



# Coding-Complete Genome Sequence of a Pollen-Associated Virus Belonging to the *Secoviridae* Family Recovered from a Japanese Apricot (*Prunus mume*) Metagenome Data Set

Andrea M. Fetters,<sup>a</sup> Paul G. Cantalupo,<sup>a</sup> Tia-Lynn Ashman,<sup>a</sup> James M. Pipas<sup>a</sup>

<sup>a</sup>Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

**ABSTRACT** We report the coding-complete genome sequence of *Japanese apricot pollen-associated secovirus 1* (JAPSV1), a virus belonging to the *Secoviridae* family, recovered from Japanese apricot (*Prunus mume*) pollen that is closely related to *Peach leaf pitting-associated virus* (PLPAV). This discovery adds to the number of known pollen-associated viruses.

The 2012 edition of *Virus Taxonomy* lists just under 900 species of plant viruses (1), yet far fewer are known to be pollen associated (2–4). The detection of such viruses in asymptomatic hosts and wild plants, however, is expected to accelerate viral discovery (1). *Secoviridae* is the fourth largest family of plant viruses, and viruses of this family are spread to plant hosts primarily by insects or nematodes (1). These viruses belong to the *Picornavirales* order and have a linear, positive-sense, single-stranded RNA (ssRNA) genome with two segments (5). The longer segment, RNA1, is approximately 6 to 8 kb and encodes all proteins necessary for cytoplasmic replication, and the shorter segment, RNA2, is 2 to 4 kb and encodes the capsid and movement proteins (5).

Pollen was collected by Akagi et al. (6) from anthers of flower buds in the balloon stage from trees in an experimental orchard at Kyoto University. They extracted the total RNA using the cold-phenol approach, and RNA sequencing (RNA-seq) libraries were constructed by the Japanese company TaKaRa using TruSeq RNA sample prep kits (Illumina). Akagi et al. (6) sequenced the RNA using the TruSeq SBS kit v3-HS on an Illumina HiSeq 2000 platform, which resulted in 35,162,899 paired-end informative *Prunus mume* cv. Kairyo-Uchida-Ume reads that were each 100 bp long (SRA run accession number [DRR002284](https://www.ncbi.nlm.nih.gov/sra/DRR002284)). We recovered *Japanese apricot pollen-associated secovirus 1* (JAPSV1) from this data set using Pickaxe, a viral discovery pipeline that detects known and novel viruses in sequence data (7, 8). Briefly, Pickaxe performs quality control on the raw reads and removes host reads (namely, *Prunus mume*, here) to obtain a set of nonhost reads. Nonhost reads are aligned to Viral RefSeq (<ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/>) and are assembled with the CLC Assembly Cell (version 5; Qiagen), using default settings. The resulting contigs are annotated with a BLAST pipeline, as described previously (7), except that the BLASTX step was replaced with RAPSearch2 (9).

We identified two contigs, 6,133 bp and 2,704 bp, as RNA1 and RNA2, respectively, by BLASTN alignments, both with 93% identity to the genome segments of *Peach leaf pitting-associated virus* (PLPAV; RNA1, GenBank accession number [MK460243](https://www.ncbi.nlm.nih.gov/nuccore/MK460243); RNA2, accession number [MK460244](https://www.ncbi.nlm.nih.gov/nuccore/MK460244)). RNA2 was increased by 690 bp through the use of an in-house script that found a substantial overlap region (94 bp) between the end of the 2,704-bp contig and the beginning of another contig, bringing the final length of RNA2 to 3,374 bp. The fact that the two contigs were not originally assembled is not surprising given the difficulty of *de novo* metagenomic assembly, since assembly programs assume the presence of one organism (10, 11). Aligning the nonhost reads to

**Citation** Fetters AM, Cantalupo PG, Ashman T-L, Pipas JM. 2019. Coding-complete genome sequence of a pollen-associated virus belonging to the *Secoviridae* family recovered from a Japanese apricot (*Prunus mume*) metagenome data set. *Microbiol Resour Announc* 8:e00881-19. <https://doi.org/10.1128/MRA.00881-19>.

**Editor** Simon Roux, DOE Joint Genome Institute

**Copyright** © 2019 Fetters et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to James M. Pipas, [pipas@pitt.edu](mailto:pipas@pitt.edu).

**Received** 16 August 2019

**Accepted** 3 September 2019

**Published** 3 October 2019

each genome segment resulted in greater than 99% coverage, with an average depth (reads/base) of 2,017 for RNA1 and 4,941 for RNA2. Both genome segments have 43% GC content. We found two open reading frames (ORFs) using the latest version of ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>), with default parameters that correspond to the *Secoviridae* genome. The longer segment of most *Secoviridae* genomes codes for helicase, proteinase, and polymerase proteins. Our conserved domain search of the longer ORF (1,920 amino acids) detected those proteins in the same sequence. Similarly, our conserved domain search of the shorter ORF (997 amino acids) detected a movement protein and two coat proteins in the same sequence as that of most *Secoviridae*. We searched the latest version of the Conserved Domain Database using the default parameters to annotate both RNA1 and RNA2 (12) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Due to the high percentage identity (98%) of the JAPSV1 polyproteins with those of PLPAV estimated using a BLASTP alignment, the similar GC content of the JAPSV1 polyproteins, and the fact that no other *Secoviridae* sequences were identified in the data set, we conclude that we detected a virus in the *Secoviridae* family that is closely related to PLPAV. Although JAPSV1 is closely related to PLPAV, its detection expands our knowledge of pollen-associated plant viruses and provides a starting point for considering how such viruses are spread, as well as plant virus host switching.

**Data availability.** This whole-genome shotgun project has been deposited in the European Nucleotide Archive under the accession numbers LR594708 and LR594709. The version described in this paper is the first version, and the coding-complete genome sequence of JAPSV1 is publicly available on the Pipas Lab website (<http://pipaslab.webfactional.com/wp/wp-content/uploads/2019/07/JAPSV1.txt>). The raw data in which we found JAPSV1 are available under SRA run number DRR002284, and the genome segment sequences of PLPAV are referenced by MK460243 and MK460244 in the National Center for Biotechnology Information.

## ACKNOWLEDGMENTS

This research was supported by funding from the Charles E. Kaufman Foundation of the Pittsburgh Foundation (grant KA2017-91786) and by the Dietrich School of Arts and Sciences First Year Graduate Fellowship at the University of Pittsburgh. The research was also supported in part by the University of Pittsburgh Center for Research Computing through the resources provided.

We thank two anonymous reviewers for their comments that improved the manuscript.

T.-L.A. and J.M.P. designed the study. P.G.C. and A.M.F. performed the analyses. All authors contributed to writing the paper.

## REFERENCES

- Roossinck MJ. 2012. Plant virus metagenomics: biodiversity and ecology. *Annu Rev Genet* 46:359–369. <https://doi.org/10.1146/annurev-genet-110711-155600>.
- Card SD, Pearson MN, Clover G. 2007. Plant pathogens transmitted by pollen. *Austral Plant Pathol* 36:455–461. <https://doi.org/10.1071/AP07050>.
- Hull R. 2014. *Plant virology*. Elsevier, Philadelphia, PA.
- Jones R. 2018. Plant and insect viruses in managed and natural environments: novel and neglected transmission pathways, p 149–187. In Malmstrom C (ed), *Advances in virus research: environmental virology and virus ecology*. Elsevier, Cambridge, MA.
- Sanfaçon H, Wellink J, Le Gall O, Karasev A, van der Vlugt R, Wetzell T. 2009. *Secoviridae*: a proposed family of plant viruses within the order *Picornavirales* that combines the families *Sequiviridae* and *Comoviridae*, the unassigned genera *Cheravirus* and *Sadwavirus*, and the proposed genus *Torradovirus*. *Arch Virol* 154:899–907. <https://doi.org/10.1007/s00705-009-0367-z>.
- Akagi T, Henry IM, Morimoto T, Tao R. 2016. Insights into the *Prunus*-specific S-RNase-based self-incompatibility system from a genome-wide analysis of the evolutionary radiation of S locus-related F-box genes. *Plant Cell Physiol* 57:1281–1294. <https://doi.org/10.1093/pcp/pcw077>.
- Starrett GJ, Marcelus C, Cantalupo PG, Katz JP, Cheng J, Akagi K, Thakuria M, Rabinowits G, Wang LC, Symer DE, Pipas JM, Harris RS, DeCaprio JA. 2017. Merkel cell polyomavirus exhibits dominant control of the tumor genome and transcriptome in virus-associated Merkel cell carcinoma. *mBio* 8:e02079-16. <https://doi.org/10.1128/mBio.02079-16>.
- Cantalupo PG, Katz JP, Pipas JM. 2018. Viral sequences in human cancer. *Virology* 513:208–216. <https://doi.org/10.1016/j.virol.2017.10.017>.
- Zhao Y, Tang H, Ye Y. 2012. RAPSearch2: a fast and memory-efficient protein similarity search tool for next generation sequencing data. *Bioinformatics* 28:125–126. <https://doi.org/10.1093/bioinformatics/btr595>.
- Cantalupo PG, Pipas JM. 2019. Detecting viral sequences in NGS data. *Curr Opin Virol* 39:41–48. <https://doi.org/10.1016/j.coviro.2019.07.010>.
- Smits SL, Bodewes R, Ruiz-Gonzalez A, Baumgärtner W, Koopmans MP, Osterhaus ADME, Schürch AC. 2014. Assembly of viral genomes from metagenomes. *Front Microbiol* 5:714. <https://doi.org/10.3389/fmicb.2014.00714>.
- Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:D200–D203. <https://doi.org/10.1093/nar/gkw1129>.