Article

Identification of Viruses and Viroids Infecting Tomato and Pepper Plants in Vietnam by Metatranscriptomics

Hoseong Choi [1], Yeonhwa Jo [1], Won Kyong Cho [1], Jisuk Yu [2], Phu-Tri Tran [3,†], Lakha Salaipeth [4], Hae-Ryun Kwak [5], Hong-Soo Choi [5] and Kook-Hyung Kim [1,2,3,*]

[1] Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea; bioplanths@snu.ac.kr (H.C.); yeonhwajo@gmail.com (Y.J.); wonkyong@gmail.com (W.K.C.)

[2] Plant Genomics and Breeding Institute, Seoul National University, Seoul 08826, Korea; mago03@snu.ac.kr

[3] Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Korea; tranphutri@gmail.com

[4] School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand; lakha.sal@kmutt.ac.th

[5] Crop Protection Division, National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Korea; hrkwakhahn@korea.kr (H.-R.K.); hschoi@korea.kr (H.-S.C.)

[*] Correspondence: kookkim@snu.ac.kr; Tel.: +82-2-880-4677

[†] Present address: Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, 11794 NY, USA.

check for
updates

Abstract: Tomato (*Lycopersicum esculentum* L.) and pepper (*Capsicum annuum* L.) plants belonging to the family *Solanaceae* are cultivated worldwide. The rapid development of next-generation sequencing (NGS) technology facilitates the identification of viruses and viroids infecting plants. In this study, we carried out metatranscriptomics using RNA sequencing followed by bioinformatics analyses to identify viruses and viroids infecting tomato and pepper plants in Vietnam. We prepared a total of 16 libraries, including eight tomato and eight pepper libraries derived from different geographical regions in Vietnam. We identified a total of 602 virus-associated contigs, which were assigned to 18 different virus species belonging to nine different viral genera. We identified 13 different viruses and two viroids infecting tomato plants and 12 viruses and two viroids infecting pepper plants with viruses as dominantly observed pathogens. Our results showed that multiple infection of different viral pathogens was common in both plants. Moreover, geographical region and host plant were two major factors to determine viral populations. Taken together, our results provide the comprehensive overview of viral pathogens infecting two important plants in the family *Solanaceae* grown in Vietnam.

Keywords: metatranscriptomics; pepper; tomato; Vietnam; virus

1. Introduction

Tomato (*Lycopersicum esculentum* L.) and pepper (*Capsicum annuum* L.) plants belonging to the family Solanaceae are cultivated worldwide. Tomato fruit, classified botanically as a berry, is consumed in various ways, such as fresh in salad or as materials for diverse dishes, sauces, and drinks [1,2]. Pepper fruits of Capsicum plants have diverse names according to their regions and types. For instance, piquant pepper varieties are referred as chili peppers, whereas peppers with large- or mid-sized fruits are referred to as bell peppers. Sometimes, colors are important factors for determining names of peppers, such as green pepper and red pepper [3].

To date, diverse viruses infecting tomato and peppers have been reported around the world. Virus infections on tomatoes and peppers have a negative impact on these crops, such as low quality and quantity of fruit production [4–8]. Among the viruses infecting solanaceous vegetables,

cucumber mosaic virus (CMV) in the genus *Cucumovirus* might be the most important virus for plants in the family *Solanceae* [9]. Tomato spotted wilt virus (TSWV) in the genus *Tospovirus* is a recently emerging plant virus that causes serious viral diseases in a wide range of important plants, including pepper and tomato [10]. In addition, tomato yellow leaf curl virus (TYLCV) in the genus *Begomovirus* causes severe damage on tomato cultures [11]. Moreover, several viroids, such as citrus exocortis viroid (CEVd), columnea latent viroid (CLVd), pepper chat fruit viroid (PCFVd), potato spindle tuber viroid (PSTVd), tomato chlorotic dwarf viroid (TCDVd) and tomato apical stunt viroid (TASVd) in the family *Pospiviroid*, infect pepper and tomato plants [9].

　　The rapid development of next-generation sequencing (NGS) technology facilitates the identification of viruses and viroids infecting plants [12–14]. NGS-based libraries have been prepared using various methods including small RNAs, double-stranded RNAs, messenger RNA, and ribosomal RNA-depleted total RNA. Hordeum vulgare endornavirus (HvEV), composed of dsRNA as a genome, was identified from barley using MiSeq [15]. In addition, barley yellow striate mosaic virus (BYSMV), a genus of *Cytorhabdovirus*, was identified and assembled into a complete genome using small RNA sequencing analysis [16]. NGS-based approaches reveal not only known viral pathogens, but also novel viral pathogens. In addition, the assembly of the viral genome, viral abundance, and virus mutation can be achieved using NGS-based approaches followed by intensive bioinformatics analyses.

　　Several viruses infecting tomato and pepper plants in Vietnam have been reported previously [4,17–22]. However, most of the reported viruses infecting tomato and pepper in Vietnam have been limited to begomoviruses. Hence, the knowledge of RNA viruses and viroids infecting tomato and pepper plants is very limited. In this study, we carried out metatranscriptomics using RNA-sequencing (RNA-Seq) followed by bioinformatics analyses to identify viruses and viroids infecting tomato and pepper plants in Vietnam. We addressed the diversity of viruses and viroids infecting the two plants in different geographical regions in Vietnam.

2. Results

2.1. Collection of Leaf Samples and Generation of Libraries for Identification of Viruses Infecting Tomato and Pepper Plants

　　To identify viruses infecting tomato and pepper plants, we collected leaf samples showing viral disease symptoms, including yellowing, mosaic, mottling, and dwarfing, in different geographical regions in Vietnam (Figure 1). We collected samples from eight and five different regions for tomatoes and peppers, respectively. In the case of pepper samples, three different kinds of pepper such as chili pepper, bell pepper, and red pepper were collected. For the pooling of the samples, we collected more than three leaves from each farm located in each individual region.

　　Samples collected from the same host plants in the same geographical regions were pooled and subjected to total RNA extraction followed by preparation of libraries for RNA-Seq. We named each library according to geographical regions and host plants. For example, the tomato sample collected from Dong Ahn was named DAT (Dong Ahn Tomato) and the chili pepper sample collected from Gia Lam was named GLCP (Gia Lam Chili Pepper). In total, we prepared 16 libraries, including eight tomato and eight pepper libraries. The prepared libraries were paired-end (2 × 100 bp) sequenced by HiSeq 2000 system.
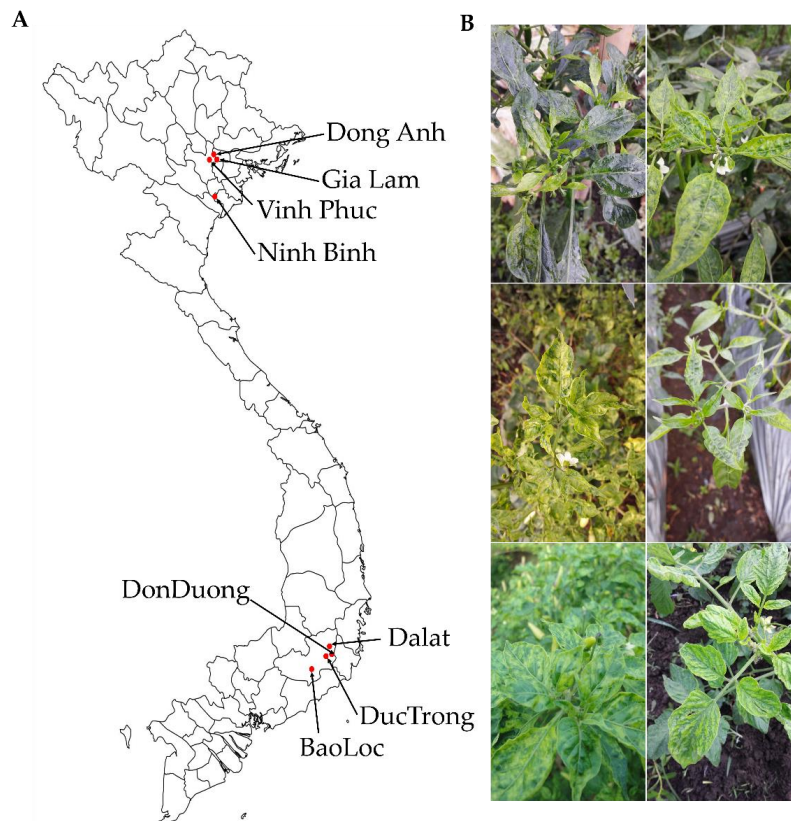
Figure 1. Locations of pepper and tomato sample collection (**A**) and typical symptoms observed in collected samples (**B**).

2.2. Identification of Virus-Associated Contigs

The numbers of raw read bases ranged from 8,473,848,288 (DDBP library) to 4,899,434,654 (NBT library). The sequenced raw data of each library were assembled using the Trinity program for de novo assembly. The number of assembled contigs ranged from 77,563 (library NBT) to 146,324 (DDRP library) (Figure 2A).
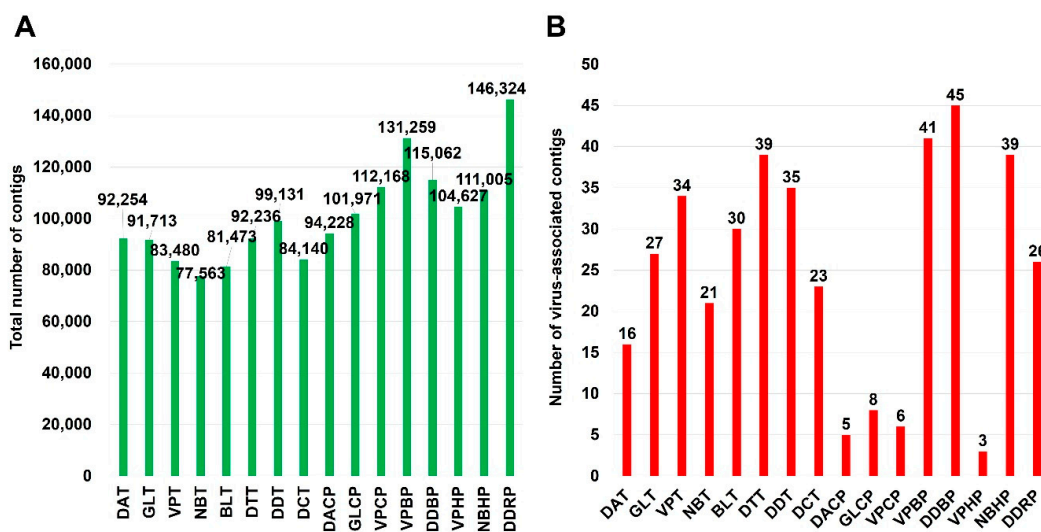

Figure 2. Number of assembled contigs and virus-associated contigs. (**A**) The total number of contigs assembled by Trinity program in each library. (**B**) The number of virus-associated contigs in each library, which was identified by BLASTN search against viral genome database.

To identify virus-associated contigs, we performed a BLASTN search using assembled contigs against virus reference genome database obtained from NCBI database. We identified a total of 602 virus-associated contigs. The number of virus-associated contigs for tomato (450 contigs) was higher than that for pepper (152 contigs) (Figure 2B). The number of identified virus-associated contigs was ranged from three (VPHP) to 45 (DDBP).

2.3. Classification of Identified Virus-Associated Contigs According to Virus Taxonomy

We identified 18 different virus species belonging to nine different viral genera (Supplementary Table S1). We identified Potato virus Y (PVY) and Chilli veinal mottle virus (ChVMV) belonging to the genus *Potyvirus*. Three viruses in the genus *Tobamovirus*—tomato mosaic virus (ToMV), pepper mild mottle virus (PepMMV), and tobacco mild green mosaic virus (TMGMV)—were identified in both plants. In addition, CMV and pepper vein yellow virus (PepVYV) from the genus *Polerovirus* were identified. Interestingly, we identified three viruses with tripartite RNA fragments such as capsicum chlorosis virus (CaCV), pepper chlorsotic spot virus (PCSV), and tomato necrotic ring virus (TNRV) from the genus *Tospovirus*. We identified four different single-stranded DNA viruses such as ageratum yellow vein China virus (AYVCNV), tomato yellow leaf curl Kanchanaburi virus (TYLCKaV), lindernia anagallis yellow vein virus (LaYVV), and tomato yellow leaf curl Vietnam virus (TYLCVNV) from the genus *Begomovirus*. Moreover, two double-stranded (ds) RNA viruses—pepper cryptic virus 2 (PCV2) in the genus *Deltapartitivirus* and southern tomato virus (STV), from the genus *Amalgavirus*—were identified. Furthermore, two viroids, CLVd and PCFVd from the genus *Pospiviroid* were identified.

2.4. Proportion of Identified Viruses and Viroids According to Virus-Associated Contigs

We identified 13 different viruses and two viroids from eight different tomato libraries. Of these, PCFVd (225 contigs) was the most frequently identified viral pathogen based on the number of identified contigs, followed by TNRV (56 contigs), TYLCKaV (26 contigs), PVY (23 contigs), and CMV (23 contigs) (Figure 3A).

In the case of pepper, we identified 12 viruses and two viroids. PeVYV (23 contigs) was the most frequently identified virus, followed by CMV (22 contigs), PCSV (22 contigs), PCV-2 (16 contigs), and TMGMV (16 contigs) (Figure 3B). We compared the identified viruses and viroids between tomato and pepper (Figure 3C). Nine viruses and two viroids (CLVd and PCFVd) were commonly identified in both plants. AYVCNV, LaYVV, STV, and TYLCVNV were identified specifically in tomato while PeVYV, PMMoV, and TMGMV were identified only in pepper. According to viral genome types, the most identified virus-associated contigs were derived from single-stranded (ss) RNA genome (73%), followed by ssDNA genome (16%), dsRNA genome (9%), and circular ssRNA genome (2%) (Figure 3D).
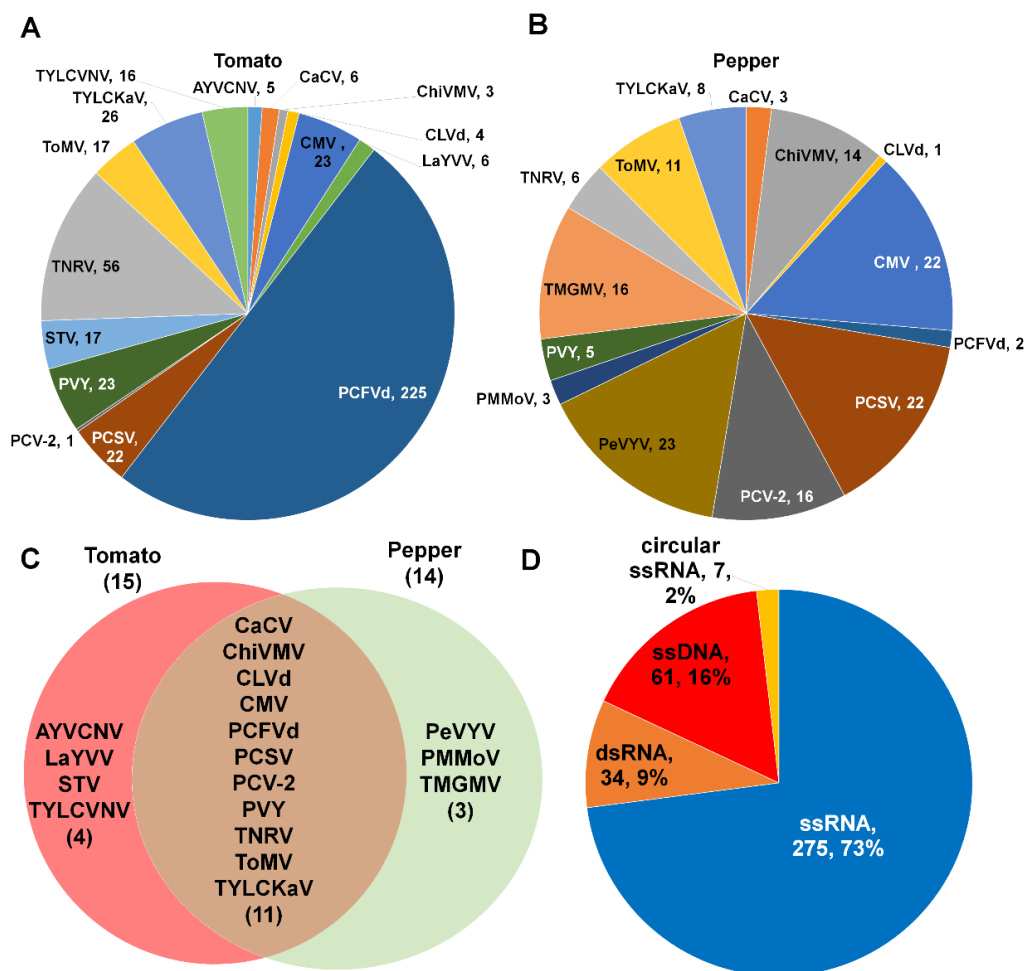
Figure 3. Number of virus-associated contigs for individual identified viral pathogen. The pie charts display the number of virus-associated contigs for identified viruses and viroids infecting tomato (**A**) and pepper (**B**) plants. (**C**) Venn diagram displays a comparison of identified viral pathogens between tomato and pepper plants. Full virus names can be found in Supplementary Table S1. (**D**) Proportion of identified viral pathogens according to viral genome type.

2.5. Proportion of Identified Viruses and Viroids According to Virus-Associated Reads

Next, we calculated the proportion of virus-associated reads in each library (Figure 4A). Except VPCP (1.048%), the proportion of virus-associated reads was less than 1% in all libraries. For eight tomato libraries, the proportion of virus-associated reads ranged from 0.012% (BLT) to 0.575% (DAT), whereas the proportion of virus-associated reads for eight pepper libraries ranged from 0.001% (DACP) to 1.048% (VPCP).

We examined the number of libraries for each virus and viroid (Figure 4B). PCV-2 was identified in eight libraries (one tomato and seven pepper libraries). AYVCNV was identified only in a single tomato library, while PeVYV and ToLCVV were identified only in a single pepper and tomato library, respectively. In contrast to other viruses, STV was identified in six tomato libraries.

We examined the number of identified viral pathogens in each library (Figure 4C). Except two libraries, DAT and VPHP, all libraries contained at least two different viral pathogens. Of these, two libraries, BLT and DDBP, contained at least seven different viral pathogens. Three tomato libraries, DTT, DDT, and DCT, and a single pepper library, DDRP, were infected by five different viral pathogens.

Based on virus-associated reads, we examined the proportion of viral pathogens in each library (Figure 4D). TNRV was the dominant virus in DAT and GLCP. CLVd was the dominant viral pathogen in GLT and NBT. PVY was the dominant virus in BLT, DTT, and VPCP. In DDT, DCT, DDBP, and DDRP,

PCSV was the dominant virus. PCV-2 was dominantly present in DACP and VPHP. ChiVMV was dominant in VPBP and NBHP.
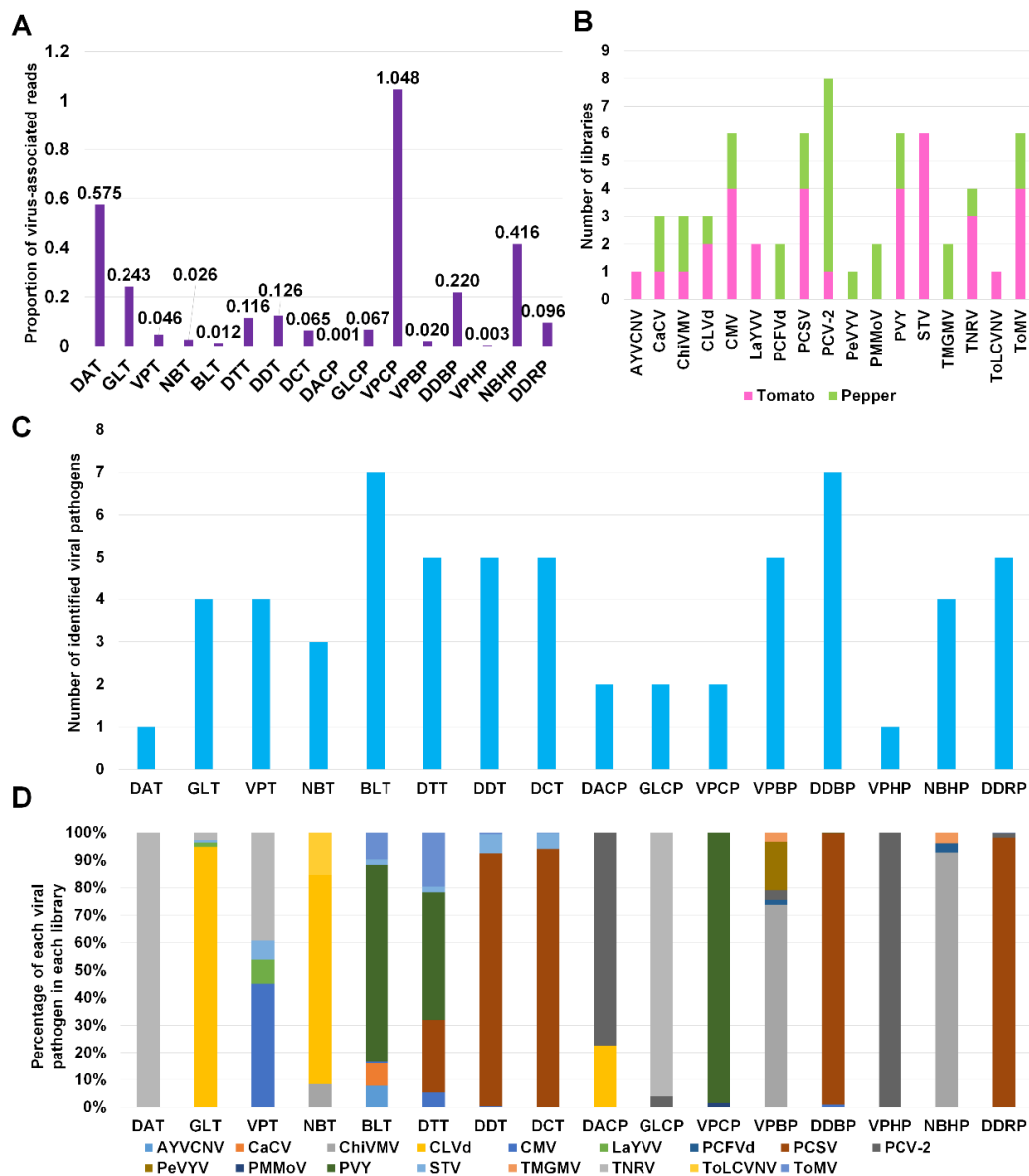

Figure 4. The viral population of identified viral pathogens in each library. (**A**) The proportion of virus-associated reads in each library. (**B**) The frequency of identified viral pathogens in different tomato and pepper libraries. Purple- and green-colored bars indicate the number of libraries for tomato and pepper plants, respectively, in which viral pathogens were infected. (**C**) The number of identified virus and viroid species in each library. (**D**) The viral population of identified viral pathogens in each library based on the number of virus-associated reads.

2.6. Proportion of Identified Viruses and Viroids According to Plant Host and Geographical Region

According to plant hosts, we examined the proportion of identified viral pathogens by combining all virus-associated reads. In tomato plants, TNRV (43.4%) was the most dominant viral pathogen, followed by PCSV (21.5%) and CLVd (20.8%) (Figure 5A). In pepper plants, PVY (47.9%) was the dominant virus, followed by PCSV (23.4%) and ChiVMV (22.5%) (Figure 5B). We next examined virus proportion according to Northern and Southern Vietnam, regardless of plant hosts. In Northern Vietnam, PCSV (82.7%) was the dominant viral pathogen, followed by PVY (9.3%) and ToMV (3.9%)

(Figure 5C). In Southern Vietnam, PVY (39%) was the dominant viral pathogen, followed by TNRV (26.7%), ChiVMV (18.4%), and CLVd (11.7%) (Figure 5D).
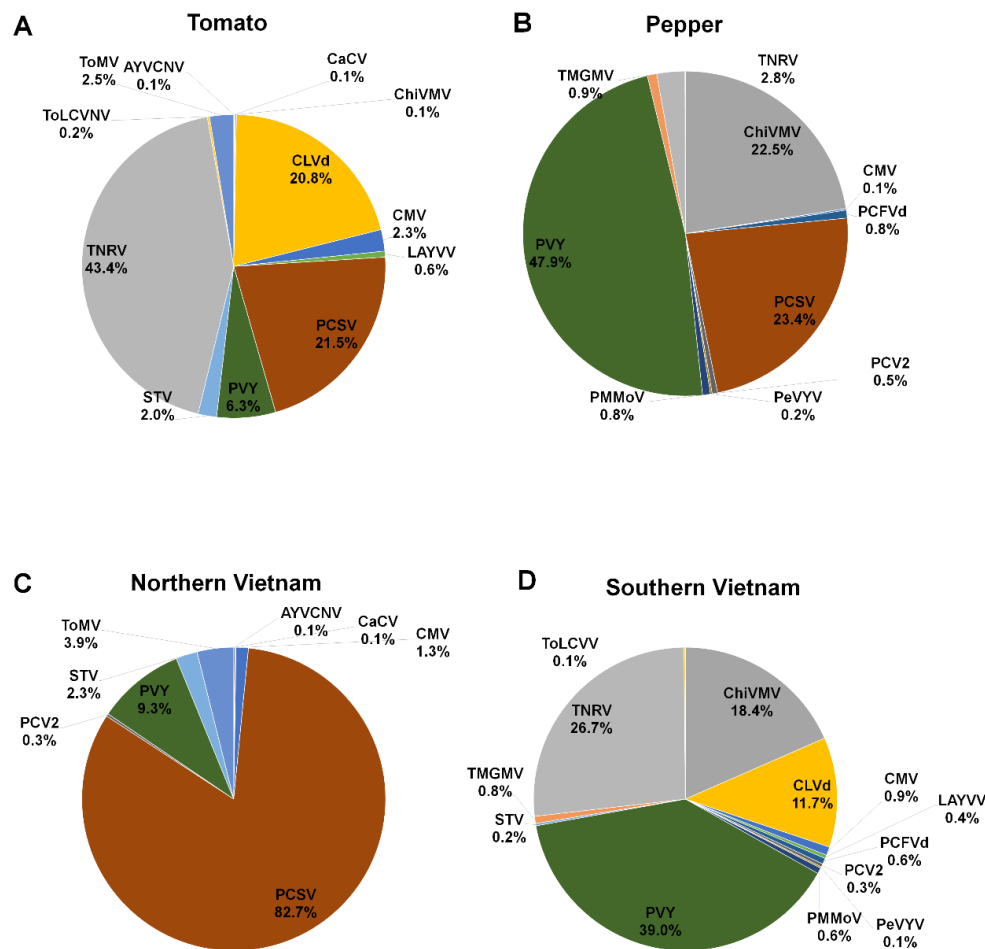

Figure 5. The proportion of identified viral pathogens according to host plants and geographical regions. Pie charts display proportion of identified viral pathogens in tomato (**A**) and pepper (**B**) plants according to virus-associated reads. The proportion of identified viral pathogens in northern (**C**) and southern Vietnam (**D**) based on virus-associated reads.

2.7. Phylogenetic Analyses for Identified Viruses and a Viroid

We assembled viral genomes for four STV isolates, one PMMoV isolate, three ToMV isolates, one TMGMV isolate, one AYVCV isolate, and two PCFVd isolates by RNA-Seq and conducted bioinformatics analyses. The assembled viral genomes, which were assembled to complete genomes, were subjected to BLASTN search to retrieve homologous viral genome sequences. After nucleotide sequence alignment, we generated six different phylogenetic trees (Figure 6). The phylogenetic tree for STV isolates revealed that isolates including GLT (MW012410) and DDT (MW012413) were closely related, while STV isolates DCT (MW012412) and DTT (MW012411) were grouped in the other clade. Notably, STV isolate DTT was classified into different groups, in contrast to other isolates (Figure 6A). The identified PMMoV isolate VPCP (MW012414) was closely related to other isolates from China (Figure 6B). In the case of ToMV, we have already reported three partial genome sequences of ToMV from tomato, pepper leaves, and chili seeds [17]. The ToMV sequence from the previous study was used for phylogenetic tree construction. The phylogenetic tree of ToMV showed two distinct groups. Three ToMV isolates (isolate DTT; MW012409, isolate DDBP; MH393623, isolate BLT; MH393621) from Vietnam were grouped with an isolate from Japan (Figure 6C). The TMGMV isolate NBHP (MW012408) was closely related to the isolate CaJO from Jordan (Figure 6D). In the case of AYVCNV, we previously

reported the assembled genome sequence as Ageratum yellow vein virus (AYVV) isolate BaoLoc [10]. However, a BLASTN search against a nucleotide database using the same identical genome sequence showed that the nearly complete genome for AYVV isolate BaoLoc showed strong genetic relationship with other known isolates from China (Figure 6E). Therefore, we renamed the virus as AYVCNV isolate BaoLoc (MW012407). Two PCFVd isolates (MW012406 and MW012415) from Vietnam were closely related with an isolate identified in tomato plants in Thailand (Figure 6F).
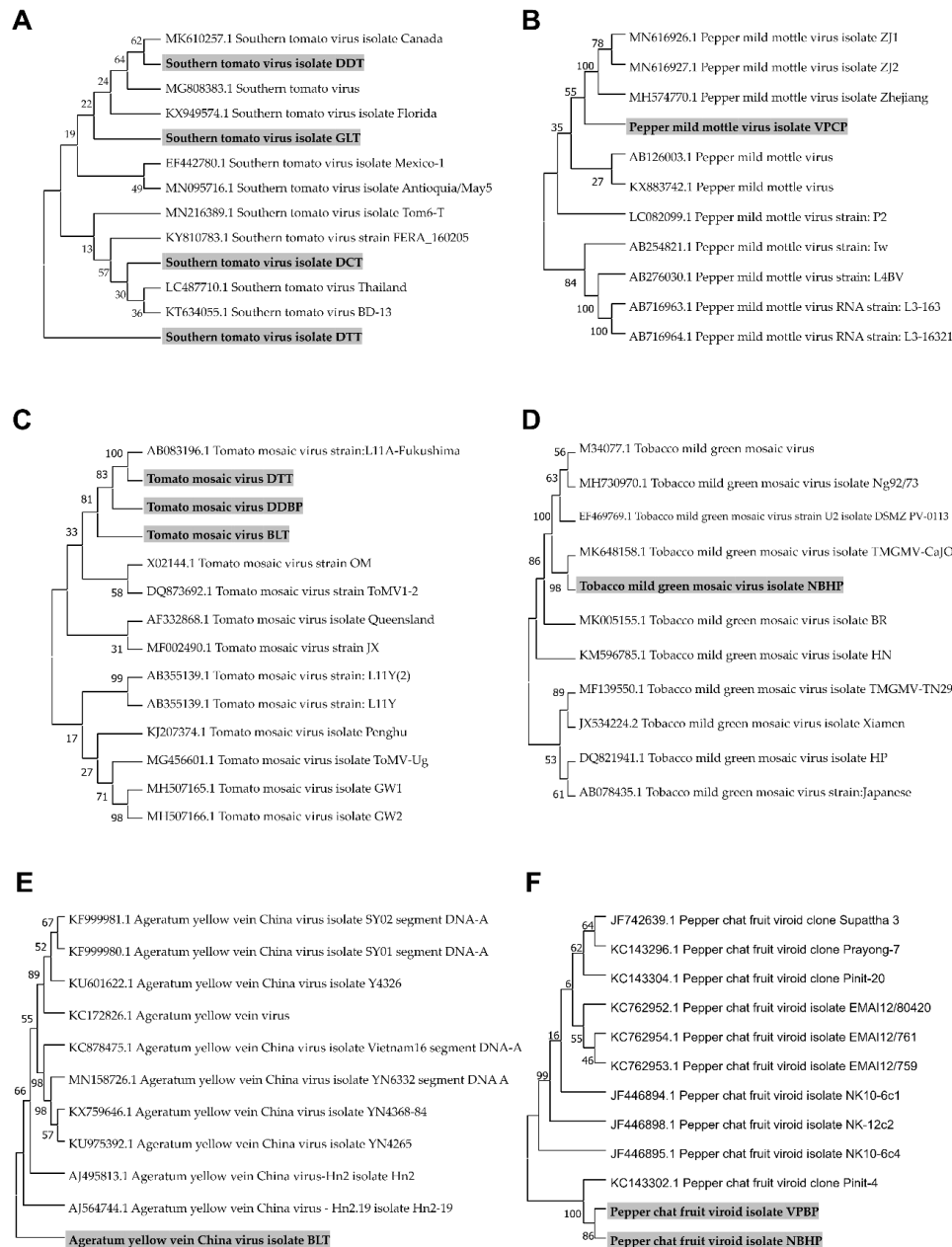

Figure 6. Phylogenetic trees for the identified five viruses and a viroid. For the construction of phylogenetic trees, we selected five viruses, including STV (**A**), PMMoV (**B**), ToMV (**C**), TMGMV (**D**), and AYVCNV (**E**), and a viroid PCFVd (**F**), in which genome sequences were assembled by RNA-Seq. Assembled viral genome sequences, as well as matched known viral genome sequences from GenBank, were subjected to the construction of phylogenetic trees. Phylogenetic trees were constructed using the MEGA7 program using the Maximum Likelihood method with a bootstrap of 1000 replicates. The identified viral isolates from this study were marked by the gray color.

2.8. Validation of Results for RNA-Seq by RT-PCR

We carried out RT-PCR using virus-specific primers to confirm results of RNA-Seq. Virus-specific primers were designed based on the identified sequence for individual virus and viroid (Table 1). The results of RT-PCR were similar to those of RNA-Seq (Figure 7). For example, infection of two viruses—TNRV and PCV in the DDRP library—was confirmed by RT-PCR. In addition, we validated infection of six viruses in the DTT library using RT-PCR.

Table 1. Information of Primer Pairs Used for RT-PCR to Validate the RNA-Seq Results.

Primers *	Nucleotide Sequence	Expected Size
PVY_diag_F	ATGGCAAATGACACAATCGATGCAG	804 bp
PVY_diag_R	TCACATGTTCTTAACTCCAAGTAGAGTATG	
CMV_diag_F	ATGGACAAATCTGAATCAACCAGTG	754 bp
CMV_diag_R	GACTGGGAGCACTCCAGATG	
ToMV_diag_F	CGAGAGGGGCAACAAACAT	317 bp
ToMV_diag_R	ACCTGTCTCCATCTCTTTGG	
AYVCNV_diag_F	CTGATGTGCCCAAAGGTTGT	406 bp
AYVCNV_diag_R	GCCTGCTCCTTAGACGCATA	
TYLCKaV_diag_F	TGCCGAAGCGTTCAATAGAT	733 bp
TYLCKaV_diag_R	TGTTGCATACACAGGATTAGAGG	
CaCV_diag_F	AAGACCTCGAAAGAGGCAAA	703 bp
CaCV_diag_R	CTTCGGAGGCAAACTATTGG	
TNRV_diag_F	TTGCTAGCTGGAGGAGAAGC	786 bp
TNRV_diag_R	TCCTCTCCTAGTTGGCTTGC	
PCV2_diag_F	TTCAATCGACGGTTTCACAA	792 bp
PCV2_diag_R	CCTTGACTTGAGGTCGTGGT	
STV_diag_F	CAAAGGGAAGACTGCTGAGG	808 bp
STV_diag_R	AGCCTCTCCATCGGGATTAT	
ChiVMV_diag_F	GCGTAAAAGGCGAAGACTCA	782 bp
ChiVMV_diag_R	GTGCCGTTCAGTGTCCTCTT	
PMMoV_diag_F	ATGGCTTACACAGTTTCCAGTGCCAA	474 bp
PMMoV_diag_R	TTAAGGAGTTGTAGCCCAGGTGAGTCC	
TMGMV_diag_F	ATGCCTTATACAATCAACTCTCCG	480 bp
TMGMV_diag_R	CTAAGTAGCCGGAGTTGTGGTC	
PeVYV_diag_F	ATGAATACGGGAGGGGTTAGG	621 bp
PeVYV_diag_R	CTATTTCGGGTTGTGCAATTGC	
LaYVV_diag_F	ATGTCGAAGCGACCTGCAGATAT	741 bp
LaYVV_diag_R	GATTTTCAGAGTAGCATACACGGGA	
CLVd_diag_F	CGGAACTAAACTCGTGGTTCCTG	370 bp
CLVd_diag_R	AGGAACCTACTGCGGTTCCA	
PCFVd_diag_F	CCGGATTCTTCTAAGGGTGCCT	317 bp
PCFVd_diag_R	AGATCCTCTCGGGTCCCGG	
ToLCVV_diag_F	GCGTTAATGCGTCCCATAAT	528 bp
ToLCVV_diag_R	GCATTAAAGTCGTGGGCAAT	
TYLCVNV_diag_F	AGAAACGCCAAGTCTGAGGA	314 bp
TYLCVNV_diag_R	GTTCGGAGACGGAGAGTTGA	

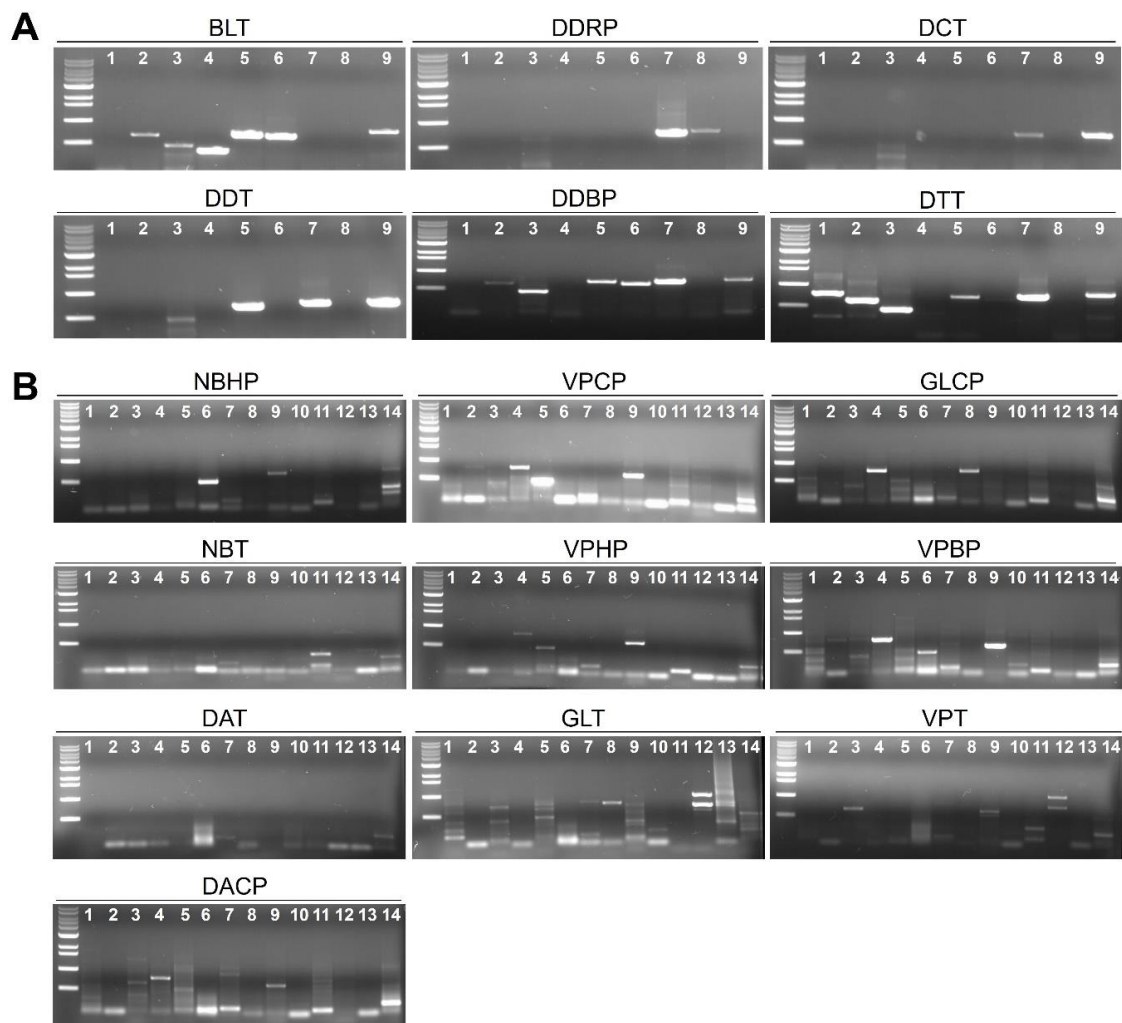* F and R indicate forward and reverse primer, respectively.

Figure 7. Confirmation of identified viruses and viroids by RT-PCR. RT-PCR results of viruses and viroids identified from the Southern (**A**) and Northern area (**B**) of Vietnam. Each sample on the gel was indicated by different number as follows (for panel A; 1-PVY, 2-CMV, 3-ToMV, 4-AYVCNV, 5-TYLCKaV, 6-CaCV, 7-TNRV, 8-PCV2, 9-STV, for panel B; 1-ChiVMV, 2-PVY, 3-CMV, 4-PCV2, 5-PMMoV, 6-TMGMV, 7-STV, 8-TNRV, 9-PeVYV, 10-ToLCVV, 11-TYLCVNV, 12-LAYVV, 13-CLVd, 14-PCFVd).

3. Discussion

Recently, a large number of viruses and viroids infecting tomato and pepper plants have been identified in a single study based on diverse NGS techniques [23–25]. Although tomato and pepper plants are widely cultivated in Vietnam and viruses infecting tomato and pepper plants can cause devastating epidemics, little is known about viruses and viroids infecting both plants. In this study, we identified 15 and 14 viral pathogens infecting tomato and pepper plants, respectively, grown in the diverse fields in Vietnam by RNA-Seq.

A previous study identified a total of 22 viruses infecting tomato derived from 170 field-grown samples in China by small RNA sequencing [25]. When we compared our results to the previous study, six viruses, including PVY, ChiVMV, CMV, ToMV, STV, and TYLCV, were commonly identified in both studies. In contrast, six viral pathogens, including AYVCNV, CaCV, PCSV, TNRV, CLVd, and LAYVV, were identified only in our study. Of these, with the exception of AYVCNV, this is the first report of five viral pathogens, including CaCV, PCSV, TNRV, CLVd, and LAYVV, infecting tomato plants in Vietnam. In addition, this is the first report of PCSV, TNRV, and LAYVV infecting tomato in the world.

Pepper viromes in two different pepper cultivars grown in India have previously been reported, revealing diverse DNA and RNA viruses infecting pepper plants [23]. Interestingly, none of the viruses were identified in both studies, suggesting geographical regions and plant varieties might be important factors for virus diversity. Both tomato and pepper plants are usually cultivated from seeds. In particular, seed transmission of several DNA viruses, such as TYLCV and pepper yellow leaf curl Indonesia virus, in pepper and chili pepper, respectively, have previously been reported [26,27]. Moreover, the seed transmission of viruses infecting tomato has been reported by several research groups [6,28,29]. In addition, seed transmission of two viroids, tomato planta macho viroid and PCFVd, in plants belonging to the family *Solanaceae* has been reported [8]. By contrast, a recent study demonstrated that TYLCV, known as a seed-transmitted begomovirus, was not seed transmitted in tomato and tobacco plants [30]. This study examined infection of TYLCV in surface-disinfected or untreated seeds, resulting in no infection of TYLCV, suggesting that most of the virus was located externally as a contaminant of the seed coat [30]. Regardless of seed-borne or seed transmission of several viruses and viroids infecting tomato and pepper plants, the seed could be a main factor for virus transmission in tomato and pepper plants in Vietnam.

The advance of NGS techniques facilitates the easy identification of known and unreported viruses in a target plant [21]. Based on RNA-seq, we were able to identify several unreported viruses and viroids in Vietnam. For example, this is the first study reporting eight viruses, including PVY, ChiVMV, CMV, CaCV, PCSV, PCV-2, TNRV, and TMGMV, as well as two viroids, CLVd and PCFVd, infecting pepper in Vietnam. Furthermore, we were able to determine the proportion of individual viral pathogen in a given sample and distribution of identified viruses and viroids in different regions and plants, as shown in the previous studies [31,32].

It is noteworthy that tomato and pepper plants are hosts for a wide range of viruses and viroids, as shown previously [23,25]. In our study, we identified a total of 18 viral species in eight genera. Moreover, the identified viruses and viroids have different kinds of genome types, such as ssDNA, dsRNA, ssRNA, and circular ssRNA. Of these, viruses with ssDNA genomes were frequently identified, suggesting that they could be major factors causing viral diseases in both plants, as suggested previously [33]. In the case of begomoviruses with ssDNA genomes, a different kind of begomovirus was identified in each region. Two viruses, PCV-2 and STV with dsRNA genome, were frequently identified in pepper and tomato plants, respectively. In fact, detailed disease symptoms caused by viruses with dsRNA genome have not been well characterized [34,35].

Furthermore, many viruses and viroids were commonly identified in both tomato and pepper plants, which are members in the family *Solanaceae* [36–38]. Based on our results, geographical region and host were important factors in determining viral population. For example, three libraries, DDT, DDBP, and DDRP, originated from different hosts but from the same region, Don Duong. The composition of viral pathogens in the three libraries was very similar, and PCSV was dominantly present in all three libraries. In the case of NBT and NBHP libraries, from tomato and pepper plants, respectively, grown in Ninh Binh, CLVd was dominant in NBT, while ChiVMV was dominant in NBHP, suggesting host-specific viral populations. We carefully supposed that the dominance of CLVd in NBT library might be associated with seed transmission of CLVd, as reported previously [9]. Some viruses, such as PVY from BLT, were associated with NGS reads, but not detected with PCR. RNA-Seq results demonstrate the number of virus-associated contig and coverage of sequence against the complete genome of each virus. To validate the RNA-Seq analysis using the conventional RT-PCR technique, a high level of sequence coverage is necessary. Although PVY in BLT library was identified by RNA-Seq with high values of viral-associated contigs, the coverage of PVY in BLT was low compared to other viruses. This disagreement between RNA-seq and RT-PCR suggests that we need to carefully check the level of coverage for successful validation by conventional RT-PCR, even though the number of virus-associated contigs was high.

Multiple infection by diverse viruses in a single host is very common. Similarly, at least seven different viral pathogens were found to infect both tomato (BLT) and pepper plants (DDBP).

Furthermore, the analysis of viral population using virus-associated reads revealed that there was, preferentially, a dominant viral pathogen in tomato and pepper plants grown in the same field. However, we could not confirm whether the dominant viral pathogen played a major role in causing typical virus symptoms. Taken together, we identified diverse viruses and viroids infecting tomato and pepper plants grown in the fields in Vietnam by RNA-Seq. Our results showed that infection by different viral pathogens was common in the two plants. However, a specific viral pathogen was dominantly present, depending on the host plants as well as on the isolated regions, suggesting that the geographical region and host plant were two major factors for determining viral populations. Since we could not confirm whether the dominant virus identified by RNA-Seq is the major pathogen in each plant to cause viral symptoms observed, further experiments and screenings for identified viruses are required to provide more information for preventing virus diseases epidemic in pepper and tomato fields. Although we did not provide direct evidence for economic losses caused by virus infections in pepper and tomato plants in Vietnam, our results provide the comprehensive overview of viral pathogens infecting two important plants in the family *Solanaceae*. Many plant viruses reported in this study could infect diverse host plants and thus present the possibility of the continuous virus disease epidemics in the fields of Vietnam and potential threat to the agricultural industry.

4. Materials and Methods

4.1. Sample Collection and RNA Sequencing

Leaf samples were collected from plants exhibiting viral symptoms in open fields of pepper and tomato in Vietnam. We pooled leaf samples collected from the same geographical regions and host plants. Description of samples used in this study was were summarized in Table 2. Leaf samples were ground using pestle and mortar in the presence of liquid nitrogen. Total RNAs were extracted using RNeasy plant mini kit (Qiagen, Hilden, Germany). Extracted total RNAs were subjected to library preparation using NEBNext Ultra RNA Library Prep Kit for Illumina according to manufacturer's instruction (NEB, Ipswich, MA, U.S.A.). Detailed library preparation is described in the previous study [39]. The prepared libraries were paired-end (2 × 100 bp) sequenced by HiSeq2000 system (Macrogen, Seoul, Korea).

Table 2. Detailed Information for RNA-Seq Libraries and RNA-Seq Results.

Region *	Host Plant	Library	Total Read Bases (bp)	Total Reads	GC (%)
Dong Anh	Tomato	DAT	5,592,948,326	55,375,726	39.75
Gia Lam	Tomato	GLT	6,595,539,572	65,302,372	42.59
Vinh Phuc	Tomato	VPT	6,078,200,604	60,180,204	42.64
Ninh Binh	Tomato	NBT	4,899,434,654	48,509,254	41.2
Bao Loc city	Tomato	BLT	6,453,511,150	63,896,150	42.39
Duc Trong	Tomato	DTT	8,004,172,432	79,249,232	42.7
Don Duong	Tomato	DDT	7,918,928,432	78,405,232	43.8
Dalat	Tomato	DCT	8,257,251,768	81,754,968	43.67
DongAnh	Chili pepper	DACP	5,522,312,764	54,676,364	41.07
Gia Lam	Chili pepper	GLCP	4,912,867,452	48,642,252	41.24
Vinh Phuc	Chili pepper	VPCP	5,229,970,486	51,781,886	40.54
Vinh Phuc	Bell pepper	VPBP	5,745,072,910	56,881,910	42.45
Don Duong	Bell pepper	DDBP	8,473,848,288	83,899,488	44.83
Vinh Phuc	Hot pepper	VPHP	5,498,428,890	54,439,890	42.86
Ninh Binh	Hot pepper	NBHP	6,356,471,966	62,935,366	42.25
Don Duong	Red pepper	DDRP	8,339,452,436	82,568,836	43.24

* Geographical regions, host plants, and library names were described. Total read bases, total reads, GC percentage for each library by RNA-Seq were also provided.

4.2. Bioinformatic Analyses

Raw sequence files from each library were de novo assembled by the Trinity program with default parameters [40]. The assembled contigs from each library were subjected to a BLASTN search against viral genome database of National Center for Biotechnology Information (NCBI). The obtained virus-associated contigs were again subjected to a BLASTX search against NCBI's non-redundant protein (NR) database to eliminate endogenous virus-like sequences. Finally, we identified viruses infecting tomato and pepper plants based on virus-associated contigs. We mapped raw sequence reads on the reference genomes of identified viruses using the Burrows–Wheeler Aligner (BWA) program with default parameters (http://bio-bwa.sourceforge.net/). The number of mapped reads for identified virus was calculated using bbmap.sh implemented in BBMap program (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/). The raw data are available at the NCBI database with the BioProject number PRJNA636575.

4.3. Construction of Phylogenetic Trees

To generate phylogenetic trees, the assembled genome sequence of each virus was subjected to a BLASTN search against NCBI's GenBank. We retrieved the top ten homologous viral genome sequences for individual virus or viroid. Nucleotide sequences were aligned by ClustalW program [41]. Phylogenetic trees were constructed by MEGA7 program using the maximum likelihood method based on the JTT matrix-based model with a bootstrap of 1000 replicates [42].

4.4. Reverse Transcription–Polymerase Chain Reaction (RT-PCR) to Validate Infection of Identified Viruses and Viroids

RT-PCR was carried out using virus-specific primers (Table 2). The RT-PCR reaction was conducted using the Diastar^{TM} Onestep RT-PCR kit (SolGent, Daejeon, Korea) following conditions based on the manufacturer's instructions. The cycling conditions were 50 °C for 30 min, 95 °C for 15 min, followed by 30 cycles at 95 °C for 20 sec, 50 °C for 40 sec, and 72 °C for 1 min, with a final extension at 72 °C for 5 min. The PCR products were confirmed by gel electrophoresis with 1Kb DNA marker (Bioneer, Daejeon, Korea) in 1% agarose gel with TAE buffer.

References

1. Blancard, D. *Tomato Diseases: Identification, Biology and Control: A Colour Handbook*; CRC Press: Boca Raton, FL, USA, 2012.

2. Ganesan, M.; Rajesh, M.; Solairaj, P.; Senthilkumar, T. Tomato as a pioneer in health management. *Int. J. Pharm. Chem. Biol. Sci.* **2012**, *2*, 210–217.

3. Faustino, J.; Barroca, M.; Guiné, R. Study of the Drying Kinetics of Green Bell Pepper and Chemical Characterization. *Food Bioprod. Process.* **2007**, *85*, 163–170. [CrossRef]

4.　Blawid, R.; Van, D.; Maiss, E. Transreplication of a Tomato yellow leaf curl Thailand virus DNA-B and replication of a DNAß component by Tomato leaf curl Vietnam virus and Tomato yellow leaf curl Vietnam virus. *Virus Res.* **2008**, *136*, 107–117. [CrossRef] [PubMed]

5.　Pernezny, K.; Roberts, P.D.; Murphy, J.F.; Goldberg, N.P. *Compendium of Pepper Diseases*; APS Press: St. Paul, MN, USA, 2003.

6.　Pospieszny, H.; Borodynko-Filas, N.; Hasiów-Jaroszewska, B.; Czerwonka, B.; Elena, S.F. An assessment of the transmission rate of Tomato black ring virus through tomato seeds. *Plant. Prot. Sci.* **2019**, *56*, 9–12. [CrossRef]

7.　Thakur, H.; Jindal, S.K.; Sharma, A.; Dhaliwal, M.S. Chilli leaf curl virus disease: A serious threat for chilli cultivation. *J. Plant. Dis. Prot.* **2018**, *125*, 239–249. [CrossRef]

8.　Yanagisawa, H.; Matsushita, Y. Host ranges European Journal of Plant Pathology and seed transmission of Tomato planta macho viroid and Pepper chat fruit viroid. *Eur. J. Plant. Pathol.* **2017**, *149*, 211–217. [CrossRef]

9.　Constable, F.E.; Chambers, G.A.; Penrose, L.; Daly, A.; Mackie, J.; Davis, K.; Rodoni, B.; Gibbs, M. Viroid-infected Tomato and Capsicum Seed Shipments to Australia. *Viruses* **2019**, *11*, 98. [CrossRef]

10.　Batuman, O.; Turini, T.A.; Oliveira, P.V.; Rojas, M.R.; Macedo, M.; Mellinger, H.C.; Adkins, S.; Gilbertson, R.L. First Report of a Resistance-Breaking Strain of Tomato spotted wilt virus Infecting Tomatoes With the Sw-5 Tospovirus-Resistance Gene in California. *Plant. Dis.* **2017**, *101*, 637. [CrossRef]

11.　Ghanim, M.; Czosnek, H. Tomato Yellow Leaf Curl Geminivirus (TYLCV-Is) Is Transmitted among Whiteflies (*Bemisia tabaci*) in a Sex-Related Manner. *J. Virol.* **2000**, *74*, 4738–4745. [CrossRef] [PubMed]

12.　Massart, S.; Olmos, A.; Jijakli, H.; Candresse, T. Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Res.* **2014**, *188*, 90–96. [CrossRef] [PubMed]

13.　Roossinck, M.J.; Martin, D.P.; Roumagnac, P. Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology* **2015**, *105*, 716–727. [CrossRef] [PubMed]

14.　Wu, Q.; Ding, S.-W.; Zhang, Y.; Zhu, S. Identification of Viruses and Viroids by Next-Generation Sequencing and Homology-Dependent and Homology-Independent Algorithms. *Annu. Rev. Phytopathol.* **2015**, *53*, 425–444. [CrossRef]

15.　Candresse, T.; Marais, A.; Sorrentino, R.; Faure, C.; Theil, S.; Cadot, V.; Rolland, M.; Villemot, J.; Rabenstein, F. Complete genomic sequence of barley (Hordeum vulgare) endornavirus (HvEV) determined by next-generation sequencing. *Arch. Virol.* **2015**, *161*, 741–743. [CrossRef] [PubMed]

16.　Yan, T.; Zhu, J.-R.; Di, D.; Gao, Q.; Zhang, Y.; Zhang, A.; Yan, C.; Miao, H.; Wang, X.-B. Characterization of the complete genome of Barley yellow striate mosaic virus reveals a nested gene encoding a small hydrophobic protein. *Virology* **2015**, *478*, 112–122. [CrossRef] [PubMed]

17.　Bae, M.; Jo, Y.; Choi, H.; Tran, P.-T.; Kim, K.-H. First report of tomato mosaic virus isolated from tomato and pepper in Vietnam. *J. Plant. Pathol.* **2018**, *101*, 181. [CrossRef]

18.　Choi, H.; Jo, Y.; Tran, P.-T.; Kim, K.-H. First report of ageratum yellow vein virus infecting tomato in Vietnam. *J. Plant. Pathol.* **2019**, *101*, 1267. [CrossRef]

19.　Cuong, H.V.; Hai, L.; Tiep, T.; Hao, N. Molecular characterization of Tomato leaf curl Hainan virus and Tomato leaf curl Hanoi virus in Vietnam. *Int. Soc. Southeast. Asian Agric. Sci. J.* **2011**, *2*, 70–82.

20.　Ha, C.; Revill, P.; Harding, R.M.; Vu, M.; Dale, J.L. Identification and sequence analysis of potyviruses infecting crops in Vietnam. *Arch. Virol.* **2007**, *153*, 45–60. [CrossRef]

21.　Maree, H.J.; Fox, A.; Al Rwahnih, M.; Boonham, N.; Candresse, T. Application of HTS for Routine Plant Virus Diagnostics: State of the Art and Challenges. *Front. Plant. Sci.* **2018**, *9*, 9. [CrossRef] [PubMed]

22.　Revill, P.; Ha, C.V.; Porchun, S.C.; Vu, M.T.; Dale, J.L. The complete nucleotide sequence of two distinct geminiviruses infecting cucurbits in Vietnam. *Arch. Virol.* **2003**, *148*, 1523–1541. [CrossRef] [PubMed]

23.　Jo, Y.; Choi, H.; Kim, S.-M.; Kim, S.-L.; Lee, B.C.; Cho, W.K. The pepper virome: Natural co-infection of diverse viruses and their quasispecies. *BMC Genom.* **2017**, *18*, 453. [CrossRef]

24.　Li, R.; Gao, S.; Hernandez, A.G.; Wechter, W.P.; Fei, Z.; Ling, K.-S. Deep Sequencing of Small RNAs in Tomato for Virus and Viroid Identification and Strain Differentiation. *PLoS ONE* **2012**, *7*, e37127. [CrossRef] [PubMed]

25.　Xu, C.; Sun, X.; Taylor, A.; Jiao, C.; Xu, Y.; Cai, X.; Wang, X.; Ge, C.; Pan, G.; Wang, Q.; et al. Diversity, Distribution, and Evolution of Tomato Viruses in China Uncovered by Small RNA Sequencing. *J. Virol.* **2017**, *91*, e00173-17. [CrossRef] [PubMed]

26. Fadhila, C.; Lal, A.; Vo, T.T.; Ho, P.T.; Hidayat, S.H.; Lee, J.; Kil, E.-J.; Lee, S. The threat of seed-transmissible pepper yellow leaf curl Indonesia virus in chili pepper. *Microb. Pathog.* **2020**, *143*, 104132. [CrossRef] [PubMed]

27. Kil, E.-J.; Park, J.; Choi, E.-Y.; Byun, H.-S.; Lee, K.-Y.; An, C.G.; Kim, C.-S. Seed transmission of Tomato yellow leaf curl virus in sweet pepper (*Capsicum annuum*). *Eur. J. Plant. Pathol.* **2017**, *150*, 759–764. [CrossRef]

28. Kil, E.-J.; Kim, S.; Lee, Y.-J.; Byun, H.-S.; Park, J.; Seo, H.; Kim, C.-S.; Shim, J.-K.; Lee, J.-H.; Kim, J.-K.; et al. Tomato yellow leaf curl virus (TYLCV-IL): A seed-transmissible geminivirus in tomatoes. *Sci. Rep.* **2016**, *6*, 19013. [CrossRef]

29. Pospieszny, H.; Borodynko, N.; Hasiów-Jaroszewska, B.; Rymelska, N.; Elena, S.F. Transmission rate of two Polish Tomato torrado virus isolates through tomato seeds. *J. Gen. Plant. Pathol.* **2018**, *85*, 109–115. [CrossRef]

30. Pérez-Padilla, V.; Fortes, I.M.; Romero-Rodríguez, B.; Arroyo-Mateos, M.; Castillo, A.G.; Moyano, C.; De León, L.; Moriones, E. Revisiting Seed Transmission of the Type Strain of Tomato yellow leaf curl virus in Tomato Plants. *Phytopathology* **2020**, *110*, 121–129. [CrossRef]

31. Jo, Y.; Bae, J.-Y.; Kim, S.-M.; Choi, H.; Lee, B.C.; Cho, W.K. Barley RNA viromes in six different geographical regions in Korea. *Sci. Rep.* **2018**, *8*, 13237. [CrossRef] [PubMed]

32. Jo, Y.; Kim, S.-M.; Choi, H.; Yang, J.W.; Lee, B.C.; Cho, W.K. Sweet potato viromes in eight different geographical regions in Korea and two different cultivars. *Sci. Rep.* **2020**, *10*, 1–16. [CrossRef] [PubMed]

33. Rojas, M.R.; Macedo, M.A.; Maliano, M.R.; Soto-Aguilar, M.; Souza, J.O.; Briddon, R.W.; Kenyon, L.; Bustamante, R.F.R.; Zerbini, F.M.; Adkins, S.; et al. World Management of Geminiviruses. *Annu. Rev. Phytopathol.* **2018**, *56*, 637–677. [CrossRef] [PubMed]

34. Sabanadzovic, S.; Valverde, R.A. Properties and detection of two cryptoviruses from pepper (*Capsicum annuum*). *Virus Genes* **2011**, *43*, 307–312. [CrossRef] [PubMed]

35. Sabanadzovic, S.; Valverde, R.A.; Brown, J.K.; Martin, R.R.; Tzanetakis, I. Southern tomato virus: The link between the families Totiviridae and Partitiviridae. *Virus Res.* **2009**, *140*, 130–137. [CrossRef] [PubMed]

36. Choi, G.-S.; Kim, J.-H.; Lee, D.-H.; Kim, J.-S.; Ryu, K.H. Occurrence and Distribution of Viruses Infecting Pepper in Korea. *Plant. Pathol. J.* **2005**, *21*, 258–261. [CrossRef]

37. Olawale, A.; Samuel, B.O.; Solomon, A.S.O.; Kumar, P.L. Surveys of virus diseases on pepper (*Capsicum* spp.) in South-west Nigeria. *Afr. J. Biotechnol.* **2015**, *14*, 3198–3205.

38. Sepúlveda, P.R.; Larraín, P.S.; Quiroz, C.E.; Rebufel, P.A.; Graña, F.S. Identification and incidence of pepper viruses in north central Chile and its association with vectors. *Agric. Tec.* **2005**, *65*, 235–245.

39. Jo, Y.; Lian, S.; Chu, H.; Cho, J.K.; Yoo, S.-H.; Choi, H.; Yoon, J.-Y.; Choi, S.-K.; Lee, B.C.; Cho, W.K. Peach RNA viromes in six different peach cultivars. *Sci. Rep.* **2018**, *8*, 1844. [CrossRef]

40. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef]

41. Larkin, M.; Blackshields, G.; Brown, N.; Chenna, R.; Mcgettigan, P.; McWilliam, H.; Valentin, F.; Wallace, I.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef] [PubMed]

42. Lewis, P.O.; Kumar, S.; Tamura, K.; Nei, M. MEGA: Molecular Evolutionary Genetics Analysis, Version 1.02. *Syst. Biol.* **1995**, *44*, 576. [CrossRef]