Metatranscriptomic Analysis of Plant Viruses in Imported Pear and Kiwifruit Pollen

Hyo-Jeong Lee and Rae-Dong Jeong © *

Department of Applied Biology, Institute of Environmentally Friendly Agriculture, Chonnam National University, Gwangju 61185, Korea

(Received on March 29, 2022; Revised on April 28, 2022; Accepted on April 28, 2022)

Pollen is a vector for viral transmission. Pollen-mediated viruses cause serious economic losses in the fruit industry. Despite the commercial importance of pollen-associated viruses, the diversity of such viruses is yet to be fully explored. In this study, we performed metatranscriptomic analyses using RNA sequencing to investigate the viral diversity in imported apple and kiwifruit pollen. We identified 665 virus-associated contigs, which corresponded to four different virus species. We identified one virus, the apple stem grooving virus, from pear pollen and three viruses, including citrus leaf blotch virus, cucumber mosaic virus, and lychnis mottle virus in kiwifruit pollen. The assembled viral genome sequences were analyzed to determine phylogenetic relationships. These findings will expand our knowledge of the virosphere in fruit pollen and lead to appropriate management of international pollen trade. However, the pathogenic mechanisms of pollen-associated viruses in fruit trees should be further investigated.

Keywords: kiwifruit, pear, pollen, RNA sequencing

*Corresponding author.
Phone) +82-62-530-2075, FAX) +82-62-530-2069
E-mail) jaed2@jnu.ac.kr

Handling Editor : Ju-Yeon Yoon

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Articles can be freely viewed online at www.ppjonline.org.
**Virome Analysis in Fruit Tree Pollen**

(Maliogka et al., 2018). The detection of viruses by molecular-biology methods such as polymerase chain reaction, quantitative real-time polymerase chain reaction (qPCR), and loop-mediated isothermal amplification is more sensitive than ELISA and offers multiple detection possibilities (López et al., 2003; Lu et al., 2018; Malandraki et al., 2017; Osman et al., 2015). However, molecular biological methods can only detect known viruses, and they have limited ability to detect recently discovered viruses and to discover novel viruses. These problems can be overcome using next generation sequencing (NGS) technology, wherein the sequence data of all putative viral agents present in the sample can be generated without the prior knowledge of the viral genome (Radford et al., 2012).

High-throughput sequencing is a highly efficient bioinformatics tool with the ability to process a larger number of samples simultaneously, thus reducing sequencing costs (Osaki et al., 1999). Additionally, the ease with which samples and libraries can be prepared is another key advantage (Huang et al., 2019). High-throughput screening can be used to simultaneously detect RNA viruses, DNA viruses, and viroids in a variety of plant samples containing very low titers. Plant viruses can be easily diagnosed and identified using this innovative technology (Qian et al., 2014).

NGS has been used in plant virology since 2009 (Adams et al., 2009; Al Rwahnih et al., 2009; Kreuze et al., 2009). This advent of NGS revolutionized plant virology as a powerful method for identifying plant viruses without any a priori knowledge (Massart et al., 2017). NGS has been used in studies including, but not limited to, the discovery of novel viruses, detection and identification of known pathogens, analysis of genomic diversity and evolution, and pathogen epidemiology (Hadidi et al., 2016). In addition, the technology has provided new insights through the identification of the causes of plant viral disease in crops, understanding the abundance and diversity of viruses, screening specific viruses in suspect plants, and detecting asymptomatic viruses (Akinyemi et al., 2016; Roossinck et al., 2015). NGS technology is progressively reaching the field of plant virus diagnostics, and is also making an impact in the field of quarantine (Martin et al., 2016; Massart et al., 2014). So far, several plant viral metagenomic studies have identified novel viruses and viroids in several crop species, including apple (Jakovljevic et al., 2017; Wright et al., 2020), grapevine (Al Rwahnih et al., 2011; Czotter et al., 2018; Jo et al., 2018c), citrus (Matsumura et al., 2017), lilies (Jo and Cho, 2018; Li et al., 2018), vanilla (Grisoni et al., 2017), barley (Jo et al., 2018a), pepper (Jo et al., 2017), and peach (Jo et al., 2018b).

In this study, comprehensive bioinformatics analyses of plant viruses in imported pear and kiwifruit pollen were performed using RNA-seq to investigate viruses and viroids transmitted through imported pollen. In addition, we revealed the diversity of viruses and viroids infecting pear and kiwifruit pollen.

### Materials and Methods

**Sample preparation.** For artificial pollination, 20 g of pear and kiwifruit pollen imported from China in 2019 were purchased from the commercial market, three pieces each. These pollen samples were pooled, grounded in liquid nitrogen, and stored at −80°C.

**RNA extraction and library preparation for RNA-seq.** Pollen RNA was extracted using the IQuick Plus Plant RNA Extraction Mini Kit (iNtRON, Seongnam, Korea) according to the manufacturer’s instructions. The quantity and quality of purified RNA samples were measured and evaluated using the Biochrom Libra S70 Double Beam Spectrophotometer (Biochrom Ltd., Cambridge, UK). The quality of RNA was measured with the BioAnalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA) and RNA 6000 nano chip. Ribosomal RNA was removed from total RNA using a Ribo-zero rRNA removal kit for plants (Ilumnina, San Diego, CA, USA), and RNA was purified and recovered using RNA Clean XP. The TruSeq stranded total RNA library prep kit (Illumina) was used to construct a cDNA library. The size, concentration, and quality of the generated library were confirmed using the Agilent D5000 ScreenTape system (Agilent Technologies) and LightCycle qPCR; and the flow cell in which the cluster was formed through the library clustering process was sequenced in Illumina Hiseq 4000, generating 151 bp paired-end reads. For the generated raw data, the base quality score was confirmed with the Fast QC program, and low-quality data (≤ Q20) and the adapter at the read end were removed using the Trim galore program before analysis.

**Transcriptome assembly and virus identification.** The raw data obtained from the library was subjected to de novo transcriptome assembly using gsAssembler v2.8 (Roche Diagnostics, Branford, CT, USA) using the default parameters (Wylie et al., 2014). From the contig obtained through the de novo transcriptome assembly, the contig blasted for e-value < 1e−10 for the mitochondria and chloroplast, and other host sequences of 200 bp or less in length was removed. To identify virus sequences in the library, the assembled contigs in the transcriptome were compared to NCBI viral reference database (https://www.ncbi.nlm.]
nih.gov/genome/viruses/) (Morgulis et al., 2008) using BLASTN, which is faster and more reliable for virus identification than other sequence similarity programs, and a cut-off E-value of $1 \times 10^{-5}$. Only virus-associated contigs were obtained for further pollen virome analyses.

**Viral sequence mapping and genome assembly.** To generate partial or complete viral genomic sequences, the pollen transcriptomes were first blasted against complete reference sequences of viruses. Contigs were selected based on the highest similarity to genomes of viruses in GenBank. For sequence alignment to the reference viral genome, virus-like reads were mapped to each reference viral genome using Geneious Prime 2020.2 (Kearse et al., 2012) and consensus sequences for each mapping file were generated based on a threshold of 95% identity.

**Phylogenetic analyses of identified viruses.** To analyze molecular genetic relationships for identified viruses, the complete or partial genome sequences of the coat protein (CP) of each virus were used in this study. For analysis of the CP gene sequence for each virus, different isolates were aligned using ClustalW in BioEdit 7.2 (Thompson et al., 2003). Phylogenetic analysis was performed using MEGA 7 (Tamura et al., 2011), and the maximum-likelihood algorithm was used to estimate the number of nucleotide substitutions per site in the Neighbor Junction (Rivarez et al., 2021) or Kimura 2 parameter method, then bootstrapped with 1,000 replications. The isolates were presented with NCBI accession numbers in the phylogenetic trees.

**Confirmation of presence of identified virus by reverse transcription polymerase chain reaction.** To confirm the presence of identified viruses by RNA-seq in pollen samples, reverse transcription polymerase chain reaction (RT-PCR) was performed using each virus-specific primer set to amplify genomic regions corresponding to the CP genes. Each of the 20 μl reactions included 0.4 μM forward primer, 0.4 μM reverse primer, 10 μl 2× SuPrimeScript RT Premix (Genet Bio, Daejeon, Korea), 2 μl template, and 6 μl DEPC-treated water. Amplification was performed under the following conditions: a reverse transcription step of 30 min at 50°C; initial denaturation at 95°C for 5 min; 35 cycles of 95°C for 30 s, 56°C for 30 s, and 72°C for 1 min; and a final extension of 5 min at 72°C. The RT-PCR products were analyzed through electrophoresis on 1.2% (w/v) agarose gel in 0.5× TBE buffer with RedSafe nucleic acid staining (iNtRON), and visualized under UV light. The size of the resulting RT-PCR product was confirmed by comparison with a 100-bp DNA ladder (iNtRON). In addition, the amplicons were cloned into the pGEM-T Easy Vector (Promega, Madison, WI, USA), and sequences were confirmed by Sanger sequencing.

**Data availability.** The raw dataset generated in the present study will be available upon publication in the NCBI Sequence Read Archive (SRA) repository under accession numbers SRR12614248 and SRR12614749. The viral genome sequences obtained from this study were also deposited in GenBank with their individual accession numbers.

## Results

### Sample collection and library generation.** Transcriptome data were analyzed by RNA-seq for the identification of viruses in pear and kiwifruit pollen. Total RNA was isolated from pear and kiwifruit pollen samples imported from China, and two different libraries were prepared. Transcriptome data were obtained using paired-end sequencing from two different libraries (pear and kiwifruit) by HiSeq 4000, and the raw data of the sequenced reads by RNA-seq were deposited in an SRA database with their respective accession numbers (SRR12614248 and SRR12614749) (Table 1). The obtained raw data size was 7.2 Gb from pear pollen and 6.15 Gb from kiwifruit pollen, respectively. The number of obtained reads were 47,858,250 from pear pollen and 10,783,664 from kiwifruit pollen. The number of contigs, 82,285 from pear pollen and 271,201 from kiwifruit pollen, were obtained through data trimming and *de novo* transcriptome assembly (Table 1).

### Identification of viruses from pollens. To identify virus-associated reads and contigs, the raw reads and assembled contigs were utilized and a BLASTN search against a virus reference genome database from NCBI was done. The number of virus-associated reads was 15,720 in pear pol-

<table>
<thead>
<tr>
<th>Name of library</th>
<th>Total read bases (bp)</th>
<th>Total reads</th>
<th>GC (%)</th>
<th>Clean up contig (e-value $\leq 1 \times 10^{-5}$)</th>
<th>SRA accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pear</td>
<td>7,226,595,750</td>
<td>47,858,250</td>
<td>48.29</td>
<td>588</td>
<td>SRR12614248</td>
</tr>
<tr>
<td>Kiwifruit</td>
<td>6,158,333,264</td>
<td>10,783,664</td>
<td>50.5</td>
<td>1,887</td>
<td>SRR12614749</td>
</tr>
</tbody>
</table>

NGS, next generation sequencing; SRA, NCBI Sequence Read Archive.
Virome Analysis in Fruit Tree Pollen

223

Fig. 1. The number and proportion of virus-associated reads and contigs in pear and kiwifruit pollen. (A) The number of virus-associated reads in fruit pollen. (B) The number of virus-associated contigs in fruit pollen. (C) The proportion of identified viruses based on the number of virus-associated contigs in pear pollen. (D) The proportion of identified viruses based on the number of virus-associated contigs in kiwifruit pollen. ASGV, apple stem grooving virus; CLBV, citrus leaf blotch virus; CMV, cucumber mosaic virus; LycMoV, lychins mottle virus.

Viral genome assembly. Virus-associated reads were mapped to each reference viral genome to obtain complete or near-complete genomes of viruses isolated from pear and kiwifruit pollen. ASGV isolated from pear pollen was mapped to 99.9% (Fig. 2A), while three different viruses isolated from kiwifruit pollen were mapped to 93.4 to 100% (Fig. 2B-D). In addition, the virus that obtained a 100% complete genome by read mapping was CLBV in kiwifruit pollen. The assembled contigs identified in each pollen were confirmed through a BLAST search, and as a result, ASGV was identified from pear pollen, and CLBV, CMV, and LycMoV were identified from kiwifruit pollen. The sequences of the CP of each virus were deposited in GenBank (Table 2).

Phylogenetic analyses of identified viruses. Phylogenetic analyses were conducted to determine the evolutionary relationship of the viruses identified from pear and kiwifruit pollen using RNA-seq. We retrieved CP sequences of 28 ASGV isolates, 29 CLBV isolates, and 21 CMV isolates including the identified isolates from this study from the NCBI database for pear and kiwifruit pollen (Fig. 3).

A phylogenetic analysis of the CP sequences of the ASGV-P01 (LC579904) identified from pear pollen, as well as other isolates from different countries, showed two separated clades; and the isolate was closely related to the ASGV isolate from China (AY886760, Pyrus serotina Reld) (Fig. 3A). For CLBV, the phylogenetic tree of CLBV showed two separate clades; with the CLBV-KP02 (LC579912) identified from kiwifruit pollen being closely related to CLBV-20180512 (MH339884, Actinidia sp.), CLBV-20180505 (MH339877, Actinidia sp.), and CLBV-20180558 (MH339930, Actinidia sp.) from China (Fig. 3B). Phylogenetic analysis of 21 CMV isolates revealed that CMV-KP01 identified in this study clustered with the American isolate (M21464) from subgroup II (Fig. 3C). In the case of LycMoV, no phylogenetic analysis was performed due to the lack of available genome sequences of LycMoV deposited in the GenBank.
Validation of the identified viruses using RT-PCR. To confirm the presence of viruses identified in pear and kiwifruit pollen by RNA-seq, RT-PCR was performed using specific primers for the detection of these viruses. Virus-specific primers were designed based on the CP regions of identified viruses (Table 3). The results of RT-PCR were similar to those of RNA-seq (Fig. 4). Infection of ASGV in pear pollen was confirmed by RT-PCR (Fig. 4A). In addition, infections of CLBV, CMV, and LycMoV were validated using RT-PCR (Fig. 4B).

Discussion

Metatranscriptomics using high-throughput sequencing is
a powerful and straightforward genome analysis technique for revealing the composition and abundance of viral communities, also referred to as the virome, in various environmental conditions, host species, and geographical distributions (Osaki et al., 1999; Rivarez et al., 2021). Due to the ease of identification of known or unknown viral pathogens from extremely large amounts of sequence data, RNA-seq is considered as an ideal technique for understanding the complexity of virus genomes and viral mutation in a certain environment. Thus, it has been introduced into more plant virome studies and diagnostic purposes (Villamor et al., 2019). Virome studies on fruit trees using RNA-seq have been extensively employed in the past few years (Jo et al., 2018a, 2020; Kim et al., 2022). However, few studies have been performed on the virome in fruit pollen samples (Jonghe et al., 2018). In this study, we used a metatranscriptomic pipeline using RNA-seq to define viruses infecting pear and kiwifruit pollen imported from China. RNA-seq revealed that the assembled virus-associated reads nearly completed assembly corresponding to target virus genomes, including ASGV, CLBV, CMV, and LycMoV. Only ASGV, one of the most common viruses infecting pear trees worldwide, was identified from pear pollen; and CLBV, CMV, and LycMoV, which are already reported in Korea, were identified from kiwifruit pollen, indicating that this study increases knowledge of viral diversity, also referred as virosphere, in fruit pollen.

From a biosecurity perspective, fruit pollen is a potential

---

**Table 3. Information of primers used for RT-PCR**

<table>
<thead>
<tr>
<th>Virus name</th>
<th>Primer name</th>
<th>Sequence 5′-3′</th>
<th>Amplicon size (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple stem grooving virus</td>
<td>ASGV-F</td>
<td>ATGAGTTTTGGAAGACGTGCTTCAA</td>
<td>714</td>
<td>Shim et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>ASGV-R</td>
<td>CTAACCCCTCCAGTTTCAAGTTACT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citrus leaf blotch virus</td>
<td>CLBV-F</td>
<td>AGAATCTTTCAACTGCCAG</td>
<td>1180</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>CLBV-R</td>
<td>ACGCCAGTTTACATCTCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cucumber mosaic virus</td>
<td>CMV-F</td>
<td>TGGTCGTCCAACATTAACCCAC</td>
<td>320</td>
<td>Kim et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>CMV-R</td>
<td>TACTGATAAACCCACGTACCAGTGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lychnis mottle virus</td>
<td>LycMoV-F</td>
<td>ATGGGTTCAGGGCCGATA</td>
<td>695</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>LycMoV-R</td>
<td>GTAACCACATAGCCAGA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 3.** Phylogenetic trees for the identified ASGV, CLBV, and CMV. CP sequences of assembled virus genome sequences, including ASGV (A), CLBV (B), and CMV (C), and other virus isolates genome sequences obtained from GenBank were used for construction of the phylogenetic trees. Phylogenetic trees were constructed using the MEGA7 program using maximum-likelihood method and bootstrap with 1,000 replicates. The genomes of identified virus isolates derived from this study are indicated by the black color. ASGV, apple stem grooving virus; CLBV, citrus leaf blotch virus; CMV, cucumber mosaic virus; LycMoV, lychnis mottle virus; CP, coat protein; MP, movement protein.
that contamination occurred during the collection process must be assessed further. Our results also raise the possibility of cross-pollination using fruit woody plants. Metatranscriptomic analyses do not provide information on the biological activity of identified viruses; thus, their viability should be assessed further. Our results also raise the possibility that contamination occurred during the collection process due to the nature of fruit pollen. This present study did not involve back inoculation (based on Koch’s postulates) in each healthy fruit tree for viruses identified in pollen, since identified viruses are generally transmitted through grafting and seeds, not by mechanical inoculation. Notably, fruit pollen may harbor pathogenicity and function as a vehicle for viral spread. ASGV is generally seed and mechanically transmitted, and no natural vectors have been reported (Kim et al., 2021). However, a recent study revealed that *Nicotiana benthamiana* plants were infected with ASGV through pollination with ASGV-infected pollen grains derived from *Malus pumila*, indicating that pollen grains possibly served as a vector for horizontal pollen transmission of ASGV (Isogai et al., 2022). To date, 13 plant viruses, including CLBV and CMV, have been proven to naturally infect kiwifruit plants (Blouin et al., 2013). In the case of CLBV, CLBV is transmitted by grafting and seed in citrus. CMV belongs to the family Bromoviridae, which has a broad host range (with more than 1,000 host plant species in over 100 families); thus, *Actinidia* species as a host is not a new finding for CMV. CMV is generally transmitted by several types of aphids, seeds, and mechanical infections. LycMoV is a tentative new species in the family Secoviridae and was first reported in *Lychnis cognate* in South Korea (Yoo et al., 2015). Even though biological features of LycMoV has not been elucidated to date, it can be transmitted by mechanical means and seeds (Shaffler et al., 2019).

Isolates of ASGV, CLBV, and CMV identified from this study were clustered with Chinese isolates based on phylogenetic analyses, indicating that these viruses have a tight genetic association with those virus isolates in China. In addition, the viruses were of mixed infection with various variants in pollen samples, which presents the complexity for recombination, genetic diversity, and evolution. This mixed infection will possibly contribute to genetic exchanges, and thus producing new hybrid variants or novel viruses (through genetic recombination or reassortment), eventually leading to the expansion of the virosphere (Nabi et al., 2022).

In the current study, we identified ASGV, CLBV, CMV, and LycMoV from imported pear and kiwifruit pollen using RNA-seq, respectively, expanding the knowledge of the complexity of viral communities in fruit pollen. However, the pathogenesis mechanisms of these viruses through pollen transmission into pear and kiwifruit trees is still up for debate. With the increase of commercial trade of fruit pollen worldwide, further studies on fruit pollen virome are urgently needed, allowing the prevention of the inflow of high-risk viruses transmitted through pollen to other countries.

**Fig. 4.** Confirmation of the identified viruses from pear and kiwifruit pollen by RT-PCR. RT-PCR results with specific primers; ASGV (A), CLBV, CMV, LycMoV (B). The amplified PCR products were visualized by 1.2% agarose gel electrophoresis. The total RNA extracted for RNA-seq were used for RT-PCR. ASGV, apple stem grooving virus; CLBV, citrus leaf blotch virus; CMV, cucumber mosaic virus; LycMoV, lychnis mottle virus; RT-PCR, reverse transcription polymerase chain reaction.
Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by a 2019 fund by Research of Animal and Plant Quarantine Agency, South Korea.

References


Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I. and Simon, R. 2009. Complete viral genome se-


Zhang, P., Liu, Y., Liu, W., Cao, M., Massart, S. and Wang, X. 2017. Identification, characterization and full-length sequence analysis of a novel polerovirus associated with wheat leaf yel-