

Complete Genome Sequence of a Novel Genotype of Squash Mosaic Virus Infecting Squash in Spain

Rugang Li,^a Shan Gao,^b Sven Berendsen,^c Zhangjun Fei,^{b,d} Kai-Shu Ling^a

U.S. Department of Agriculture-Agricultural Research Service, U.S. Vegetable Laboratory, Charleston, South Carolina, USA^a; Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York, USA^b; Rijk Zwaan Breeding B.V., De Lier, the Netherlands^c; U.S. Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Cornell University, Ithaca, New York, USA^d

The complete genome sequence of a new isolate of squash mosaic virus (SqMV) infecting squash plants in Spain was obtained using deep sequencing of small RNAs. The low nucleotide sequence identities, with only 87 to 88% for RNA1 and 84 to 86% for RNA2 to known SqMV isolates, suggested that this isolate belongs to a novel genotype.

Received 28 December 2014 Accepted 5 January 2015 Published 12 February 2015

Citation Li R, Gao S, Berendsen S, Fei Z, Ling K-S. 2015. Complete genome sequence of a novel genotype of squash mosaic virus infecting squash in Spain. *Genome Announcements* 3(1):e01583-14. doi:10.1128/genomeA.01583-14.

Copyright © 2015 Li et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Kai-Shu Ling, kai.ling@ars.usda.gov.

Squash mosaic virus (SqMV), of the genus *Comovirus* in the family *Secoviridae* (1), is a seed-borne virus infecting cucurbit crops worldwide (2–4). SqMV has a bipartite single-stranded RNA genome (RNA1 and RNA2) encapsidated separately with two capsid proteins. Both RNA molecules contain a genome-linked viral protein (VPg) and a poly(A) tail. Currently, two genotypes (or subgroups) that share only 88 to 89% genome nucleotide sequence identity have been recognized (2). In the present study, a new isolate (RZ-SqMV) collected in 2010 in Spain on a field for squash germplasm evaluation could not be detected by the real-time reverse transcription-PCR (RT-PCR) that was designed for SqMV (5), suggesting a greater genetic diversity than previously known among SqMV isolates. In order to obtain a full-genome sequence from the isolate RZ-SqMV, we employed deep sequencing of small RNAs (sRNAs) and assembly technology (6, 7). An sRNA library was constructed according to the published protocol (8) and sequenced in an Illumina HiSeq 2000. The sRNA sequence reads were processed and assembled based on the bioinformatics pipeline (7), resulting in two near-full genomes of RNA1 and RNA2. The authenticity of the newly assembled genomic sequences was confirmed through Sanger sequencing of additional RT-PCR products along the two genomic RNAs. The 5′ termini in both viral RNAs were obtained using rapid amplification of cDNA ends (RACE) technology. Upon treatment to a purified plant total RNA preparation with calf intestine alkaline phosphatase to dephosphorylate the RNAs and proteinase K to remove the VPg, 5′-adaptor ligation to the treated RNAs was performed either with a 5′-RACE adaptor from the Ambion First-Choice RLM-RACE kit (Austin, TX, USA) or with the Illumina 5′-RNA adapter (5′-GUUCAGAGUUCUACAGUCCGACGAUC-3′). The 5′-terminal sequences obtained by RACE for RNA1 and RNA2 were in agreement using either ligation adaptor. The two SqMV RNAs began with an identical consensus sequence (UAU-UAAA), in agreement with the 5′-terminal sequences in other comoviruses, except the Chinese isolate of SqMV, which is reported to have a unique terminal end (4). The complete genome sequence for the

isolate RZ-SqMV was obtained. With a total of 5,858 nucleotides (nt), excluding a poly(A) tail, RNA1 contains a single open reading frame (ORF) encoding a polyprotein of 1,858 amino acids (aa). RNA2 has 3,370 nt encoding two overlapping polyproteins (1,009 aa and 920 aa, respectively) initiating either at nucleotide position 219 (AUG219) or 486 (AUG486) and terminating at position 3248 with a UAG stop codon. Using the complete genome sequences of RNA1 and RNA2, RZ-SqMV was found to share only 87 to 88% sequence identities in RNA1 and 84 to 86% sequence identities in RNA2 to all known SqMV isolates from both recognized SqMV genotypes. Such low levels of sequence identity suggest that RZ-SqMV should be considered a third genotype of SqMV. The complete genome sequence obtained from this novel genotype will allow us to better understand the genetic diversity in SqMV and to improve the molecular-based detection.

Nucleotide sequence accession numbers. The full genomic sequences of RZ-SqMV were deposited in GenBank under the accession numbers [KP223323](https://www.ncbi.nlm.nih.gov/nuclot/KP223323) and [KP223324](https://www.ncbi.nlm.nih.gov/nuclot/KP223324).

ACKNOWLEDGMENTS

The sRNA deep sequencing was conducted by the Genomics Resources Core Facility at the Weill Cornell Medical College in New York City. Sanger sequencing was carried out by Functional Biosciences (Madison, WI).

We thank Andrea Gilliard and Alan Wilder for their excellent technical assistance.

This work was supported in part by the USDA, National Institute of Food and Agriculture, and SCRI 2010-600-25320 to K.-S.L. and SCRI 2012-01507-229756 to K.-S.L. and Z.F.

REFERENCES

- Sanfaçon H, Iwanami T, Karasev AV, van der Vlugt R, Wellink J, Wetzel T, Yoshikawa N. 2012. Secoviridae, p 881–899. In King AMO, Lefkowitz E, Adams MJ, Carstens EB (ed), *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier/Academic Press, London, United Kingdom.
- Haudenschild JS, Palukaitis P. 1998. Diversity among isolates of squash mosaic virus. *J Gen Virol* 79:2331–2341.

3. Han SS, Yoshida K, Karasev AV, Iwanami T. 2002. Nucleotide sequence of a Japanese isolate of squash mosaic virus. Brief report. *Arch Virol* 147: 437–443. <http://dx.doi.org/10.1007/s705-002-8332-5>.
4. Hu J, Zhou T, Liu L, Peng B, Li H, Fan Z, Gu Q. 2009. The genomic sequences of a Chinese isolate of Squash mosaic virus with novel 5' conserved ends. *Virus Genes* 38:475–477.
5. Ling K.-S, Wechter WP, Walcott RR, Keinath AP. 2011. Development of a real-time RT-PCR assay for squash mosaic virus useful for broad spectrum detection of various serotypes and its incorporation into a multiplex seed health assay. *J Phytopathol* 159:649–656. <http://dx.doi.org/10.1111/j.1439-0434.2011.01814.x>.
6. Kreuze JF, 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* 388:1–7. <http://dx.doi.org/10.1016/j.virol.2009.03.024>.
7. Li R, Gao S, Hernandez AG, Wechter WP, Fei Z, Ling K-S. 2012. Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PLoS One* 7:e37127. <http://dx.doi.org/10.1371/journal.pone.0037127>.
8. Chen Y-R, Zheng Y, Liu B, Zhong S, Giovannoni J, Fei Z. 2012. A cost-effective method for Illumina small RNA-Seq library preparation using T4 RNA ligase 1 adenylated adapters. *Plant Methods* 8:41. <http://dx.doi.org/10.1186/1746-4811-8-41>.