



# Nanopore-Based Complete Genome Sequence of a Sri Lankan Cassava Mosaic Virus (*Geminivirus*) Strain from Thailand

Ana M. Leiva,<sup>a</sup> Wanwisa Siriwan,<sup>b</sup> Diana Lopez-Alvarez,<sup>a</sup> Israel Barrantes,<sup>c</sup> Nuannapa Hemniam,<sup>b</sup> Kingkan Saokham,<sup>d</sup>  
Wilmer J. Cuellar<sup>a</sup>

<sup>a</sup>Virology Laboratory, Agrobiodiversity Research Area, International Center for Tropical Agriculture, Cali, Colombia

<sup>b</sup>Department of Plant Pathology, Faculty of Agriculture, Kasetsart University, Bangkok, Thailand

<sup>c</sup>Institute for Biostatistics and Informatics in Medicine and Ageing Research, Rostock, Germany

<sup>d</sup>Center for Agricultural Biotechnology, Kasetsart University, Nakhon Pathom, Thailand

**ABSTRACT** Sri Lankan cassava mosaic virus is an emerging pathogen in Southeast Asia. Here, we report the complete genome of a Thai isolate obtained using Nanopore technology. The isolate was collected in 2019 from the northeastern province of Surin, soon after disease eradication was reported in the country.

Sri Lankan cassava mosaic virus (SLCMV) is a circular, bipartite single-stranded DNA virus belonging to the family *Geminiviridae* (genus *Begomovirus*). It is one of several geographically distinct virus species causing cassava mosaic disease (CMD), a major disease of cassava (*Manihot esculenta* Crantz), in Africa and Asia (1). SLCMV was characterized for the first time in Sri Lanka in 1998 (2); it also occurs in India, and since 2015, it has emerged in Southeast Asia (3, 4). The occurrence of SLCMV in Thailand, the world's largest exporter of cassava, has not yet been confirmed, although unofficial reports indicate that eradication activities have been taking place in the northeastern region of the country since 2018. Here, we present the complete genome sequence of a Thai isolate of SLCMV, which was obtained using an Oxford Nanopore Technologies MinION sequencer.

The complete genome of SLCMV was obtained from an infected cassava plant collected in the province of Surin in February 2019 from a field presenting a 0.7% incidence of CMD. A phi29 rolling-circle amplification protocol (New England Biolabs, USA) was carried out with 60 ng of total DNA, which had been extracted using cetyl trimethylammonium bromide (CTAB) (5). After a 16-h reaction, 1  $\mu$ g of the amplified product was randomly sheared using a g-TUBE device (Covaris, USA), and the library was prepared according to the manufacturer's instructions (genome DNA by ligation sequencing kit, product number SQK-LSK109 [Oxford Nanopore Technologies]). Sequencing was conducted using a FLO-MIN106D (R9.5) flow cell, and base calling was performed in real time using MinKNOW v2.0.

The assembly was performed in two parallel ways (using default parameters), i.e., *de novo* assembly, using the raw Nanopore reads that passed quality control (fastq pass) with Canu v1.8 (6), and reference assembly, using Minimap2 (7) and Pilon (8) with distinct SLCMV genome sequences (isolates Attur 2 [GenBank accession numbers [KP455484](#) and [KP455485](#)], Sreekaryam 1 [[MK404225](#) and [MK404226](#)], and SLCMV-Col [[AJ314737](#) and [AJ314738](#)]). A consensus was created using the reference assembly results, and the quality was checked using Qualimap v2.2.1 (9). The resulting contigs consisted of 72,800 reads for DNA-A, with an average coverage of 15,000 $\times$ , and 70,681 reads for DNA-B, with an average coverage of 6,000 $\times$ . The assembled sequences were validated by Sanger sequencing (Macrogen, South Korea) of overlapping amplicons covering the entire circular genome components and were deposited in GenBank.

Sequence Demarcation Tool analysis (10) of the assembled DNA-A (2,759 nucleotides;

**Citation** Leiva AM, Siriwan W, Lopez-Alvarez D, Barrantes I, Hemniam N, Saokham K, Cuellar WJ. 2020. Nanopore-based complete genome sequence of a Sri Lankan cassava mosaic virus (*Geminivirus*) strain from Thailand. *Microbiol Resour Announc* 9:e01274-19. <https://doi.org/10.1128/MRA.01274-19>.

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

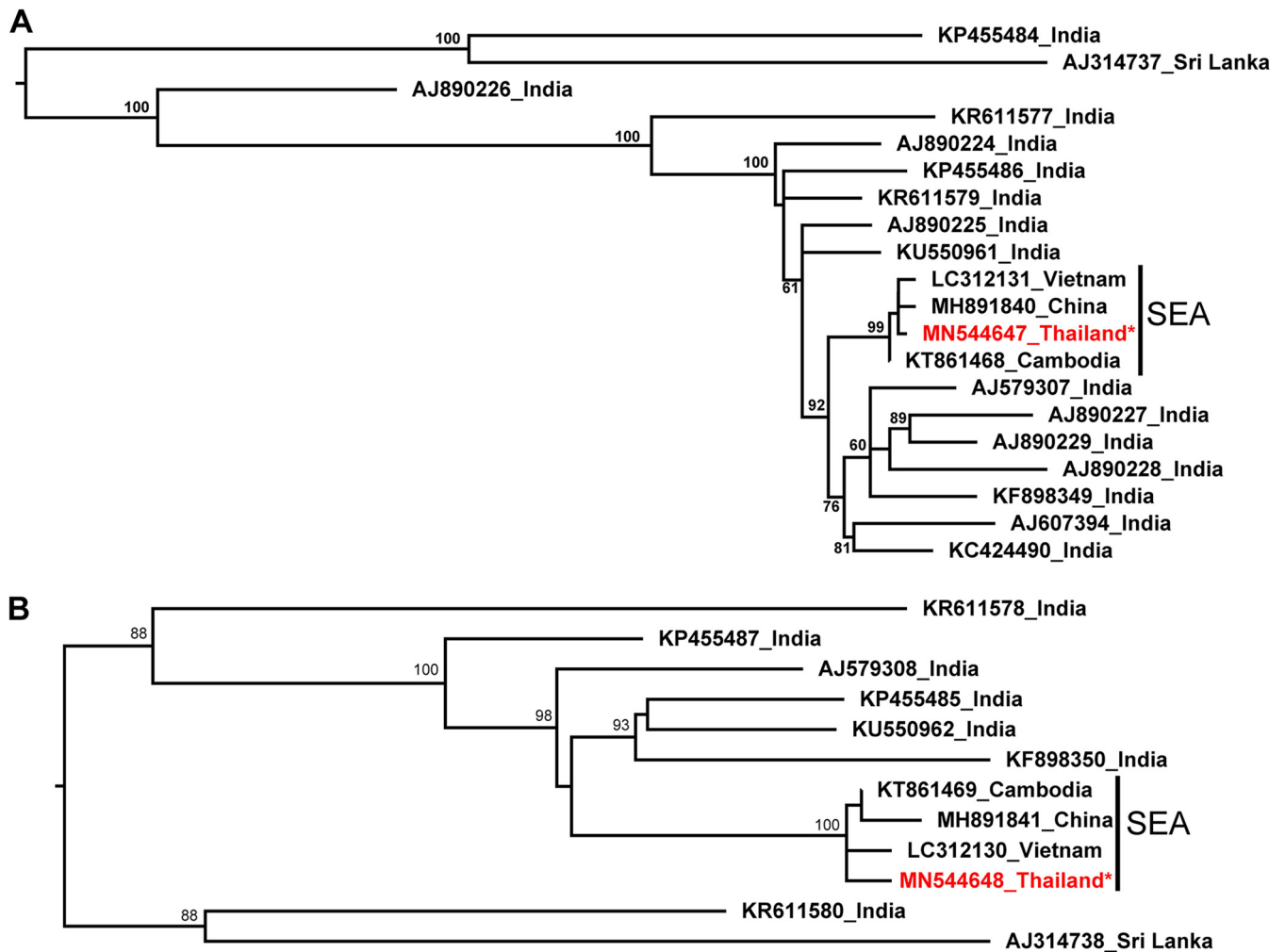
**Copyright** © 2020 Leiva et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Wilmer J. Cuellar, [w.cuellar@cgiar.org](mailto:w.cuellar@cgiar.org).

**Received** 27 October 2019

**Accepted** 8 January 2020

**Published** 6 February 2020



**FIG 1** Maximum likelihood tree based on a GTR+G model (confirmed with jModelTest2), using a bootstrap value of 1,000, comparing full SLCMV genome sequences for DNA-A (A) and DNA-B (B) components with those of the Surin1 isolate described in this report (in red). SEA, Southeast Asia.

GC content of 45.9%) and DNA-B (2,737 nucleotides; GC content of 42.7%) showed that both components share ~99.9% nucleotide identity with other Southeast Asian isolates reported in GenBank. The intergenic regions present a 5'-GGAGA-3' conserved direct repeat motif known as the iteron, with the corresponding "iteron-related domain" (FRIQS KNIFLTYPKC) at the N terminus of the Rep protein (11). A characteristic of most isolates from Southeast Asia is the presence of a G-to-A transition at position 1569 of DNA-A, which introduces a stop codon that eliminates a 7-amino-acid domain associated with virulence (12) from the C terminus of the Rep protein (AC1). Phylogenetic analysis of all available SLCMV genome sequences using IQ-TREE and jModelTest2 (13, 14) indicated that Southeast Asian isolates form a monophyletic cluster, with >99% bootstrap values, for both DNA components (Fig. 1).

**Data availability.** Genome sequences and Nanopore reads were deposited in GenBank under accession numbers [MN544647](https://www.ncbi.nlm.nih.gov/nuccore/MN544647) (DNA-A) and [MN544648](https://www.ncbi.nlm.nih.gov/nuccore/MN544648) (DNA-B) and in the Sequence Read Archive under accession number [PRJNA587722](https://www.ncbi.nlm.nih.gov/sra/PRJNA587722). A phylogenetic tree based on Nextstrain analysis (15) is available online (<https://nextstrain.org/community/pestdisplace/CMDASIA?c=virus&r=location>).

## ACKNOWLEDGMENTS

We are grateful to the staff from the Virology Laboratory, the Cassava Program at the International Center for Tropical Agriculture, and the Thai Tapioca Development Institute for supporting this work.

The research by W.S. and K.S. was partly funded by the Center of Excellence on Agricultural Biotechnology, Ministry of Higher Education, Science, Research, and Innovation (Bangkok, Thailand). This research was undertaken as part of, and was cofunded by, the CGIAR Research Program on Roots, Tubers, and Bananas and was supported by CGIAR Trust Fund contributors (<https://www.cgiar.org/funders>).

## REFERENCES

1. Legg JP, Lava Kumar P, Makesh Kumar T, Tripathi L, Ferguson M, Kanju E, Ntawuruhunga P, Cuellar W. 2015. Cassava virus diseases: biology, epidemiology and management. *Adv Virus Res* 91:85–142. <https://doi.org/10.1016/bs.aivir.2014.10.001>.
2. Saunders K, Salim N, Mali VR, Malathi VG, Briddon R, Markham PG, Stanley J. 2002. Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* 293:63–74. <https://doi.org/10.1006/viro.2001.1251>.
3. Wang HL, Cui XY, Wang XW, Liu SS, Zhang ZH, Zhou XP. 2016. First report of Sri Lankan cassava mosaic virus infecting cassava in Cambodia. *Plant Dis* 100:1029. <https://doi.org/10.1094/PDIS-10-15-1228-PDN>.
4. Uke A, Hoat TX, Quan MV, Liem NV, Ugaki M, Natsuaki KT. 2018. First report of *Sri Lankan cassava mosaic virus* infecting cassava in Vietnam. *Plant Dis* 102:2669. <https://doi.org/10.1094/PDIS-05-18-0805-PDN>.
5. Lozano I, Leiva AM, Jimenez J, Fernandez E, Carvajal-Yepes M, Cuervo M, Cuellar WJ. 2017. Resolution of cassava-infecting alphaflexiviruses: molecular and biological characterization of a novel group of potexviruses lacking TGB3. *Virus Res* 241:53–61. <https://doi.org/10.1016/j.virusres.2017.03.019>.
6. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
7. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
8. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
9. Okonechnikov K, Conesa A, Garcia-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:292–294. <https://doi.org/10.1093/bioinformatics/btv566>.
10. Muhire BM, Varsani A, Martin DP. 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9:e108277. <https://doi.org/10.1371/journal.pone.0108277>.
11. Argüello-Astorga GR, Ruiz-Medrano R. 2001. An iteron-related domain is associated to motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol* 146:1465–1485. <https://doi.org/10.1007/s007050170072>.
12. Wang D, Zhang X, Yao X, Zhang P, Fang R, Ye J. 2020. A 7-amino-acid motif of Rep protein essential for virulence is critical for triggering host defense against Sri Lankan cassava mosaic virus. *Mol Plant Microbe Interact* 33:78–86. <https://doi.org/10.1094/MPMI-06-19-0163-FI>.
13. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
14. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772. <https://doi.org/10.1038/nmeth.2109>.
15. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34:4121–4123. <https://doi.org/10.1093/bioinformatics/bty407>.