 doi:10.20944/preprints202003.0336.v1

Complete genome sequence of a distinct isolate of *Cassava common mosaic virus* (CsCMV) infecting cassava in Hainan, China

Nai-tong Yu[1,2]*, Yi Yang[3], Jun-hua Li[1], Wei-li Li[1], Zhi-xin Liu[1]*

1. Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture and Rural Affairs, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China.
2. Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Haikou 571101, China.
3. Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

*Corresponding author:
Nai-tong Yu, Email: yunaitong@163.com
Zhi-xin Liu, Email: liuzhixin@itbb.org.cn
Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences
Tel: 086-898-66890770

Email addresses:
NTY: yunaitong@163.com
YY: yiyang569@163.com
JHL: 1031631267@qq.com
WLL: 1124047447@qq.com
ZXL: liuzhixin@itbb.org.cn

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2020 doi:10.20944/preprints202003.0336.v1

Abstract: The complete genomic sequence of a *Cassava common mosaic virus* Linggao isolate (CsCMV-LG) was determined from cassava (*Manihot esculenta* Crantz) with mild leafy mosaic symptom to no symptom in China. Excluding the poly(A) tail, the CsCMV-LG genome (GenBank accession No. MT038420) is 6374 nucleotides (nts) in length, with five major open reading frames encoding a 1450-amino acids (aa) RNA-dependent RNA polymerase (RdRp), three triple gene block (TGB) proteins (231-aa, 110-aa and 95-aa), and a 229-aa coat protein (CP). Phylogenetic analysis indicated that the complete genome of the CsCMV-LG is closely related to that of CsCMV-Brazilian which has been assigned to the genus *Potexvirus*, but the sequence identity shared only 88.0%. Notable, the mild CsCMV-LG isolate can also infect *Nicotiana benthamiana* in laboratory through rub inoculation causing mild vein yellowing at 15-day post inoculation. This is the first full-length genome sequence of a distinct isolate of *Cassava common mosaic virus* (CsCMV) infecting cassava in Hainan, China.

Keywords: *Manihot esculenta* Crantz, potexvirus, *Cassava common mosaic virus*, genomics

Cassava (*Manihot esculenta* Crantz) is a tropical plant in the genus *Manihot* in the family *Euphorbiaceae*. It is the sixth largest food crop in the world, mainly grown in sub-Saharan Africa, South Asian and South America. With a production of over 270 million tonnes, it provides staple food for more than 500 million people (FAO, 2017). Besides used as a staple, cassava is also a potential raw material for bioenergy and bio-based materials. In China, cassava is an important tropical crop in southern provinces, with a planting area of 500,000 hectares and a total output of 10 million tons (FAO, 2017). Viral diseases such as Cassava mosaic disease (CMD) and Cassava brown steak disease (CBSD) are the main causes of severe economic losses in cassava plantation worldwide (Kuon et al., 2000; Mulenga et al., 2018). However, there are other cassava infecting viruses that also potential threat to cassava yields.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2020 doi:10.20944/preprints202003.0336.v1

Cassava common mosaic virus (CsCMV) is a re-emergent virus affecting cassava production in South American countries such as Brazil, Peru, Colombia and Paraguay (Legg et al., 2015). The virus has also been reported to infect Chaya (*Cnidoscolus aconitifolius*), a plant in the same family as cassava, in Mexico (Jones et al., 1998). The genome of CsCMV has been sequenced from Brazilian isolate (CsCMV-Brazilian) and it was classified into genus *Potexvirus* on the basis of viral particle morphology, genome and serology. CsCMV-Brazilian consisted of 6376 nts long single-strand RNA which potentially encoded RNA-dependent RNA polymerase (RdRp), triple gene block 1-3 (TGB1-3) and coat protein (CP) just as many other *Potexvirus* (Calvert et al., 1996).

In 2019, 16 symptomatic cassava samples with mild leafy mosaic and 12 asymptomatic samples were collected from Linggao County, Hainan Province, China. To ascertain the presence of potexvirus, total RNAs of these samples were extracted by using a Quick RNA isolation kit (Bioteke, Beijing, China) and reverse transcribed into the first strand cDNA by using *EasyScript* Reverse Transcriptase (Trans, China). PCR targeting CsCMV *RDRP* gene was performed by using specific primers CsCMV 3280F/3892R (Table 1) and the result showed that the DNA fragments of the expected sizes were obtained from all 16 diseased samples (100%) and three asymptomatic samples (25%). The DNA fragments were purified by an OMEGA gel recovery kit (Bio-Tek, USA). Each DNA fragment was cloned into the pMD18-T vector (Takara, Dalian, China), and then transformed into *E. coli* DH5a competent cells (2nd Lab, Shanghai, China). Three positive clones were randomly selected for Sanger sequencing at Invitrogen (Guangzhou, China). Bioinformatic analysis showed that the sequence was highly identical (99.9 % sequence similarity) and highly similar to the CsCMV by blastn.

 doi:10.20944/preprints202003.0336.v1

To determine the complete nucleotide sequence of this virus, another four pair-primers (Table 1) were further designed base on the complete genome sequence of CsCMV-Brazilian (GenBank accession no. U23414.1). The PCR reaction and amplification cycle were conducted as described by Li et al (Li et al., 2020). The five overlapping fragments were edited by ChromasPro software (Technelysium Pty. Ltd., Australia) and were used to assemble into the complete genome of CsCMV-LG by BioEdit software. The 5' end sequence was further determined using a 5' Rapid Amplification of cDNA Ends (5' RACE) kit (Invitrogen, Shanghai, China) according to the manufacturer's instructions. The complete genome of CsCMV-LG was deposited in GenBank under accession number MT038420. The CsCMV-LG isolate comprises 6374 nts (excluding the poly(A) tail) and has similar genome organization to potexvirus (Fig. 1A). Complete genome comparison between CsCMV-Brazilian and CsCMV-LG shows the two isolates sharing only 88.0 % sequence similarity in full length, but 94.7% and 92.1% in 76-nt 5' UTR and 115-nt 3' UTR, respectively (Table 2). Another two Potexviruses that close related to CsCMV-LG are *Tulip virus X* (AB066288.1) and *Hydrangea ringspot virus* (LC107516.1), sharing 55.3% and 55.2% sequence similarity, respectively.

Five open reading frames (ORFs) were predicted to encode proteins RdRp, TGB1, TGB2, TGB3, and CP in CsCMV-LG at nt positions 77–4429, 4432–5127, 5039–5425, 5280–5567, and 5570–6259, respectively. The similarities of these putative proteins and the corresponding ones between CsCMV-LG and CsCMV-Brazilian were 87.30-92.0% at nucleotide level and 91.7-97.8 % at amino acid level, with highest aa similarity found in CP coding region (Table 2). RdRp of CsCMV-LG is a 164.91 kDa protein and shares 94.30 % sequence identity with RdRp of CsCMV-Brazilian. The RdRp contains conserved motifs

 doi:10.20944/preprints202003.0336.v1

like in those of other potexviruses. As expected, CsCMV-LG RdRp has a viral methyltransferase motif at aa 39–336, a viral RNA helicase domain at aa 733–963, and a RdRp motif at aa 1236–1343, which are conserved motifs found in potexviruses RdRp.

In order to further determine the homology and evolutionary relationships of CsCMV-LG with other potexvirus, two phylogenetic trees based on the complete genomes nucleotide sequences were constructed by using two methods in MEGA6.0 software (Tamura et al., 2013). The two methods used were Neighbor-Joining method (Saitou and Nei, 1987) and maximum likelihood method (Tamura and Nei, 1993), each with bootstrap values of 1000 replications. The result showed that both trees showed that CsCMV-LG was the highest homology to the CsCMV-Brazilian and clearly distinct from the other potexviruses.

CsCMV is a re-emergent virus significantly affecting cassava production in many countries in South America. The main symptoms of CsCMV infection in cassava were mosaic and chlorotic symptoms in previous reports (Zanini et al., 2018). However, the diseased leaves of cassava plant present only very mild mosaic symptoms in this study and some of them were even asymptomatic (Fig. 1B). The diseased fresh leaves were ground in PBS buffer and inoculated into tobacco (*Nicotiana benthamiana*) leaves. The result showed that the mild vein yellowing symptoms were observed on tobacco leaves varying at 10-15 dpi (most plants presented the symptoms at 15 dpi) (Fig. 1C). However, CsCMV could be detected by RT-PCR starting at day-5 in the systematic leaves. This biological experiment further showed that the CsCMV-LG is a distinct isolate of the CsCMV-Brazilian.

Currently, the disease caused by CsCMV has spread to other cassava plantation areas in southern

China. It is likely that the disease would further spread without prevention and proper management. Further studies are needed in order to clarify the genetic diversity, biological characterization, and epidemiology of CsCMV in different geographic regions. Due to the wide distribution of CsCMV in Hainan and other southern provinces in China, the genome information presented in this work will allow the design of molecular tests aimed at detecting CsCMV in future.

Funding

This work was funded by the National Key R&D Program of China (2019YFD1000500), National Natural Science Foundation of China (31261140363) and Young Elite Scientists Sponsorship Program by CSTC (Project No. CSTC-QN201704).

Declarations

Consent for publication

All authors agreed to the publication of this manuscript.

Competing interests

The authors declare that they have no conflict of interest.

References

[dataset] Food and Agriculture Organization [FAO]. Food Outlook. Biannual report on global food markets. 2017; Available at: http://www.fao.org/3/a-I8080e.pdf.

Calvert, L.A., Cuervo, M.I., Ospina, M.D., Fauquet, C.M., Ramirez, B.C. 1996. Characterization of *Cassava common mosaic virus* and a defective RNA species. J Gen Virol. 77 (Pt 3): 525-530.

Jones, P., Devonshire, J., Dabek, A., Howells, C. 1998. First report of *Cassava common mosaic potexvirus* infecting Chaya (*Cnidoscolus chayamansa*) in Tuvalu. Plant Dis. 82(5): 591.

 doi:10.20944/preprints202003.0336.v1

Kuon, J.E., Qi, W., Schläpfer, P., Hirsch-Hoffmann, M., von Bieberstein, P.R., Patrignani, A., Poveda, L., Grob, S., Keller, M., Shimizu-Inatsugi, R., Grossniklaus, U., Vanderschuren, H., Gruissem, W. 2000. Haplotype-resolved genomes of geminivirus-resistant and geminivirus-susceptible African cassava cultivars. BMC Biol. 17(1): 75.

Legg, J.P., Lava Kumar, P., Makeshkumar, 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.

Li, W.L., Yu, N.T., Wang, J.H., Li, J.C., Liu, Z.X. 2020. The complete genome of *Banana streak GF virus* Yunnan isolate infecting Cavendish *Musa* AAA group in China. PeerJ. 8: e8459.

Mulenga, R.M., Boykin, L.M., Chikoti, P.C., Sichilima, S., Ng'uni, D., Alabi, O.J. 2018. Cassava brown streak disease and *Ugandan cassava brown streak virus* reported for the first time in Zambia. Plant Dis. 102(7): 1410-1418.

Saitou, N., Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 4(4): 406-425.

Tamura, K., Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10(3): 512-526.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30(12): 2725-2729.

Zanini, A.A., Cuellar, W.J., Celli, M.G., Luque, A.V., Medina, E.D., Conci, V.C., Di Feo, L.V. 2018. Distinct strains of the re-emergent *Cassava common mosaic virus* (genus: *Potexvirus*) infecting cassava in Argentina. 67(8): 1814-1820.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2020 doi:10.20944/preprints202003.0336.v1

159 **Table 1** Primers used for CsCMV-LG amplification in this study.

Fragment	Primer name	Sequences (5'-3')
5' Race	CsCMV 770R	CTCTATAAGACCAACACCTAGCCAGCCC
P1	CsCMV 1F	CAAAACAAAACAGAACAAAACAATCTAACTAACTC
	CsCMV 1760R	GGCTTTCCTCCTTCTGCTCAGC
P2	CsCMV 1700F	CGTCCAGGAGGAAGAAAGAG
	CsCMV 3340R	CGGTCCAGTTGTGTTCCTTATG
P3	CsCMV 3280F	GGGAAACCATTAAGGCCAGGCT
	CsCMV 3892R	CCTTCAAGCACCCATTCAGGGAT
P4	CsCMV 3855F	AGCAAAGCACCACAACATCCC
	CsCMV 5490R	CCTGGAGCAAAGCTTTCGC
P5	CsCMV 5430F	CTCATAGTAAGGGCGGTTGTTTG
	M4T	GTTTTCCCAGTCACGACTTTTTTTTTTTTTTTTT

160

161

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2020 doi:10.20944/preprints202003.0336.v1

Table 2 The sequence identities of viral gene nucleotide and amino acid between CsCMV-Brazilian and CsCMV-LG.

Gene name	Nucleotide (%)	Amino acid (%)
Complete genome	88.0	—
5’ UTR	94.7	—
RDRP	87.40	94.30
TGB1	87.30	94.30
TGB2	92.0	92.80
TGB3	90.8	91.7
CP	88.5	97.8
3’ UTR	92.1	—

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2020 doi:10.20944/preprints202003.0336.v1

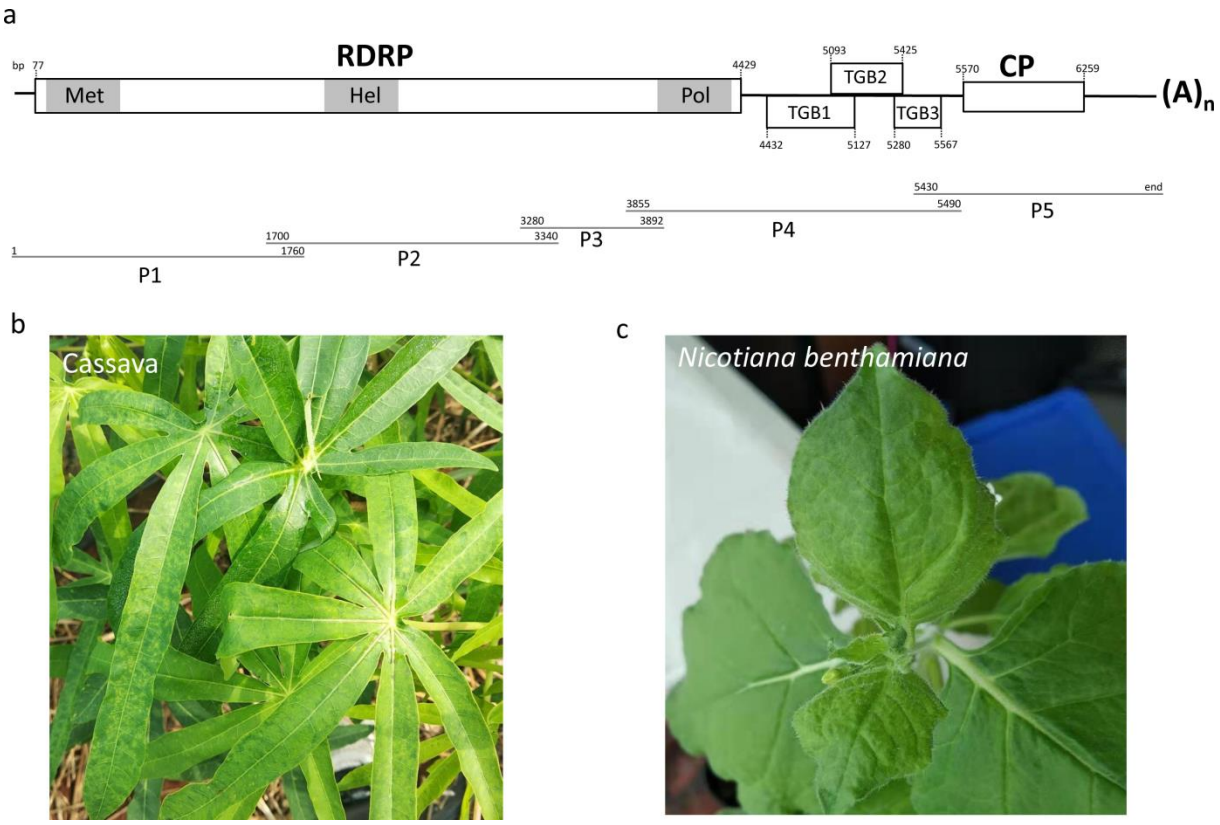

Fig. 1 Schematic representation of the genome organization of CsCMV-LG.
The 5'- and 3'-untranslated regions (UTR) are represented by a solid line. The viral proteins RDRP, TGB1-3 and CP are indicated by boxes in white. The putative domains Met, Hel and Pol with RDRP protein are indicated in gray (**A**). Symptoms of mild mosaic on the diseased cassava plant (**B**). Symptoms of mild vein yellowing on the diseased tobacco plants at 15 day post inoculation (**C**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2020 doi:10.20944/preprints202003.0336.v1

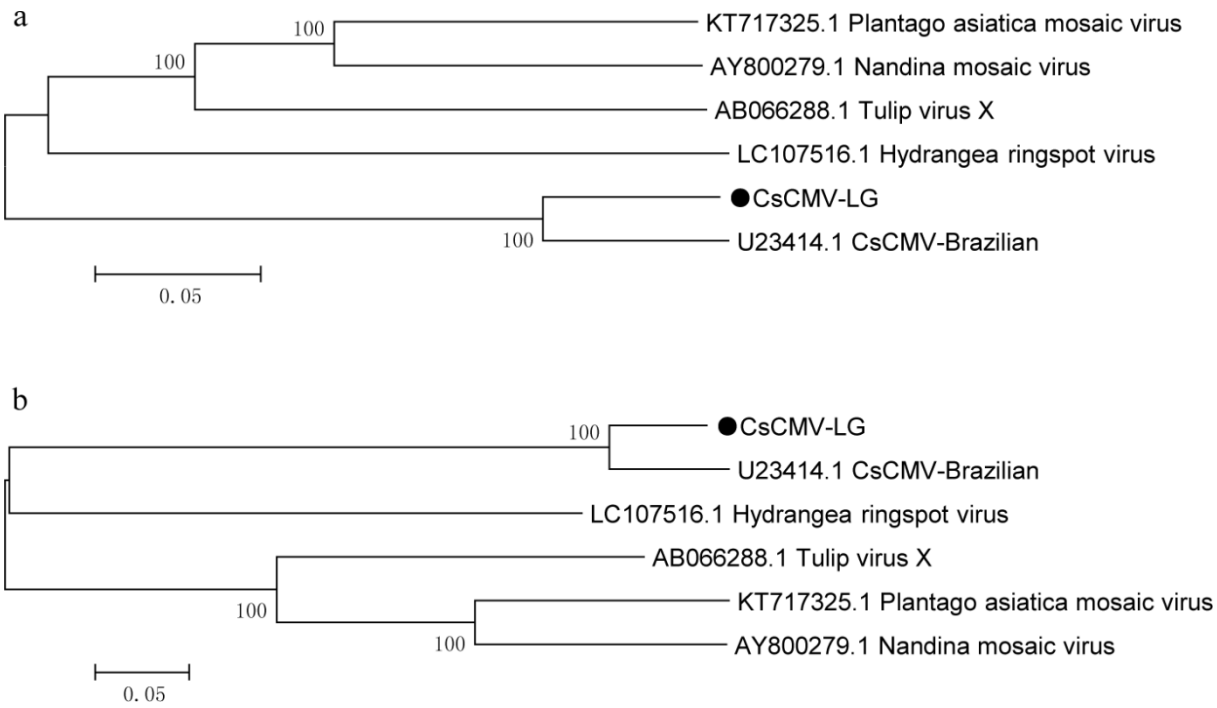

Fig. 2 Phylogenetic analysis of the CsCMVs and other potexviruses based on the complete genome nucleotide sequences.

The Neighbor-Joining method (**A**) and the Maximum Likelihood method (**B**) were used to construct the phylogenetic trees. The bootstrap values (NJ/ML: 1000 replications) of identical branches in each tree constructed by the two methods are shown above and below the branches. All positions containing gaps and missing data were eliminated. Evolutionary analysis was conducted using MEGA6. The following viruses were included in the analysis: CsCMV-LG (MT038420), CsCMV-Brazilian (U23414.1), *Hydrangea ringspot virus* (LC107516.1), *Tulip virus X* (AB066288.1), *Plantago asiatica mosaic virus* (KT717325.1) and *Nandina mosaic virus* (AY800279.1).