA Sequencing and Illumina rRNA-Depleted totRNA Sequencing

The analysis of complete datasets showed that Illumina sequencing of rRNA-depleted totRNA resulted in a markedly higher relative fractions of viral reads compared to MinION direct RNA sequencing of totRNA (**Figure 3A**), which was

Sequencing approach	Sample number	Virus (Baltimore classification)	% of mapped reads	Average depth (reads)	Fraction of reference covered (reads)	Consensus sequence identity (%)	Average MinION reads identity	Average MinION contigs identity
Illumina sequencing rRNA-depleted totRNA	I	TYLCV (ssDNA)	0.0809	9.15	68.68	100.00	NA	
		ToCV (ssRNA+)	0.1584	31.87	99.22	100.00		
		PepMV-CH (ssRNA+)	12.3547	4787.66	100.00	98.96		
		PepMV-EU (ssRNA+)	10.1102	3905.12	100.00	98.51		
		ToMV (ssRNA+)	0.0139	5.90	92.26	99.80		
		STV (dsRNA)	0.0069	4.01	87.66	99.89		
		CLVd (viroid)	0.5523	2087.28	100.00	100.00		
	II	CaMV (dsDNA-RT)	3.4109	6387.25	100.00	99.99		
	III	CCyV1 (ssRNA-)	0.0381	42.26	99.90	100.00		
	IV	TSWV (ssRNA-)	4.5122	5173.80	99.99	99.97		
	V	TASVd (viroid)	0.0083	244.95	100.00	100.00		
		PVeV2 (dsRNA)	0.0116	11.52	99.84	99.99		
MinION direct RNA sequencing of totRNA	I	TYLCV (ssDNA)	0.0020	3.51	54.19	97.26	89.63	95.26
		ToCV (ssRNA+)	0.0007	0.26	28.59	/	90.53	/
		PepMV-CH (ssRNA+)	1.1528	2807.55	99.75	98.74	86.05	96.77
		PepMV-EU (ssRNA+)	0.5638	1289.09	99.64	98.35	85.45	95.64
		ToMV (ssRNA+)	0.0000	0.00	0.00	/	/	/
		STV (dsRNA)	0.0001	0.06	5.70	/	90.82	/
		CLVd (viroid)	0.0005	2.06	91.18	89.46	88.65	/
	II	CaMV (dsDNA-RT)	0.7548	429.28	99.70	99.61	88.21	97.16
	III	CCyV1 (ssRNA-)	0.0022	0.71	51.79	87.56	89.44	91.83
	IV	TSWV (ssRNA-)	0.2101	139.94	99.72	99.65	88.53	97.89
	V	TASVd (viroid)	0.0002	1.04	86.46	81.58	83.80	/
		PVeV2 (dsRNA)	0.0043	8.01	99.68	99.05	87.64	97.20
MinION direct RNA sequencing of rRNA-depleted totRNA	I	TYLCV (ssDNA)	0.0006	0.43	39.55	87.55	88.81	/
		ToCV (ssRNA+)	0.0442	15.08	99.60	98.52	87.71	97.03
		PepMV-CH (ssRNA+)	0.6243	685.49	99.77	98.78	87.30	97.25
		PepMV-EU (ssRNA+)	0.4552	415.00	99.66	98.37	86.74	/
		ToMV (ssRNA+)	0.0097	13.27	91.37	93.27	87.68	92.75
		STV (dsRNA)	0.0030	5.56	96.13	98.27	84.76	95.74
		CLVd (viroid)	0.0063	19.10	99.47	99.12	89.26	92.75
MinION cDNA-PCR sequencing of totRNA	I	TYLCV (ssDNA)	0.0486	17.15	58.97	99.40	89.38	98.35
		ToCV (ssRNA+)	0.0773	4.87	91.97	95.51	89.48	96.32
		PepMV-CH (ssRNA+)	27.1932	12346.89	99.89	98.89	87.07	93.96
		PepMV-EU (ssRNA+)	15.7794	6553.26	99.95	98.52	86.73	94.99
		ToMV (ssRNA+)	0.0847	50.77	93.58	99.78	87.38	99.52
		STV (dsRNA)	0.0006	0.09	9.11	/	90.43	/
		CLVd (viroid)	0.0040	3.14	99.73	93.57	90.85	93.16
MinION cDNA-PCR sequencing of rRNA-depleted totRNA	I	TYLCV (ssDNA)	0.0132	8.00	65.44	99.60	90.93	98.53
		ToCV (ssRNA+)	0.1303	24.47	99.87	99.86	91.38	99.73
		PepMV-CH (ssRNA+)	11.2630	6788.83	99.88	98.86	90.90	97.24
		PepMV-EU (ssRNA+)	9.7091	4884.39	99.94	98.46	90.36	96.61
		ToMV (ssRNA+)	0.0452	32.33	99.91	99.88	91.61	99.69
		STV (dsRNA)	0.0029	1.49	87.23	94.21	93.91	96.38
		CLVd (viroid)	0.0054	10.24	99.73	99.42	91.52	95.08

% of mapped reads	Average depth (reads)	Fraction of reference covered (reads)	Consensus sequence identity (%)	Average MinION reads identity	Average MinION contigs identity
0	0	0	81	81	81
0.01	1	10	82	82	82
0.025	5	20	83	83	83
0.05	10	30	84	84	84
0.075	15	40	85	85	85
0.1	25	50	86	86	86
1	50	60	87	87	87
3	75	70	88	88	88
6	100	80	89	89	89
9	1000	90	90	90	90
11	1500	91	91	91	91
13	3000	92	92	92	92
15	4500	93	93	93	93
17	6000	94	94	94	94
19	7500	95	95	95	95
21	9000	96	96	96	96
23	10500	97	97	97	97
25	12000	98	98	98	98
27	13500	99	99	99	99
29	15000	100	100	100	100

Heatmap is calculated for every section separately (% of mapped reads / Average depth (reads) / Fraction of reference covered (reads) / Consensus sequence identity (%) / Average MinION reads identity / Average MinION contigs identity) across all sequencing approaches. For every calculation (section - including legend) yellow represents the lowest value in the section, green represents the 50 percentile between the lowes and the highest value, and blue represents the highest value in the section.

FIGURE 2 | Heatmap of the samples included in the comparison with corresponding results: percentage of mapped reads, average depth (reads), fraction of reference covered (reads), consensus sequence identity (%), and average MinION reads/contigs identity (/ no data; NA not applicable).

expected due to the inclusion of the ribosomal RNA depletion step in the Illumina approach. These differences were reflected also when performing additional analyses on rarefied datasets (**Figure 3**); however, they were dependent on the amount of virus reads present in the original dataset.

As noted in section Nanopore Sequencing Using MinION Device Resulted in Comparable Detection of Plant Viruses/Viroids as Illumina Sequencing, no reads of ToMV were detected by MinION direct RNA sequencing of totRNA even in the complete dataset. Comparisons of rarefied subsamples further showed that MinION direct RNA sequencing performed comparably well to Illumina sequencing in cases in which a high fraction of specific viral reads was present in the samples (**Figure 3**). For MinION direct RNA datasets, in which virus sequences were present at more than 0.5% (PepMV in sample I and CaMV in sample II), relatively high fractions of corresponding viral genomes were covered by reads and contigs even in the smallest subsamples (**Figure 3**). Rarefaction analysis also showed very similar performance between the two approaches for PVeV2 (sample V)–in both cases, the fraction of genome covered by reads or contigs dropped correspondingly with the decreased dataset sizes. For TSWV, with 0.2% of the reads mapped to the viral genome, sharp drops in fractions of reference covered by reads or contigs were observed at smaller subsample data sizes for MinION direct RNA sequencing approach. For TYLCV, none of the two approaches enabled reconstruction of the complete genome, and Illumina sequencing approach performed only slightly better considering the two investigated parameters. In this analysis, for the remaining viruses and viroids (ToMV, ToCV, CCyV1, STV, CLVd, and TASVd), the MinION direct RNA sequencing method resulted in lower fraction of the genome covered by reads and contigs when compared to Illumina sequencing. In several cases, *de novo* assembly did not produce any contigs for the corresponding viruses, even for the complete datasets (data not shown).

Choice of a Suitable MinION Sequencing Approach Can Improve Detection of Plant Viruses and Viroids

Direct RNA sequencing using the MinION enabled detection of most of the plant viruses or viroids that were previously confirmed in the same samples with rRNA-depleted totRNA

A

	ssRNA+				ssRNA-		dsRNA		dsDNA -RT	ssDNA	circularRNA	
% of mapped reads	Pep-MV CH (Sample I)	Pep-MV EU (Sample I)	ToMV (Sample I)	ToCV (Sample I)	CCyV1 (Sample II)	TSWV (Sample III)	STV (Sample I)	PVeV2 (Sample V)	CaMV (Sample II)	TYLCV (Sample I)	CLVd (Sample I)	TASVd (Sample IV)
MinION seq. directRNA of totRNA	1.1528	0.5638	0.0000	0.0007	0.0022	0.2101	0.0001	0.0043	0.7548	0.0020	0.0005	0.0002
Illu. rRNA-dep. totRNA seq.	12.3547	10.1102	0.0139	0.1584	0.0381	4.5122	0.0069	0.0116	3.4109	0.0809	0.5523	0.0083

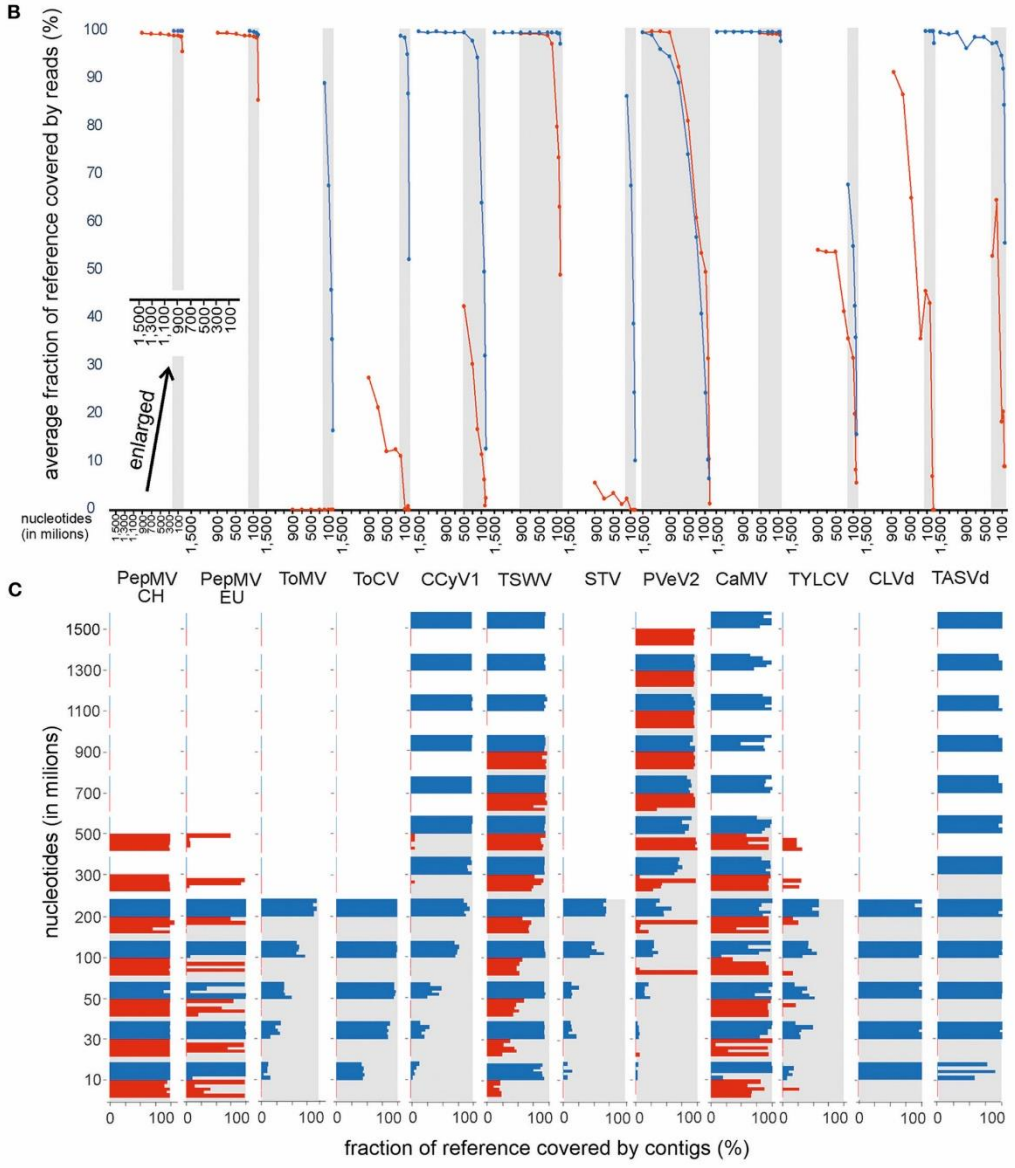

FIGURE 3 | Comparison of MinION direct RNA sequencing of totRNA (represented in red) and Illumina sequencing of rRNA-depleted totRNA (represented in blue) using data size-normalized subsamples. Results for each virus included in the analysis are shown along the x-axis and are grouped according to Baltimore

(Continued)

FIGURE 3 | classification. **(A)** Percentage of specific virus reads in trimmed and filtered complete HTS datasets. **(B)** Average fraction of reference covered by reads (%) at different subsample sizes. Dots represent the average value of analysis of 5 replicate subsamples. Different subsample sizes were used (10, 30, 50, 100, 200, 300, 500, 700, 900, 1,100, 1,300, 1,500 million nts–note the enlarged x-axis in the lower left part of panel 3B for a clearer view). **(C)** Fraction of reference covered by contigs (%) at different subsample sizes. Every bar represents the result of analysis for separate replicate subsamples. In **(B,C)**, gray areas designate the range in which the subsamples were available for both approaches compared.

Illumina sequencing. However, it showed somewhat inferior performance in the systematic comparisons described in a previous section. Thus, a single sample, containing five viruses and one viroid (sample I), was selected and used to further explore the performance of three additional sample preparation or nanopore sequencing library preparation approaches, which included either rRNA depletion, sequencing of PCR-amplified cDNA, or both (**Figure 1**), to investigate how much of the performance deficit is due to the sample preparation or library preparation method and how much is due to the platform.

Of these approaches, the one which is most comparable to the method used for Illumina sequencing (cDNA-PCR sequencing of rRNA-depleted totRNA) resulted in very similar fractions of viral reads in the sequenced datasets (**Figure 4**). This approach resulted in lower fractions of specific virus or viroid reads for almost all viruses than observed for both nanopore cDNA-based sequencing approaches or for the Illumina sequencing approach. Compared to MinION direct RNA sequencing without ribosomal RNA depletion, it resulted in an increased fraction of specific viral reads for four out of 7 viruses or viroids, including the detection of one virus (ToMV), which was not detected using direct RNA sequencing alone.

MinION sequencing of cDNA-PCR without ribosomal RNA depletion also resulted in relatively high fractions of specific viral reads. For four out of seven viruses or viroids, the numbers of corresponding reads were even higher than in the rRNA-depleted dataset sequenced by the same method (**Figure 4**.)

Moreover, even though for most viruses, the fractions of viral reads did not increase, when including rRNA depletion to direct RNA sequencing, the rarefaction analysis showed improved performance also in this case, for all but one virus, TYLCV (**Figure 4**.)

For viroids, the Illumina rRNA-depleted totRNA sequencing approach resulted in higher fractions of viroid reads than any of the MinION sequencing approaches (**Figures 3A**, **4A**.)

De novo assembly of reads from sample I, which contained two strains of PepMV in some subsamples, resulted in an assembly of contigs corresponding only to one strain (**Figures 3C**, **4C**.) The effect was observed, when using either Illumina or nanopore sequencing approaches. After the removal of the reads of one or the other PepMV strains and performing *de novo* assembly again (as described in Section Comparison of the Performance of Nanopore and Illumina Sequencing for Detection of Viruses on Rarefied Datasets), the artifact was no longer observed. Additionally, for some subsamples of MinION data, the assembled contigs were longer than the reference genome sequence (**Figure 4C**). Further investigation (visual inspection of nanopore reads mapping to the corresponding contigs) revealed mistakes or artifacts in some mapped reads, which likely led to artifactual *de novo* assembly of the corresponding contigs.

Different Nanopore Sequencing Approaches Using MinION Device Result in Different Accuracy of Reads; However, Generated Consensus Sequences Show Relatively High Accuracies in all Cases, When Sequencing Depths Are High Enough

Closer investigation and comparison of average read identities (proxy of sequencing error rate) for different MinION sequencing approaches for sample I revealed the lowest read identities when using direct RNA sequencing of rRNA-depleted totRNA (minimum 84.67%, average 87.43%). The highest average read identities were observed when using cDNA-PCR sequencing of the rRNA-depleted totRNA approach (minimum 90.36%, average 91.5%). For the same approach, also the highest average contig identities were observed (minimum 95.08%, average 99.73%). Even though the inclusion of the ribodepletion step resulted in a slightly decreased mean quality score (**Supplementary Table S2**), as was already observed in other studies (Wongsurawat et al., 2019), the calculation of average MinION read identity (**Figure 2**) did not show any marked differences.

As expected, the pairwise identities of consensus viral sequences compared to generated or selected references were higher in the case of Illumina sequencing approach for all viruses or viroids for which a calculation was possible (>99.5% in all cases). For MinION sequencing approaches, pairwise identities of consensus viral sequences compared to generated or selected references were higher than 98%, when average sequencing depth values were 5x or more, except in the case of ToMV (sample I), where, despite the sequencing depth of 13.27, this was 93.27%. Upon visual inspection of the mapping files, we observed that this was a consequence of uneven coverage by reads (the 5‘ of the viral genome was covered by very few reads), which contributed to the final lower average pairwise identity of consensus viral genome sequence.

Rapid Generation of MinION Data Needed for Detection of Viruses Present in Plants in Moderate Titres

A relatively short time (>30 min) was needed for retrieving sufficient data (covering at least 50% of viral genomic sequences), in cases, in which fractions of specific viral reads in samples were higher than 0.2% (PepMV, CaMV, TSWV; **Figures 3**, **4**; **Table 2**) using any of the sample preparation approaches. Further, in the case of PVeV2, for which 0.0043% of reads in the sample were mapped to the virus reference sequence, ∼2.5 h was needed to retrieve sufficient amount of data. For cases, in which virus or viroid reads were present in samples in very small fractions, a

A

	ssRNA+				dsRNA	ssDNA	viroid
% of mapped reads	PepMV CH	PepMV EU	ToMV	ToCV	STV	TYLCV	CLVd
MinION directRNA seq. of totRNA	1.1528	0.5638	0.0000	0.0007	0.0001	0.0020	0.0005
MinION directRNA seq. of rRNA-dep. totRNA	0.6243	0.4552	0.0097	0.0442	0.0030	0.0006	0.0063
MinION cDNA- PCR seq. of totRNA	27.1932	15.7794	0.0847	0.0773	0.0006	0.0486	0.0040
MinION cDNA-PCR seq.of rRNA-dep. totRNA	11.2630	9.7091	0.0452	0.1303	0.0029	0.0132	0.0054
Illumina rRNA-dep. totRNA seq.	12.3547	10.1102	0.0139	0.1584	0.0069	0.0809	0.5523

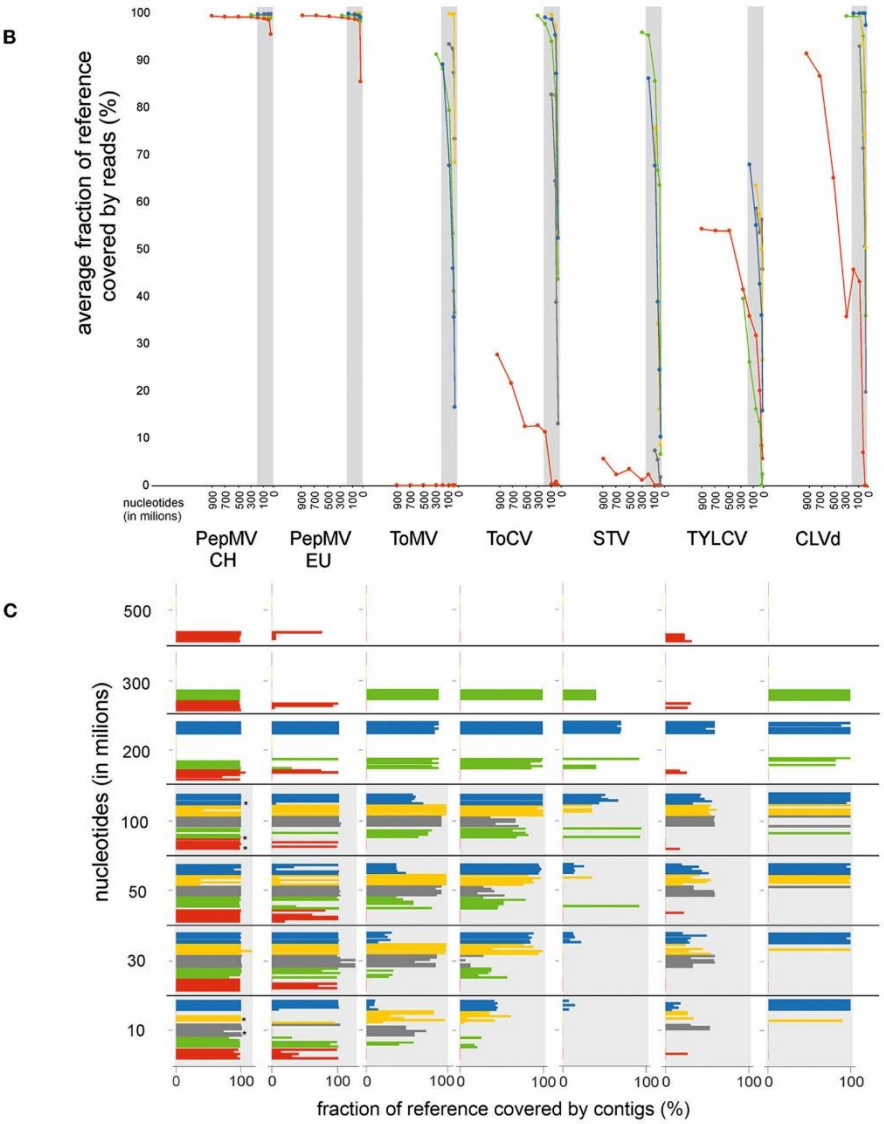

FIGURE 4 | Comparison of MinION sequencing: direct RNA sequencing of totRNA, direct RNA sequencing of rRNA-depleted totRNA, cDNA-PCR sequencing of totRNA, cDNA-PCR sequencing of rRNA-depleted totRNA and Illumina sequencing of rRNA-depleted totRNA using data size-normalized subsamples. Results for

(Continued)

FIGURE 4 | each virus included in the analysis are shown along the x-axis and are grouped according to Baltimore classification. **(A)** Percentage of specific virus reads in trimmed and filtered complete HTS datasets. **(B)** Average fraction of reference covered by reads (%) at different subsample sizes. Dots represent the average value of analysis of 5 replicated subsamples. Different subsample sizes were used (10, 30, 50, 100, 200, 300, 500, 700, 900, 1,100, 1,300, 1,500 million nts). **(C)** Fraction of reference covered by contigs (%) at different subsample sizes. Each bar represents the result of analysis for a separate replicate subsample. Bars with * indicate chosen subsample for additional analysis explained in section Comparison of the Performance of Nanopore and Illumina Sequencing for Detection of Viruses on Rarefied Datasets. In **(B,C)**, red represents MinION direct RNA sequencing of totRNA, green represents direct RNA sequencing of rRNA-depleted totRNA, gray represents cDNA-PCR sequencing of totRNA, yellow represents cDNA-PCR sequencing of rRNA-depleted totRNA, and blue represents the results for Illumina sequencing of rRNA-depleted totRNA. Gray areas designate the range in which the subsamples were available for both compared approaches.

TABLE 2 | Cumulative yield of nucleotides sequenced in time.

Sample number	Virus/viroid present	Nucleotides (millions)	Time (minutes)	Nucleotides (millions)	Time (minutes)	Nucleotides (millions)	Time (minutes)	Nucleotides (millions)	Time (minutes)
		MinION direct RNA sequencing of totRNA		MinION direct RNA sequencing of rRNA-depleted totRNA		MinION cDNA-PCR sequencing of totRNA		MinION cDNA-PCR sequencing of rRNA-depleted totRNA	
I	TYLCV	500	520	/	/	30	50	50	220
	ToCV	/	/	30	80	50	130	10	20
	PepMV	10	10	10	30	10	10	10	20
	ToMV	/	/	50	140	10	10	10	20
	STV	/	/	30	80	/	/	100	600
	CLVd	700	710	30	80	30	50	10	20
II	CaMV	10	10	NA	NA	NA	NA	NA	NA
	CCyV1	/	/	NA	NA	NA	NA	NA	NA
III	TSWV	20	30	NA	NA	NA	NA	NA	NA
IV	TASVd	200	220	NA	NA	NA	NA	NA	NA
V	PVeV2	200	170	NA	NA	NA	NA	NA	NA

Data in table represent the time when enough nucleotides would be sequenced to cover at least 50% of virus/viroid reference sequence by mapping reads. / no data: 50% coverage of virus/viroid reference by reads were not achieved. NA, not applicable.

relatively long time was needed to retrieve sufficient amount of data; e.g., for CLVd (0.0005% of reads were mapped to viroid reference sequence), the sequencing should last at least around 12 h to retrieve enough data. Equivalent observations were made for several other viruses or viroid, sequenced within the sample I (**Table 2**). In some cases, especially when very low fractions of specific viral reads were observed in the samples, retrieving enough data to cover >50% of genome sequence was not achieved even if the sequencing lasted for 46–48 h (**Table 2**.)

DISCUSSION

The systematic comparison of different approaches of nanopore and Illumina sequencing performed in this study demonstrated the effectiveness of nanopore sequencing using the MinION platform (Oxford Nanopore Technologies) for fast and sensitive detection of plant viruses, when the most optimal library preparation approach is used. Besides the ability to detect viruses, the accuracy and time efficiency of the approach were evaluated.

All viruses present in investigated samples, except ToMV, which was present in a sample in extremely low titer (Pecman et al., 2017), were identified using all of the employed approaches. ToMV was not detected using direct RNA sequencing of totRNA with the MinION sequencer (**Figure 2**). In general, across all datasets, inferior performance of direct RNA sequencing of totRNA using MinION compared to Illumina-based rRNA-depleted totRNA sequencing for detection of plant viruses is evident: between 10.7 and 1,104.6 (PepMV and CLVd, respectively) times, fewer virus reads (average 244.8) were observed in the complete direct RNA MinION datasets compared to the Illumina datasets. Rarefaction analyses showed a similar picture: the fractions of reference viral genomes covered by viral reads or contigs dropped markedly with reduced dataset sizes for direct RNA MinION for viruses, which were present in the original datasets in low amounts (e.g., below 0.2%) (**Figure 3**.) Moreover, in some cases, no contigs were recovered after *de novo* assembly for these viruses (**Figure 3C**), which indicates that the approach would not be very efficient for detection of (new) viruses present in plants in low titres.

We primarily included direct RNA sequencing of totRNA on MinION in this comparative study due to the speed and straightforward nature of this approach (according to SQK-RNA002, only 115 min is needed for library preparation). However, results of this first comparison showed that such an approach has a reduced performance for detection of plant viruses, compared to the Illumina-based sequencing of rRNA-depleted totRNA, which we take as a 'golden standard' in this study. This is likely due to both the high error rate of the sequencing itself and the lack of the rRNA depletion step. Thus, in a second part of the study, we performed analysis of a selected sample (containing a diverse assembly of plant viruses)

using nanopore sequencing MinION device, but with several improvements, which were shown to improve the performance of nanopore sequencing for virus detection.

Including rRNA depletion prior to the direct RNA sequencing approach using the MinION platform resulted in increased fractions of specific virus reads in most of the cases (**Figure 2**.) The fractions of specific virus reads changed between 0.3 and 63.1 (TYLCV and ToCV, respectively) times, and on average, the fraction increased by 17.8 times for all viruses tested.

The protocol using cDNA-PCR sequencing of totRNA (with no rRNA depletion) also resulted in improvements in the fractions of specific virus reads in the dataset (**Figure 2**). The fractions of virus reads increased between 6.0 and 110.4 (STV and ToCV, respectively) times compared to direct RNA sequencing of totRNA and on average 33.3 times for all viruses tested. For three viruses (PepMV-EU, PepMV-CH, and ToMV), this approach resulted in a greater proportion (1.5x, 2.2x, and 6.09x, respectively) of virus sequences in the MinION dataset than sequencing rRNA-depleted totRNA using the Illumina platform.

Moreover, incorporating both rRNA depletion and reverse transcription prior to nanopore sequencing using MinION platform (i.e., using cDNA-PCR sequencing of rRNA-depleted totRNA) led to the greatest increases in the observed fractions of specific virus reads. The proportion of virus reads increased between 6.6 and 186.1 (TYLCV and ToCV, respectively) times and on average 43.2 times for all viruses tested compared to direct RNA sequencing. For one virus (ToMV), this approach resulted in a greater proportion (3.2 x) of virus sequences obtained in the respective MinION dataset compared to the dataset generated from rRNA-depleted totRNA using the Illumina platform. Using cDNA-PCR protocol for sequencing rRNA-depleted totRNA also resulted in the highest observed consensus sequence identities, and the comparison of calculated average viral reads identities revealed that this approach resulted in most "accurate" sequencing reads (**Figure 2**.)

As mentioned above, in a number of cases, no contigs of the corresponding viruses were generated from complete (data not shown) or near complete MinION direct RNA datasets (**Figures 3**, **4**). This was also observed for two viruses (TYLCV, PepMV-EU) when direct RNA sequencing of rRNA-depleted RNA was used and one virus (STV) when cDNA-PCR sequencing of totRNA was used. This suggests that detection of such viruses could be missed if only contigs are analyzed, and also, that unknown viruses would be missed if present in samples in similarly low titres. Using the cDNA-PCR protocol for sequencing rRNA-depleted totRNA, contigs were assembled for all viruses in the datasets tested and they showed the highest average contig identity of all of the nanopore sequencing approaches compared (**Figures 2**, **4**.)

The performance of the direct RNA sequencing of totRNA using MinION was notably poorer for viroids than viruses. Compared to Illumina sequencing, there was 1104,6x and 41,5x smaller proportion of CLVd and TASVd viroid reads when sequencing totRNA directly using the MinION and for both viroids included in the comparison, only partial genomes were obtained. When using either cDNA-PCR sequencing of rRNA-depleted RNA or direct RNA sequencing of rRNA-depleted totRNA, complete viroid genomes were recovered (**Figure 4**.) Inferior performance of the direct sequencing of totRNA for the detection of viroids is most likely a consequence of the circular genome which could not be polyadenylated in the first steps of the protocol. Only damaged or intermediate replication forms of viroids could be polyadenylated. Moreover, it is possible that the nanopore sequencing could be adversely affected by the secondary structure of viroid RNA (Flores et al., 2014), and this may be partly overcome by the larger amounts of target present in rRNA-depleted totRNA. This study is the first using nanopore sequencing for viroids, and further improvements in template preparation, such as fragmentation of the input RNA, could be envisaged in this case.

Regardless of the approach used (Illumina or all of the tested template preparation methods using the MinION device), we observed artifacts in an assembly of PepMV genomes. PepMV was present in sample I in a mixed infection of two strains. In several cases, the genome of one of the two strains was not assembled in the process. After subtracting the reads of one of the strains and repeating the *de novo* assembly, the artifact was no longer observed. This suggests that special attention should be applied to analysis pipelines to resolve observed assembly artifacts to ensure detection of multiple strains of the same virus.

Different types of input material could possibly also affect the quality and the amount of the generated sequencing data. Most of the samples used in this study were frozen leaves; however, also, fresh leaves and the dry seed samples were used in some cases. Although in our study, the sample size for different input materials is not large enough to draw any conclusions, we observed the highest amount of generated data when using fresh leaf material (**Supplementary Table S1**). More tests would be needed to study the impact of the input material on the sequencing results and likely specific adaptations of extraction and sample preparation procedures could be implemented to ascertain optimal results for different sample types.

One of the main advantages of the nanopore sequencing is the speed with which sequence data are generated (Bronzato et al., 2018; Fellers et al., 2019). In this study, the libraries for MinION sequencing were prepared and applied to the flow cell within 1 day. Since the data could be analyzed in real time, estimations of the time needed to generate sequences, which would cover 50% or more of a specific viral genome using read mapping were made. The time in which these thresholds were achieved depended on the amount of virus reads present in the complete datasets (a proxy of viral titer). Our estimates suggest the 50% threshold would be achieved in 10–30 min for PepMV using any of the approaches, or on the other end of the spectrum in 11 h and 50 min for CLVd (using MinION direct RNA sequencing of totRNA). This suggests that a diagnostic workflow could be established with <1 day to perform RNA extraction and library preparation, followed by an overnight run on the MinION device and bioinformatic analysis of the following day.

The speed and accessibility of the methodology has led to exploration of the technique as an in-field diagnostic tool (Boykin et al., 2019). Though possible for the detection of high-titer viruses using rapid sample preparation and simplified analysis, this would currently rely on transferring several items

of laboratory equipment to the site where testing is being performed and would also need the end-users to be skilled in molecular biology protocols. For diagnostic applications where potentially low-titer infections are likely, the method would require significantly more time to run the flow cell and analyse the data, making it less practical for in-field use. Furthermore, adoption as a field test is dependent on the time criticality of the actions taken based on the outcome of the test and how these will be improved if the results are generated more quickly. The practical benefits described do make the approach suited to routine diagnostic laboratories, where the larger sequencers may be too expensive and impractical to run, leading laboratories to rely on outsourcing to HTS providers. The speed and scalability of MinION sequencing make it well suited to smaller numbers of samples in diagnostic laboratories and, in particular, where rapid turnaround of results is needed especially, given the results are approaching the quality generated by Illumina sequencing.

One of the challenges for introducing this method into routine laboratory use is the constant and rapid development, introduction and withdrawal of flow cells, kits, protocols, and bioinformatic tools by Oxford Nanopore Technologies. This leads to uncertainty with incorporating nanopore into routine testing protocols, in particular those run within a quality certification scheme (e.g., ISO17025). To overcome this obstacle, the use of internal negative control (healthy plant) and in particular a standardized positive control suited for the entire workflow is needed. For example, *Phaseolus vulgaris*, cv. Black Turtle infected with endornaviruses has been used as a positive control in other studies (Kesanakurti et al., 2016) and was successfully sequenced in this study (Sample V).

Finally, when comparing the costs for sample, library preparation, and sequencing per sample using either MiSeq (Illumina) or MinION flow cell (Oxford Nanopore Technologies), comparable prices can be estimated if using multiplexing of samples. In the case of MiSeq (Illumina) sequencing, the estimated cost is 189 €/sample if sequencing 24 samples in the same run (Vazquez-Iglesias et al., 2022). On the other hand, the estimated cost for barcoding and sequencing 12 or 24 samples on one MinION flow cell using cDNA-PCR barcoding kit and including ribodepletion step is 215 €/sample or 170 €/sample, respectively. Due to the increasing throughput of the MinION flow cells, both multiplexing options should now provide enough data for reliable detection of most of the viruses present in the tested samples. More detailed calculation is described in **Supplementary Data 4**.

To conclude, the results of this study indicate that, when appropriate library preparation and sequencing protocols are selected, nanopore sequencing using the MinION device gave equivalent detection of a range of viruses and viroids than a commonly used Illumina sequencing approach. In this regard, the performance of MinION direct RNA sequencing of totRNA was lower than Illumina sequencing, but improved significantly when rRNA depletion was incorporated or when cDNA amplification or both were incorporated. Whilst these slowed down the sample preparation, they facilitated detection of the lowest titer virus infection included in the study.

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

AP, DK, MR, and NB designed the experiment. AP performed laboratory part of the experiment and analyzed the data with the assistance of IA and DK and wrote the draft of the manuscript. All authors significantly contributed with reviewing and editing the manuscript.

FUNDING

The study was financially supported by Slovenian Research Agency thought project NanoPhyto (L7-2632), core financing (P4-0165, P4-0407), and young researcher grant (AP), Euphresco (VIRFAST) project and by COST Action CA15223, thought STSM (short-term scientific mission).

ACKNOWLEDGMENTS

We thank Sam McGreig for the help with bioinformatic analysis, i.e., sharing tested parameters for *de novo* assembly of MinION reads, Mike Rott for seeds of *Phaseolus vulgaris*, cultivar Black Turtle, in this study analyzed as Sample V, and Wouter De Coster for script: get_cumulative_yield_table.py (**Supplementary Data 3**.)

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.883921/full#supplementary-material

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

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

Beddoe, N., Adams, I. P., McGreig, S., and Forsyth A. (2020). Identification and full genomic sequence of nerine yellow stripe virus. *Arch. Virol.* 16, 2967–2971. doi: 10.1007/s00705-020-04776-3

Boykin, L. M., Sseruwagi, P., Alicai, T., Ateka, E., Mohammed, I.U., Stanton, J.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

Branton, D., and Deamer, D. (2019). "The Development of Nanopore Sequencing" in *Nanopore Sequencing*, eds. D. Branton, and D. Deamer (London: World Scientific, Pub Co Inc. doi: 10.1142/10995

Bronzato, B. 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, 1648–52. doi: 10.1094/PDIS-04-17-0488-RE

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods.* 12, 59–60. doi: 10.1038/nmeth.3176

Chalupowicz, L., Dombrovsky, A., Gaba, V., Luria, N., Reuven, M., Beerman, A., et al. (2019). Diagnosis of plant diseases using the nanopore sequencing platform. *Plant Pathol.* 68, 229–238. doi: 10.1111/ppa.12957

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. 34, 2666–2669. doi: 10.1093/bioinformatics/bty149

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

Fellers, J. P., Webb, C., Fellers, M. C., Shoup Rupp, J., and Wolf. E. De. (2019). Wheat virus identification within infected tissue using nanopore sequencing technology. *Plant Dis*. 103, 2199–2203 doi: 10.1094/PDIS-09-18-1700-RE

Filloux, D., Fernandez, E., Loire, E., Claude, L., Galzi, S., Candresse, T., et al. (2018). Nanopore-based detection and characterization of yam viruses. *Sci. Rep.* 8. doi: 10.1038/s41598-018-36042-7

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

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., et al. (2018). Highly Parallel Direct RNA Sequencing on an Array of Nanopores. *Nat. Method*. 15, 201–206. doi: 10.1038/nmeth.4577

Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., et al. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* 7, 1–13. doi: 10.1186/s13073-015-0220-9

Huson, D. A., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenome data. *Gennome Res.* 17, 377–386. doi: 10.1101/gr.5969107

Ip, C., Loose, M., Tyson, J. R., de Cesare, M., Brown, B. L., Jain, M., et al. (2015). MinION analysis and reference consortium: phase 1 data release and analysis. *F1000Research*. 4, 1075. doi: 10.12688/f1000research.7201.1

Jones, R. A. C., and Naidu, R. A. (2019). Global dimensions of plant virus diseases : current status and future perspectives. *Annu. Rev. Virol.* 6, 387–409. doi: 10.1146/annurev-virology-092818-015606

Kesanakurti, P., Belton, M., Saeed, H., Rast, H., Boyes., I., et al. (2016). Screening for plant viruses by next generation sequencing using a modified double strand RNA extraction protocol with an internal amplification control. *J. Virol. Methods*. 236, 35–40. doi: 10.1016/j.jviromet.2016.07.001

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

Li, H. (2016). Sequence analysis minimap and miniasm : fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*. 32, 2103–2010. doi: 10.1093/bioinformatics/btw152

Li, H. (2018). Minimap2 : pairwise alignment for nucleotide sequences. *Bioinformatics*. 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Meredith, L. W., Hamilton, W. L., Warne, B., Houldcroft, C. J., Hosmillo, M., Jahun, A. S., et al. (2020). Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* 20, 1263–1272. doi: 10.1016/S1473-3099(20)30562-4

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. doi: 10.3389/fmicb.2017.01998

Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., et al. (2017). Multiplex PCR method for minion and illumina sequencing of zika and other virus genomes directly from clinical samples. *Nat. Protocols*. 12, 1261–1266. doi: 10.1038/nprot.2017.066

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., et al. (2016). Ebola surveillance. *Nature*. 65, 35–43. doi: 10.1038/nature16996

Tan, S., Dvorak, C. M.T., and Murtaugh, M. P. (2019). Rapid, unbiased PRRSV strain detection using MinION direct RNA sequencing and bioinformatics tools. *Viruses*. 11. doi: 10.3390/v11121132

Van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet*. 34, 666–681. doi: 10.1016/j.tig.2018.05.008

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Vazquez-Iglesias, I., McGreig, S., Pufal, H., Robinson, R., Clover, G. R. G., Fox, A., et al. (2022). A novel high-throughput sequencing approach reveals the presence of a new virus infecting Rosa : Rosa Ilarvirus-1 (RIV-1). *J. Virol. Methods*. 300. doi: 10.1016/j.jviromet.2021.114417

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

Visser, M., Bester, R., Burger, J. T. and Maree, H. J. (2016). Next-generation sequencing for virus detection: covering all the bases. *Virol J.* 13, 85. doi: 10.1186/s12985-016-0539-x

Wongsurawat, T., Jenjaroenpun, P., Taylor, M. K., Lee, J., Tolardo, A. L., Parvathareddy, J., et al. (2019). Rapid sequencing of multiple RNA viruses in their native form. *Front. Microbiol.* 10, 1–8. doi: 10.3389/fmicb.2019.00260

Yamagishi, J., Runtuwene, L. R., Hayashida, K., Mongan, A. E., Nguyen Thi, L. A., Nguyen Thuy, L., et al. (2017). Serotyping dengue virus with isothermal amplification and a portable sequencer. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-03734-5

Conflict of Interest: IA and AF are employed by Fera Science Ltd.

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.

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.

Chapter 3

Discussion

3.1 The Comparison of Plant Virus and Viroid Detection Using sRNA and rRNA-Depleted totRNA High Throughput Sequencing Approach Showed Comparable Results

We compared the effectiveness of two widely established HTS protocols for plant virus detection using Illumina technology: sRNA and rRNA-depleted totRNA. When comparing the amount of virus/viroid nucleotides recovered by one or the other protocol, we observed different results for different viruses/viroids: in some cases, more viral/viroid nucleotides were recovered using sRNA while in others it was rRNA-depleted totRNA sequencing the one who performed better (Pecman et al., 2017). Detailed inspection of the read mapping results suggested higher recovery of ssDNA virus and viroid nucleotides when using the sRNA protocol. In the case of viroids, this could be the consequence of different ration between induced RNA silencing (Itaya et al., 2001; Martínez de Alba et al., 2002; Papaefthimiou et al., 2001) and multiplication of the messenger RNA, as viroids do not code for proteins (Flores et al., 2014) and "long" RNAs are generated solely for the purpose of replication. Similarly, in the case of viruses with a circular ssDNA genome organization, longer RNA transcripts are generated only transiently during the transcription step, which could be the reason for the lower recovery of viral reads in the analysis of rRNA-depleted totRNA pool. In contrast, sRNAs could be amplified by the action of RNA- dependent RNA polymerase 6 (Borges & Martienssen, 2015) during the production of secondary sRNAs with a positive effect for the detection with the sRNA protocol. For most of the other investigated viruses, the sequencing of rRNA-depleted totRNA resulted in a larger proportion of reads mapping to the viral genomes compared with sRNA. However, a few exceptions were observed, potato virus Y being the most notable with many more viral reads present in the sRNA dataset (Pecman et al., 2017). The high abundance of virus-derived sRNA has been reported for PVY (Kutnjak et al., 2015) and other potyviruses (Kreuze et al., 2009) even though they encode strong RNA silencing suppressors (Ivanov et al., 2016; Yelina et al., 2002). Nevertheless, both protocols successfully identified two divergent strains of pepino mosaic virus, which, for the case of the sRNA protocol was previously confirmed for a closely related potexvirus (Kutnjak et al., 2014). When comparing the de novo assembly of sequencing reads, the rRNA-depleted totRNA protocol was generally more efficient than the sRNA protocol; this was confirmed by a higher proportion of viral genomes covered by de novo generated contigs from rRNA-depleted totRNA datasets (Pecman et al., 2017). Poorer coverage of viral genomes by de novo assembled contigs from sRNA reads is likely the consequence of more difficult assembly of

very short sRNA reads into longer contigs, which has been already observed (Kutnjak et al., 2015; Massart et al., 2018; Visser et al., 2016). An important difference between the compared protocols was observed when the genome assembly of a previously un-described cytorhabdovirus was attempted. The latter was identified from the rRNA-depleted totRNA reads, following de novo assembly and BLASTx analysis, whilst just virus reads could be found in the sRNA sequence data and only after a post-hoc analysis (Pecman et al., 2017).

3.2 High Throughput Sequencing Successfully Facilitate Novel Virus Strain Identification and Characterization

The knowledge about HTS applicability for generic plant virus detection and identification gained in the first study led us to its application for discovery and characterization of a novel strain of henbane mosaic virus. Here, firstly, HTS (sRNA and rRNA-depleted totRNA protocols) was used as a generic detection technique for identifying the presence of known and unknown viruses in a field sample, which resulted in the detection of HMV for the first time in tomato and the first time in any plant in Slovenia. Since the virus was present in a mixed infection and few genomic information was available at the time, it was probably overlooked by conventional diagnostic methods (Pecman et al., 2018). Secondly, HTS (sRNA protocol) enabled us to rapidly generate complete genomic sequences of several HMV isolates, followed by their genome annotation, pairwise comparison, and phylogenetic analyses, which taxonomically classified the HMV isolates within the *Potyvirus* genus. According to the sequence similarities with other *Potyvirus* species (Adams et al., 2005) all sequenced isolates belong to three different strain groups within the same species. The partially sequenced HMV-PHYS/H (AM184113.1) isolate detected in Hungary clustered to a 4th group. The observed sequence diversity among the low number of analyzed HMV isolates is surprisingly high and could be explained by several scenarios. The investigated isolates were from different countries, different host plants, and stored in collections at different decades over the previous 80 years (Pecman et al., 2018). "To estimate the current diversity among different HMV isolates in nature, additional studies, including several samples collected from the natural environment, from different host plants, and from a broader geographic range, should be completed" (Pecman et al., 2018). Susceptible hosts of an HMV-SI/L isolate, which expressed disease symptoms were closely related plant species, e.g., *H. niger*, *Nicotiana* spp. and *S. lycopersicum* (different varieties) (Pecman et al., 2018). According to our host range analysis and listed literature (Hamilton, 1932; Lovisolo & Bartels, 1970; Salamon, 2018), different isolates caused different symptoms in different hosts. "Nevertheless, all the previously reported host range analysis should be interpreted with caution. The composition of those inoculums was not confirmed with HTS, so they could have contained mixed infections of viral species, which could explain some of the contradictory results" (Pecman et al., 2018) of disease symptoms on different host plants from literature (Hamilton, 1932; Lovisolo & Bartels, 1970; Salamon, 2018). The host range analysis of HMV done in our study is the first one in which single infection of the inoculum was confirmed. HMV was separated from other co-infecting viruses, through repeated passages on different test plants. Furthermore, when comparing the observed symptoms of mixed infection from a field tomato sample with symptoms of HMV-SI/L single infection of tomato plant in the greenhouse, the symptomatology was different, possibly because the virus was present in mixed infection in the originally sampled tomato plant (Pecman et al., 2018). The study of HMV demonstrated the need for a quick generic diagnostic method, which could detect expected and unexpected emerging plant pathogens with limited or absent genomic information and demonstrated the importance of well-designed biological characterization of novel virus or strains identified by HTS.

3.3 Nanopore High Throughput Sequencing has a Potential for Plant Virus and Viroid Detection in a Reduced Time and with Potential for On-Site Deployment in Comparison to Illumina

As shown above, for improved investigation of complex relationship between the presence of different plant viruses and symptoms, fast detection followed by identification and characterization of the causing agent/s is needed. In plant pathology, nanopore technology was already successfully used for detection of bacteria, fungi, and viruses (Boykin et al., 2019; Bronzato Badial et al., 2018; Chalupowicz et al., 2019; Fellers et al., 2019). In some studies (Beddoe et al., 2020; Filloux et al., 2018; Vazquez-Iglesias et al., 2022), both nanopore and Illumina technology were used for plant virus detection; however, a systematic comparison between those two approaches for the detection of a wide array of viruses with different genome types was still lacking (Pecman et al., 2022). With this in mind, we decided to systematically compare the well-established Illumina technology, using benchtop sequencer MiSeq and rRNA-depleted totRNA as RNA input, and the continuously evolving nanopore platform (Oxford Nanopore Technologies), using a portable MinION sequencer. In the nanopore platform, we additionally compared four library preparation combinations as RNA input: direct RNA sequencing of total RNA, direct RNA sequencing of rRNA-depleted total RNA, cDNA-PCR sequencing of total RNA, and cDNA-PCR sequencing of rRNA-depleted total RNA. In the first comparison, all viruses were sequenced with nanopore technology, using MinION sequencer and direct RNA sequencing of totRNA as RNA input. The results of this first comparison showed that such an protocol has a reduced performance for detection of plant viruses, compared to the Illumina-based sequencing of rRNA-depleted totRNA, which is accepted as the "gold standard" (Pecman et al., 2022). Nanopore sequencing, while successfully detecting most of virus/viroids present in the samples, failed to identify one virus (tomato mosaic virus) that was present in one of the analyzed samples at extremely low titer (Pecman et al., 2017). Moreover, no contigs were recovered after de novo assembly for some viruses, which indicates that the direct RNA nanopore sequencing using totRNA protocol may not be efficient enough for the detection of (new) viruses present in plants in low titers (Pecman et al., 2022). Thus, in the second part of the study, we performed the analysis of a selected sample (containing a diverse assembly of plant viruses) using a nanopore sequencing MinION device but with several modifications, which were shown to improve the performance of nanopore sequencing for virus detection. Including rRNA depletion prior to the direct RNA sequencing protocol using nanopore technology (MinION sequencer) resulted in increased fractions of specific virus reads in the datasets in most of the cases. The protocol using cDNA-PCR sequencing of totRNA (with no rRNA depletion) also resulted in improved fractions of specific virus reads in the dataset (Pecman et al., 2022). "Moreover, incorporating both rRNA depletion and reverse transcription prior to nanopore sequencing using the MinION platform (i.e., using cDNA-PCR sequencing of rRNA-depleted totRNA) led to the greatest increases in the observed fractions of specific virus reads" (Pecman et al., 2022). This sequencing strategy also resulted in the highest observed consensus sequence identities and the comparison of calculated average viral reads identities revealed that this protocol resulted in the most "accurate" sequencing reads among all four nanopore protocols (Pecman et al., 2022). Furthermore, cDNA-PCR of rRNA-depleted totRNA was the only protocol that allowed contig assembly for all viruses in the tested datasets and the assembled contigs showed the highest identity when mapped to consensus sequence of

all of the nanopore sequencing protocols compared (see Figure 2, (Pecman et al., 2022)). Thus, based on our results, this protocol would be the most convenient for virus discovery applications. Finally, pioneer nanopore sequencing for viroid detection resulted in complete viroid genome recovery, when including rRNA depletion prior cDNA-PCR or direct RNA sequencing, although several improvements could be envisaged and possibly applied, e.g. random primed double-stranded cDNA as input for nanopore library preparation (Liefting et al., 2021). One of the main advantages of nanopore sequence is the rapidity of the procedure (Bronzato et al., 2018; Fellers et al., 2019). In our case, the libraries for MinION sequencing were prepared and loaded to the flow cell within one day; since the data can be analysed in real time, the estimation of the time needed to generate sequences, which would cover 50% or more of a specific viral genome using read mapping, was calculated. According to this estimation the longest time for reaching a 50% threshold would be achieved in 11 h and 50 min for CLVd (using MinION direct RNA sequencing of totRNA). The time for achieving this threshold, however, depends on the amount of virus sequences present in the complete datasets (a proxy of viral titer). Taking together, RNA extraction and library preparation could be performed with <1 day, followed by an overnight run on the MinION device and bioinformatic analysis on the following day (Pecman et al., 2022). Furthermore, due to the portability of the MinION sequencer, on-site deployment for detection of plant viruses can be implemented (Boykin et al., 2019), but the implementation of library preparation is still impractical. For on-site work, several items must be transferred from the laboratory to the site where testing is performed. In the case of potentially low-titer infections, the method would require more time to run the flow cell and analyse the data, making it less practical for end users. Nevertheless, the practical benefits described do make the protocol suited to diagnostic laboratories, where running larger sequencers may be too expensive and impractical, leading laboratories to outsource HTS services to different providers (i.e., Illumina) (Pecman et al., 2022). The speed, comparable sequencing cost to Illumina (see Publication 2.3), and scalability of nanopore sequencer (MinION) make it well suited to laboratories with smaller numbers of samples and, in particular, where rapid turnaround of results is needed.

Chapter 4

Conclusion

In this thesis, we started with the comparison of two most generic HTS protocols for plant virus/viroid detection on the Illumina platform. Next, the obtained knowledge was used for the identification and characterization of a new plant virus strain. Finally, with all acquired expertise, we proceeded further to expand the use of high throughput sequencing for plant virus/viroid detection to the faster and more accessible nanopore technology. The latter technology enables either on-site use or use in smaller laboratories near sampling sites. The outcomes from the first paper partially confirmed the first hypothesis, since by using both tested HTS protocols, sRNA and rRNA-depleted totRNA, all included known viruses/viroids were identified. However, a putative novel cytorhabdovirus was only detected by analyzing the data generated from rRNA-depleted totRNA and not by sRNA protocol, since the small amount of short reads prevented the de novo contig assembly. In the second paper, HTS was used for the identification of unexpected known viral species on a symptomatic field tomato plant. The successful detection of henbane mosaic virus with HTS thus confirmed the second hypothesis. Regardless, the identified known virus had limited genomic information, so in the next step, HTS enabled the first complete genome assembly and annotation of the henbane mosaic virus. Following the virus isolation, the biological characterization and host range analysis was also conducted confirming the third hypothesis. The results of last study confirmed the final, fourth hypothesis, indicating that, when appropriate library preparation and sequencing protocols are selected i.e., cDNA-PCR sequencing of rRNA-depleted totRNA, nanopore sequencing using the MinION device resulted in the equivalent detection of a range of viruses and viroids as with commonly used Illumina sequencing. According to the nanopore sequencing results, the time needed for nanopore HTS workflow was estimated to two working days, which indeed is shorter in comparison to Illumina. On-site use is possible, but still impractical since several items from laboratory need to be transferred to the site. Nevertheless, the scalability and cost range of nanopore MinION technology makes it accessible to most, including small diagnostic labs near sampling sites.

To conclude, the comparisons between different HTS platforms and sample preparation protocols described in this thesis and their application for novel virus strain identification and characterization, delivered useful findings that will serve as guide for plant virus (diagnostic) laboratories in the selection of an HTS protocol, appropriate for their specific application.

Moreover, the results of this doctorate raise some conclusions but also questions and consequently new ideas for future experiments and optimizations that should deal with:

i) Expanding the research on HMV to studying its prevalence, distribution and consequently its threat for tomato crop across Slovenia. This point was partially addressed, by the study conducted by Rivarez et al., 2022, where 436 samples of tomato plants and surrounded weeds were collected (in summer 2019 and 2020), sequenced with HTS and analyzed. The latter resulted in HMV detection in one bulk sample in 2020, demonstrating its presence but not high prevalence.

ii) Improvement of bioinformatics pipeline for detection of different strains of the same virus using nanopore sequencing datasets (as pointed out in Pecman et al., 2022). For

example, Diamond (Buchfink, et al., 2014) could be used for metagenomics analysis of reads and contigs for strain identification followed by BLASTN confirmation and additional analysis on reads and/or contigs if needed. Furthermore, the reads analysis using metagenomics approach may result in detection of new virus present in low reads amount, which may not be assembled into contigs. In this case further analysis for confirmation could be implemented subsequently.

iii) Improvement of viroid detection by nanopore technology through testing and comparison of different sample preparation protocols (i.e., enzymatic versus fragmentation by sonication), libraries (Liefting et al., 2021) and new bioinfomatics pipelines for viroid detection (Rivarez et al., 2022).

iv) Optimization of cost/sample ratio with introduction of barcodes (as discussed in Pecman et al., 2022) and/or pooling samples when using MinION flow cell for nanopore sequencing or testing outputs of Flongle flow cell on single sample.

v) Introducing nanopore technology for generic in-house diagnostics of plant viruses and viroids, which will help when fast results of lower number of samples are needed. Adaptation to plant virus diagnostics environment should be preceded by a proper assessment the sensitivity, repeatability, reproducibility and when applicable inclusivity/exclusivity of the approach.

vii) Assessing the need for additional controls (e.g., an internal, alien control) within HTS library preparation and sequencing for more reliable usage of HTS for the detection of plant pathogens and pests (Massart et al., 2022).

.

Appendix A

Supplementary Material of Included Publications

A.1 Supplementary Material for Publication 2.1

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.01998/full#supplementary-material

A.2 Supplementary Material for Publication 2.2

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02739/full#supplementary-material

A.3 Supplementary Material for Publication 2.3

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.883921/full#supplementary-material

Appendix B

Permission for Reproduction of Included Publications

B.1 Permission for Reproduction of Publication 2.1

Keywords: next generation sequencing, small RNA, ribosomal RNA depleted total RNA, detection, plant viruses, plant viroids

Citation: Pecman A, Kutnjak D, Gutiérrez-Aguirre I, Adams I, Fox A, Boonham N and Ravnikar M (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

Received: 17 July 2017; **Accepted:** 28 September 2017;
Published: 13 October 2017.

Edited by:
Guenther Witzany, Telos - Philosophische Praxis, Austria

Reviewed by:
Claudio Ratti, Università di Bologna, Italy
Carmen Hernandez, Instituto de Biología Molecular y Celular de Plantas (CSIC), Spain

***Correspondence:** Anja Pecman, anja.pecman@nib.si
Denis Kutnjak, denis.kutnjak@nib.si

B.2 Permission for Reproduction of Publication 2.2

Keywords: potyvirus, henbane mosaic virus, tomato, high-throughput sequencing, host range analysis, phylogeny

Citation: Pecman A, Kutnjak D, Mehle N, Žnidarič MT, Gutiérrez-Aguirre I, Pirnat P, Adams I, Boonham N and Ravnikar M (2018) High-Throughput Sequencing Facilitates Characterization of a "Forgotten" Plant Virus: The Case of a Henbane Mosaic Virus Infecting Tomato. *Front. Microbiol.* 9:2739. doi: 10.3389/fmicb.2018.02739

Received: 29 August 2018; **Accepted:** 26 October 2018;
Published: 19 November 2018.

Edited by:
John P. Carr, University of Cambridge, United Kingdom

Reviewed by:
Jesús Navas-Castillo, Instituto de Hortofruticultura Subtropical y Mediterránea La Mayora (IHSM), Spain
F Murilo Zerbini, Universidade Federal de Viçosa, Brazil

***Correspondence:** Anja Pecman, anja.pecman@nib.si Maja Ravnikar, maja.ravnikar@nib.si

B.3 Permission for Reproduction of Publication 2.3

Keywords: high-throughput sequencing, plant virus/viroid detection, comparison, nanopore MinION sequencing, Illumina MiSeq sequencing

Citation: Pecman A, Adams I, Gutiérrez-Aguirre I, Fox A, Boonham N, Ravnikar M and Kutnjak D (2022) Systematic Comparison of Nanopore and Illumina Sequencing for the Detection of Plant Viruses and Viroids Using Total RNA Sequencing Approach. *Front. Microbiol.* 13:883921. doi: 10.3389/fmicb.2022.883921

Received: 25 February 2022; **Accepted:** 31 March 2022;
Published: 11 May 2022.

Edited by:
Elvira Fiallo-Olivé, La Mayora Experimental Station (CSIC), Spain

Reviewed by:
Pasquale Saldarelli, National Research Council (CNR), Italy
Miguel A. Aranda, Spanish National Research Council (CSIC), Spain

***Correspondence:** Anja Pecman, anja.pecman@nib.si; Denis Kutnjak, denis.kutnjak@nib.si

References

Adams, I., & Fox, A. (2016). Diagnosis of Plant Viruses Using Next-Generation Sequencing and Metagenomic Analysis. *Current Research Topics in Plant Virology*, 323–335. https://doi.org/10.1007/978-3-319-32919-2_14

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

Adams, M. J., Antoniw, J. F., & Fauquet, C. M. (2005). Molecular criteria for genus and species discrimination within the family Potyviridae. *Archives of Virology*, *150*(3), 459–479. https://doi.org/10.1007/s00705-004-0440-6

Al Rwahnih, M., Daubert, S., Golino, D., & 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*(2), 395–401. https://doi.org/10.1016/j.virol.2009.02.028

Al Rwahnih, M., Dave, A., Anderson, M. M., Rowhani, A., Uyemoto, J. K., & Sudarshana, M. R. (2013). Association of a DNA virus with grapevines affected by red blotch disease in California. *Phytopathology*, *103*(10), 1069–1076. https://doi.org/10.1094/PHYTO-10-12-0253-R

Baltimore, D. (1971). Expression of Animal Virus Genomes. *Bacteriological Reviews*, *35*(3), 235–241.

Barba, M., Czosnek, H., & Hadidi, A. (2013). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses*, *6*(1), 106–136. https://doi.org/10.3390/v6010106

Beddoe, N., Adams, I. P., McGreig, S., & Forsyth, A. (2020). Identification and full genomic sequence of nerine yellow stripe virus. *Archives of Virology*, *165*(12), 2967–2971. https://doi.org/10.1007/s00705-020-04776-3

Bester, R., Cook, G., Breytenbach, J. H. J., Steyn, C., De Bruyn, R., & Maree, H. J. (2021). Towards the validation of high-throughput sequencing (HTS) for routine plant virus diagnostics: measurement of variation linked to HTS detection of citrus viruses and viroids. *Virology Journal*, *18*(1), 1–19. https://doi.org/10.1186/s12985-021-01523-1

Boonham, N., Kreuze, J., Winter, S., van der Vlugt, R., Bergervoet, J., Tomlinson, J., & Mumford, R. (2014). Methods in virus diagnostics: From ELISA to next generation sequencing. *Virus Research*, *186*, 20–31. https://doi.org/10.1016/j.virusres.2013.12.007

Borges, F., & Martienssen, R. A. (2015). The expanding world of small RNAs in plants.

Nature Reviews Molecular Cell Biology, 1–12. https://doi.org/10.1038/nrm4085

Boykin, L. M., Sseruwagi, P., Alicai, T., Ateka, E., Mohammed, I. U., Stanton, J. L., … Ndunguru, J. (2019). Tree Lab : Portable Genomics for Early Detection of Plant Viruses and Pests in Sub-Saharan Africa.

Branton, D., & Deamer, D. (2018). Nanopore Sequencing. *Nanopore Sequencing*, *1*, 1–16. https://doi.org/10.1142/10995

Bronzato Badial, A., Sherman, D., Stone, A., Gopakumar, A., Wilson, V., Schneider, W., & King, J. (2018). Nanopore Sequencing as a Surveillance Tool for Plant Pathogens in Plant and Insect Tissues. *Plant Disease*, *102*(8), 1648–1652. https://doi.org/10.1094/PDIS-04-17-0488-RE

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*, 59. Retrieved from https://doi.org/10.1038/nmeth.3176

Chalupowicz, L., Dombrovsky, A., Gaba, V., Luria, N., Reuven, M., Beerman, A., … Manulis-Sasson, S. (2019). Diagnosis of plant diseases using the Nanopore sequencing platform. *Plant Pathology*, *68*(2), 229–238. https://doi.org/10.1111/ppa.12957

Clark, M. F., Adams, A. N., Malling, E., Clark, M. F., Adams, A. N., & Malling, E. (1977). Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J. Gen. Virol.*, *34*, 475–483. https://doi.org/10.1099/0022-1317-34-3-475

Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE*, *16*(10 October). https://doi.org/10.1371/journal.pone.0257521

Donaire, L., Wang, Y., Gonzalez-Ibeas, D., Mayer, K. F., Aranda, M. A., & Llave, C. (2009). Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology*, *392*(2), 203–214. https://doi.org/10.1016/j.virol.2009.07.005

Fellers, J. P., Webb, C., Fellers, M. C., Rupp, J. S., & Wolf, E. De. (2019). Wheat Virus Identification Within Infected Tissue Using Nanopore Sequencing Technology, *1*(September), 2199–2203. https://doi.org/10.1094/PDIS-09-18-1700-RE

Filloux, D., Fernandez, E., Loire, E., Claude, L., Galzi, S., Candresse, T., … Roumagnac, P. (2018). Nanopore-based detection and characterization of yam viruses. *Scientific Reports*, *8*(1), 17879. https://doi.org/10.1038/s41598-018-36042-7

Flores, R., Gago-Zachert, S., Serra, P., Sanjuán, R., & Elena, S. F. (2014). Viroids: Survivors from the RNA World? *Annual Review of Microbiology*, (June), 395–414. https://doi.org/10.1146/annurev-micro-091313-103416

Gaafar, Y. Z. A., & Ziebell, H. (2020). Comparative study on three viral enrichment approaches based on RNA extraction for plant virus/viroid detection using high-throughput sequencing. *PLoS ONE*, *15*(8 August 2020), 1–17. https://doi.org/10.1371/journal.pone.0237951

Hamilton, A. M. (1932). On three new virus diseases of Hyoscyamus niger. *The Annals of Applied Biology*, *19*(4), 550–567.

Hull, R. (2014). Introduction. *Plant Virology*, 3–14. https://doi.org/10.1016/B978-0-12-384871-0.00001-7

Ip, C. L. C., Loose, M., Tyson, J. R., de Cesare, M., Brown, B. L., Jain, M., … Olsen, H.

E. (2015). MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research*, *4*. https://doi.org/10.12688/f1000research.7201.1

Itaya, A., Folimonov, A., Matsuda, Y., Nelson, R. S., & Ding, B. (2001). Potato spindle tuber viroid as inducer of RNA silencing in infected tomato. *Molecular Plant-Microbe Interactions : MPMI*, *14*(11), 1332–1334. https://doi.org/10.1094/MPMI.2001.14.11.1332

Ivanov, K. I., Eskelin, K., Bašic, M., De, S., Lõhmus, A., Varjosalo, M., & Mäkinen, K. (2016). Molecular insights into the function of the viral RNA silencing suppressor HCPro. *Plant Journal*, *85*(1), 30–45. https://doi.org/10.1111/tpj.13088

Jones, R. A. C., & Naidu, R. A. (2019). Global Dimensions of Plant Virus Diseases: Current Status and Future Perspectives. *Annual Review of Virology*, *6*, 387–409. https://doi.org/10.1146/annurev-virology-092818-015606

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

Kutnjak, D., Rupar, M., Gutierrez-Aguirre, I., Curk, T., Kreuze, J. F., & Ravnikar, M. (2015). Deep sequencing of virus derived small interfering RNAs and RNA from viral particles shows highly similar mutational landscape of a plant virus population. *Journal of Virology*, *89*(9), 4760–4769. https://doi.org/10.1128/JVI.03685-14

Kutnjak, D., Silvestre, R., Cuellar, W., Perez, W., Müller, G., Ravnikar, M., & Kreuze, J. (2014). Complete genome sequences of new divergent potato virus X isolates and discrimination between strains in a mixed infection using small RNAs sequencing approach. *Virus Research*, *191*, 45–50. https://doi.org/10.1016/j.virusres.2014.07.012

Lebas, B., Adams, I., Al Rwahnih, M., Baeyen, S., Bilodeau, G. J., Blouin, A. G., … Massart, S. (2022). Facilitating the adoption of high-throughput sequencing technologies as a plant pest diagnostic test in laboratories: A step-by-step description. *EPPO Bulletin*, *52*(2), 394–418. https://doi.org/https://doi.org/10.1111/epp.12863

Liefting, L. W., Waite, D. W., & Thompson, J. R. (2021). Application of oxford nanopore technology to plant virus detection. *Viruses*, *13*(8). https://doi.org/10.3390/v13081424

Lovisolo, O., & Bartels, R. (1970). On a New Strain of Henbane mosaic virus from Physalis alkekengi By, (Cdv).

Malapi-Wight, M., Adhikari, B., Zhou, J., Hendrickson, L., Maroon-Lango, C. J., McFarland, C., … Hurtado-Gonzales, O. P. (2021). Hts-based diagnostics of sugarcane viruses: Seasonal variation and its implications for accurate detection. *Viruses*, *13*(8). https://doi.org/10.3390/v13081627

Maree, H. J., Fox, A., Rwahnih, M. Al, & Boonham, N. (2018). Application of HTS for Routine Plant Virus Diagnostics : State of the Art and Challenges, *9*(August), 1–4. https://doi.org/10.3389/fpls.2018.01082

Martínez de Alba, a E., Flores, R., Hernández, C., de Alba, A. E., Flores, R., Hernandez, C., & Martinez de Alba, A. E. (2002). Two chloroplastic viroids induce the accumulation of small RNAs associated with posttranscriptional gene silencing. *J Virol*, *76*(24), 13094–13096. https://doi.org/10.1128/JVI.76.24.13094-13096.2002

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., … Wetzel, T. (2017). A Framework for the Evaluation of Biosecurity, Commercial, Regulatory, and Scientific Impacts of Plant Viruses and Viroids Identified by NGS Technologies. *Frontiers in Microbiology*, *8*(January). https://doi.org/10.3389/fmicb.2017.00045

Massart, S., Chiumenti, M., De Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., … Candresse, T. (2018). Virus Detection by High-Throughput Sequencing of Small RNAs: Large-Scale Performance Testing of Sequence Analysis Strategies. *Phytopathology*, *109*(3), 488–497. https://doi.org/10.1094/phyto-02-18-0067-r

Massart, S., Olmos, A., Jijakli, H., & Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Research*, *188*, 90–96. https://doi.org/10.1016/j.virusres.2014.03.029

Massart Sebastien, Adams I., Al Rwahnih M., Baeyen S., Bilodeau G.J., Blouin A., Boonham N., Candresse T., Chandelier A., De Jonghe K., Fox A., Gaafar Y.Z.A., Gentit P., Haegeman A., Ho W., Hurtado-Gonzales, Jonkers W., Kreuze J., Kutnjak D., … Lebas B.S.M. (2022). Guidelines for the reliable use of high throughput sequencing technologies to detect plant pathogens and pests. Retrieved from https://doi.org/10.5281/zenodo.6637519

Mehle, N., Kutnjak, D., Jakoš, N., Seljak, G., Pecman, A., Massart, S., & Ravnikar, M. (2019). First Report of Cucurbit Aphid-Borne Yellows Virus in Cucurbita pepo and Cucurbita maxima in Slovenia. *Plant Disease*, *104*(2), 599. https://doi.org/10.1094/PDIS-07-19-1524-PDN

Mehle, N., Kutnjak, D., Tušek Žnidarič, M., & Ravnikar, M. (2018). First Report of Apium Virus Y and Carrot Thin Leaf Virus in Parsley in Slovenia. *Plant Disease*, *103*(3), 592. https://doi.org/10.1094/PDIS-04-18-0690-PDN

Meredith, L. W., Hamilton, W. L., Warne, B., Houldcroft, C. J., Hosmillo, M., Jahun, A. S., … Goodfellow, I. (2020). Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *The Lancet Infectious Diseases*, *20*(11), 1263–1272. https://doi.org/10.1016/S1473-3099(20)30562-4

Papaefthimiou, I., Hamilton, A., Denti, M., Baulcombe, D., Tsagris, M., & 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 Research*, *29*(11), 2395–2400. https://doi.org/10.1093/nar/29.11.2395

Pecman, A., Kutnjak, D., Gutiérrez-Aguirre, I., Adams, I., Fox, A., Boonham, N., & Ravnikar, M. (2017). Next generation sequencing for detection and discovery of plant viruses and viroids: Comparison of two approaches. *Frontiers in Microbiology*, *8*(OCT). https://doi.org/10.3389/fmicb.2017.01998

Pecman, A., Adams, I., Gutiérrez-Aguirre, I., Fox, A., Boonham, N., Ravnikar, M., & Kutnjak, D. (2022). Systematic Comparison of Nanopore and Illumina Sequencing for the Detection of Plant Viruses and Viroids Using Total RNA Sequencing Approach. *Frontiers in Microbiology*, *13*(May), 1–14. https://doi.org/10.3389/fmicb.2022.883921

Pecman, A., Kutnjak, D., Mehle, N., Žnidarič, M. T., Gutiérrez-Aguirre, I., Pirnat, P., … Ravnikar, M. (2018). High-Throughput Sequencing Facilitates Characterization of a "Forgotten" Plant Virus: The Case of a Henbane Mosaic Virus Infecting Tomato.

Frontiers in Microbiology, *9*(November), 1–11. https://doi.org/10.3389/fmicb.2018.02739

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., … Michel, J. (2016). Ebola surveillance. *Nature*. https://doi.org/10.1038/nature16996

Rivarez, M. P. S., Kogej, Z., Jakoš, N., Pecman, A., Seljak, G., Vučurović, A., … Kutnjak, D. (2021). First Report of Ranunculus White Mottle Ophiovirus in Slovenia in Pepper with Yellow Leaf Curling Symptom and in Tomato. *Plant Disease*, *106*(7), 2003. https://doi.org/10.1094/PDIS-08-21-1624-PDN

Rivarez, M. P. S., Pecman, A., Bačnik, K., Maksimović Carvalho Ferreira, O., Vučurović, A., Seljak, G., … Kutnjak, D. (2022). In-depth study of tomato and weed viromes reveals undiscovered plant virus diversity in an agroecosystem. *BioRxiv*, 2022.06.30.498278. https://doi.org/10.1101/2022.06.30.498278

Rivarez, M. P. S., Vučurović, A., Mehle, N., Ravnikar, M., & Kutnjak, D. (2021). Global Advances in Tomato Virome Research: Current Status and the Impact of High-Throughput Sequencing. *Frontiers in Microbiology*, *12*(May). https://doi.org/10.3389/fmicb.2021.671925

Roossinck, M. J., Martin, D. P., & Roumagnac, P. (2015). Plant virus metagenomics: Advances in virus discovery. *Phytopathology*, *105*(6), 716–727. https://doi.org/10.1094/PHYTO-12-14-0356-RVW

Rott, M., Xiang, Y., Boyes, I., Belton, M., Saeed, H., Kesanakurti, P., … Rast, H. (2017). Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroids. *Plant Disease*, *101*(8), 1489–1499. https://doi.org/10.1094/PDIS-03-17-0306-RE

Salamon, P. (2018). TERMESZTETT ÉS VADON ÉLŐ BURGONYAFÉLÉK VÍRUSOS BETEGSÉGEI ÉS VÍRUSAI MAGYARORSZÁGON. 9. A BELÉNDEK MOZAIK VÍRUS (HENBANE MOSAIC VIRUS, HMV) REZERVOÁR GAZDANÖVÉNYEI ÉS PATOLÓGIAI VÁLTOZÉKONYSÁGA. *NÖVÉNYVÉDELEM 2018, 79*(54), 161–171.

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics*, *34*(9), 666–681. https://doi.org/10.1016/j.tig.2018.05.008

Vazquez-Iglesias, I., Mcgreig, S., Pufal, H., Robinson, R., Gerard, R., Clover, G., … Adams, I. P. (2022). A novel high-throughput sequencing approach reveals the presence of a new virus infecting Rosa : rosa ilarvirus-1 (RIV-1), *300*(December 2021).

Villamor, D. E. V., Ho, T., Rwahnih, M. Al, Martin, R. R., & Tzanetakis, I. E. (2019). High Throughput Sequencing For Plant Virus Detection and Discovery, 1–10. https://doi.org/10.1094/PHYTO-07-18-0257-RVW

Visser, M., Bester, R., Burger, J. T., & Maree, H. J. (2016). Next-generation sequencing for virus detection: covering all the bases. *Virology Journal*, *13*(1), 85. https://doi.org/10.1186/s12985-016-0539-x

Vunsh, R., Rosner, A., & Stein, A. (1990). The use of the polymerase chain reaction (PCR) for the detection of bean yellow mosaic virus in gladiolus. *Annals of Applied Biology*, *117*(3), 561–569.

Wu, Q., Ding, S.-W., Zhang, Y., & 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. https://doi.org/10.1146/annurev-phyto-080614-120030

Yelina, N. E., Savenkov, E. I., Solovyev, A. G., Morozov, S. Y., & Valkonen, J. P. T. (2002). Long-distance movement, virulence, and RNA silencing suppression controlled by a single protein in hordei- and potyviruses: complementary functions between virus families. *Journal of Virology*, *76*(24), 12981–12991. https://doi.org/10.1128/JVI.76.24.12981-12991.2002

Bibliography

Publications Related to the Thesis

Pecman, A., Kutnjak, D., Gutiérrez-Aguirre, I., Adams, I., Fox, A., Boonham, N., & Ravnikar, M. (2017). Next generation sequencing for detection and discovery of plant viruses and viroids: Comparison of two approaches. *Frontiers in Microbiology*, *8*(OCT). https://doi.org/10.3389/fmicb.2017.01998

Pecman, A., Kutnjak, D., Mehle, N., Žnidarič, M. T., Gutiérrez-Aguirre, I., Pirnat, P., Adams, I., Boonham, N., Ravnikar, M. (2018). High-Throughput Sequencing Facilitates Characterization of a "Forgotten" Plant Virus: The Case of a Henbane Mosaic Virus Infecting Tomato. *Frontiers in Microbiology*, *9*(November), 1–11. https://doi.org/10.3389/fmicb.2018.02739

Pecman, A., Adams, I., Gutiérrez-Aguirre, I., Fox, A., Boonham, N., Ravnikar, M., & Kutnjak, D. (2022). Systematic Comparison of Nanopore and Illumina Sequencing for the Detection of Plant Viruses and Viroids Using Total RNA Sequencing Approach. *Frontiers in Microbiology*, *13*(May), 1–14. https://doi.org/10.3389/fmicb.2022.883921

Biography

Anja Pecman is a PhD student, currently employed at National Institute of Biology, Ljubljana, Slovenia. She started her BSc study (program Food Science and Nutrition) in 2007 at Biotechnical faculty, Ljubljana, Slovenia and finished it in 2011. In the same year, she continued with MSc study in the same faculty, program Biotechnology. In 2013 she visited the joined laboratory of European Centre for Environment and Human Health and University of Exeter Medical School in Penryn Campus, UK where she was performing the analysis on Bacillus subtilis quorum sensing mechanism under the supervision of Dr Michiel Vos. In 2015 she finished her MSc Degree (Microcystin isolation, electrochemical degradation and toxicity test of byproducts) at the Biotechnical faculty in collaboration with National Institute of Biology, Department of Genetic Toxicology and Cancer Biology, Ljubljana, Slovenia. The results were also scientifically published: Meglič et al., 2016, doi: 10.1016/j.jes.2016.02.016. Later in 2015, she was selected as a PhD candidate under the supervision of Prof Dr Maja Ravnikar and Dr Ion Gutierrez Aguirre at the Department of Biotechnology and Systems Biology in National Institute of Biology, Ljubljana, Slovenia. She started her postgraduate studies in Nanosciences and Nanotechnologies at the Jožef Stefan International Postgraduate School in Ljubljana, Slovenia where she was also the member of their student council (2016-2018). Her research work is focused on detection and identification of different plant viruses with high throughput sequencing technologies (HTS). In 2016, 2017, 2018 and 2019 she visited FERA ltd. Institute (UK) where she has done part of the research work with HTS technology under the supervision of Prof Dr Neil Boonham, Dr Ian Adams and Dr Adrian Fox. From the collaboration of the first two visits two paper were published, e.g.: Pecman et al., 2017, doi: 10.3389/fmicb.2017.01998 and Pecman et al., 2018, doi:10.3389/fmicb.2018.02739. The visit in 2018 and 2019 were in the frame of collaborative research project, resulting in the third research paper: Pecman et al., 2022, doi: 10.3389/fmicb.2022.883921. From 2021 she is involved in the work of Laboratory for plant health diagnostics where she is using HTS for the detection of plant viruses and viroids. She attended workshops/webinars for the data analysis and for the use of HTS in diagnostics, furthermore she presented her work on different local and international conferences. She was working mentor of seven students, two MSc students, two BSc students and three high school students.