ANNOTATED SEQUENCE RECORD

An isolate of sweet potato chlorotic stunt virus from Brazil with a distinct genome organization

Maurício Rossato[1] · Caroline Amaral Souza[1] · Fernando Lucas Melo[1] · Anelise Franco Orilio[2,5] ·
Renato Oliveira Resende[1,2] · Simone Graça Ribeiro[3] · Fernanda Yuri Borges Naito[1] · Gilvan Pio Ribeiro[4] ·
Genira Pereira Andrade[4] · Rita Cássia Pereira-Carvalho[1]

Abstract

Sweet potato chlorotic stunt virus (SPCSV; genus *Crinivirus,* family *Closteroviridae*), is an economically important pathogen of sweet potato. In the present work, the nucleotide sequences of two RNA segments of SPCSV (isolate SPCSV-UNB-01) were determined by MiSeq Illumina sequencing of samples of sweet potato plants grafted onto *Ipomoea setosa*. A comparative analysis of the genome organization of SPCSV-UNB-01 and other SPCSV sequences showed that RNA1 was lacking p22, and p5.1 and that p5.2. was absent in RNA2, indicating a unique genomic pattern. SPCSV-UNB-01 contained longer p6 and p5 regions, with little similarity to orthologous sequences. Sequence comparison did not reveal any previously identified functional domains within these open reading frames (ORFs). No recombination or rearrangement events were detected. Phylogenetic analysis suggested the possibility of separate entries of SPCSV into South America based on the genetic distance between SPCSV-UNB-01 and the Peruvian isolate m2-47. Samples from northeastern Brazil (State of Pernambuco) were positive for SPCSV when tested using specific primers for the major coat protein (CP) gene. This is the first full-length genome sequence of SPCSV-UNB-01 from Brazil.

Sweet potato chlorotic stunt virus (SPCSV) is an important pathogen of sweet potato (*Ipomoea batatas*) and is capable of causing a 52% reduction in yield [1]. In Brazil, detailed information about the impact of SPCSV is limited because symptoms of infection are often confused with those caused by begomoviruses, which share the same vector, *Bemisia tabaci.* SPCSV belongs to the genus *Crinivirus* in the family *Closteroviridae* and has a bipartite, single-stranded

plus-sense RNA genome. This virus has been sequenced and contains two RNA segments, RNA1 and RNA2, which have a variable number of open reading frames (ORF) and share similarities with other members of the genus *Crinivirus* [2]. In RNA1, five ORFs have been reported, including polyprotein 1a, RNA-dependent RNA polymerase (RdRp), RNase3, p7, and p22. In RNA2, 11 ORFs have been described, including p6, p6.1, p5, p5.1, p5.2, heat shock protein 70 homolog, p60, p8, CP (major coat protein), CPm (minor coat protein), and p28 [3].

The lack of a full-length Brazilian genome sequence of this virus prompted this research. In the present study, we describe the complete genome sequence of an SPCSV isolate from Brazil denoted "SPCSV-UNB-01" with GenBank accession numbers MH614269 (RNA1) and MH614270 (RNA2) and compared its genetic organization and diversity to nine other full-length SPCSV genome sequences available in the GenBank database.

Symptomatic sweet potato plants were grafted onto *Ipomoea setosa* and kept in a greenhouse for 21 days, followed by virus enrichment and RNA extraction. Total RNA was extracted from four pooled leaf samples, each with 10 leaves of 10 *I. setosa* plants, using TRIzol reagent, following

---

Handling Editor: Robert H.A. Coutts.

✉ Rita Cássia Pereira-Carvalho
  rcpcarvalho@unb.br

1 Departamento de Fitopatologia, Universidade de Brasília, Brasília, Brazil

2 Departamento de Biologia Celular, Universidade de Brasília, Brasília, Brazil

3 Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brazil

4 Departamento de Fitopatologia, Universidade Federal Rural de Pernambuco, Pernambuco, Brazil

5 Present Address: Departamento de Fitopatologia, Universidade Federal de Viçosa, Minas Gerais, Brazil

the manufacturer's instructions, and sequenced on an Illumina MiSeq platform. The sequencing generated 6,862,788 paired-end reads, which were then processed for quality using CLC Genomic Workbench version 8.0 (QIAGEN Bioinformatics) and assembled in Geneious R8.

A blastN search of the resulting contigs showed that two contigs displayed a high degree of similarity to RNA1 and RNA2 sequences of SPCSV. The RNA1-related contig of 8,473 nt was assembled from 3,466 reads and showed 91% sequence identity with 100% coverage to the isolate SPCSV_Can181-9/AM-MB2. The RNA2-derived contig was 8,016 nt long and was covered by 4,236 reads. It spanned 97% of the SPCSV_Can181-9/AM-MB2 RNA2 sequence, with 86% identity. For identification of this virus, both RNA segments were aligned to 20 sequences of 12 criniviruses using MAFFT. Phylogeny was inferred using MrBayes [4], implemented in Geneious R8 with 1,000,000 generations and 25% burn-in, using the GTR+G+I model. Phylogenetic analysis of both RNA1 and RNA2 of SPCSV-UNB-01 grouped this virus with other SPCSV isolates in a clade with a maximum value of posterior probability (data not shown).

Annotation was performed using the complete genome sequence of SPCSV, retrieved from GenBank, as a reference (accession numbers NC_004124.1 and NC_004123.1). New ORFs were searched using Geneious R8, and annotations were manually checked for quality. The percent identity values for pairs of sequences were extracted from the MAFFT alignment.

Compared to the other nine SPCSV sequences, SPCSV-UNB-01 RNA1 did not show any differences in the number of ORFs (Fig. 1), except for the silencing suppressor p22 [5], which is present only in the Uganda isolate (NC_004123). Four ORFs were identified, including those for polyprotein 1a (5,964 nt), with 89.2% (nt) and 92.9% (aa) identity to the orthologous ORF from the isolate Can181-9/AM-MB2; RdRp (1518 nt), with 95.1% (nt) and 98.2% (aa) identity to the corresponding ORF of isolate Can181-9/AM-MB2; RNase3 (690 nt), with 93.0% identity (nt and aa) to the corresponding ORF of Sichuan-12-8, and p7 (168 nt), showing 92.9% (nt) and 89.1% identity to the corresponding ORF of isolate Can181-9/AM-MB2.

Within RNA2, greater differences were found (Fig. 1). p6 was found to be present in the SPCSV m2-47, Guangdong, and Uganda isolates but absent from the other sequences. Furthermore, the SPCSV-UNB-01 p6 is longer, with 174 nt versus 150 nt in the other SPCSV isolates. The ORF p6 has the highest sequence identity, 75.3% (nt) and 63.2% (aa), to the isolate Guangdong p6, and SPCSV-UNB-01 p5 is also longer, with 162 nt, compared to 135 nt for the other isolates, and has 83.0% (nt), and 47.2% (aa) sequence identity to the isolate Sichuan-12-12. p5 is present in Can181-9, Can181-9/AM-MB2, Sichuan-8-12, Sichuan-12-12, Jiangsu, and Chongqing-12-8 and is absent in the m2-47,

Guangdong, and Uganda isolates. The lack of information on both SPCSV p5 and p6 ORFs makes it difficult to predict whether these alterations may have biological consequences. The additional six ORFs identified in RNA2 include HSP70h, which is 1665 nt in length, and shares 87.7% (nt), and 98.0% (aa) identity with SPCSV isolate Can181-9/AM-MB2; p60, with 1557 nt and 88.1% (nt) and 92.9% (aa) identity to SPCSV-Sichuan-12-12; p8 with 222 nt and 89.2% (nt) and 95.9% (aa) identity to SPCSV isolate Can181-9/AM-MB2; the major coat protein, with 774 nt and 86.6% (nt) and 91.4% (aa) identity to SPCSV isolate Jiangsu; the CPm, with 774 nt and 83.4% and 86.0% (aa) identity to SPCSV isolate Sichuan-12-12; and p28, with 729 nt and 89.2% (nt) and 96.7% (aa) identity to SPCSV isolate Can181-9/AM-MB2.

To further characterize these ORFs, we performed a search for functional domains in p5 and p6 of SPCSV-UNB-01 and other isolates using CDART [6], but no domains were found. Furthermore, because SPCSV-UNB-01 shares genome characteristics with two distinct groups of SPCSV isolates, one with p6 but lacking p5 and the other lacking both p5 and p6, the possibility that recombination had occurred in the genome of SPCSV-UNB-01 was considered. Recombination analysis was performed in RDP4 [7], using all of the available methods, and no recombination events were detected in SPCSV-UNB-01. Phylogenetic trees based on RNA1 and RNA2 were congruent with each other, suggesting that no reassortment events had occurred in the evolutionary history of these lineages. The region between the 5´UTR and ORF HSP70h contains several putative small ORFs, suggesting that *de novo* gene emergence is common in these lineages, but the biological impact of this variation in gene content remains to be determined.

Phylogenetic analysis using only SPCSV isolates showed that SPCSV-UNB-01 is in a different clade than the other South American isolate, m2-47 from Peru (Fig. 1). This raises the possibility that the Brazilian strain entered South America separately.

PCR reactions with the specific primers RNA2_4.359F (ATGGCCGATAGTAACAAAACAG) and RNA2_5.132R (CGATCACGAACCAAAAAGGC), targeting the CP gene (amplicon of 773 bp), confirmed the presence of the virus in one of the 40 sequenced *I. setosa* samples. Furthermore, other greenhouse-kept *I. setosa* (n = 3) and *I. batatas* (n = 4) plants from other unsequenced germplasm from northeastern Brazil (State of Pernambuco) were also positive for SPCSV-UNB-01.

In conclusion, MiSeq sequencing of grafted *I. setosa* plants led to the identification of the first full-length genome of a Brazilian isolate of SPCSV. The genome of SPCSV-UNB-01 differs in the number of ORFs and its amino acid sequence from other reported SPCSV genomes. Recombination and/or rearrangement events
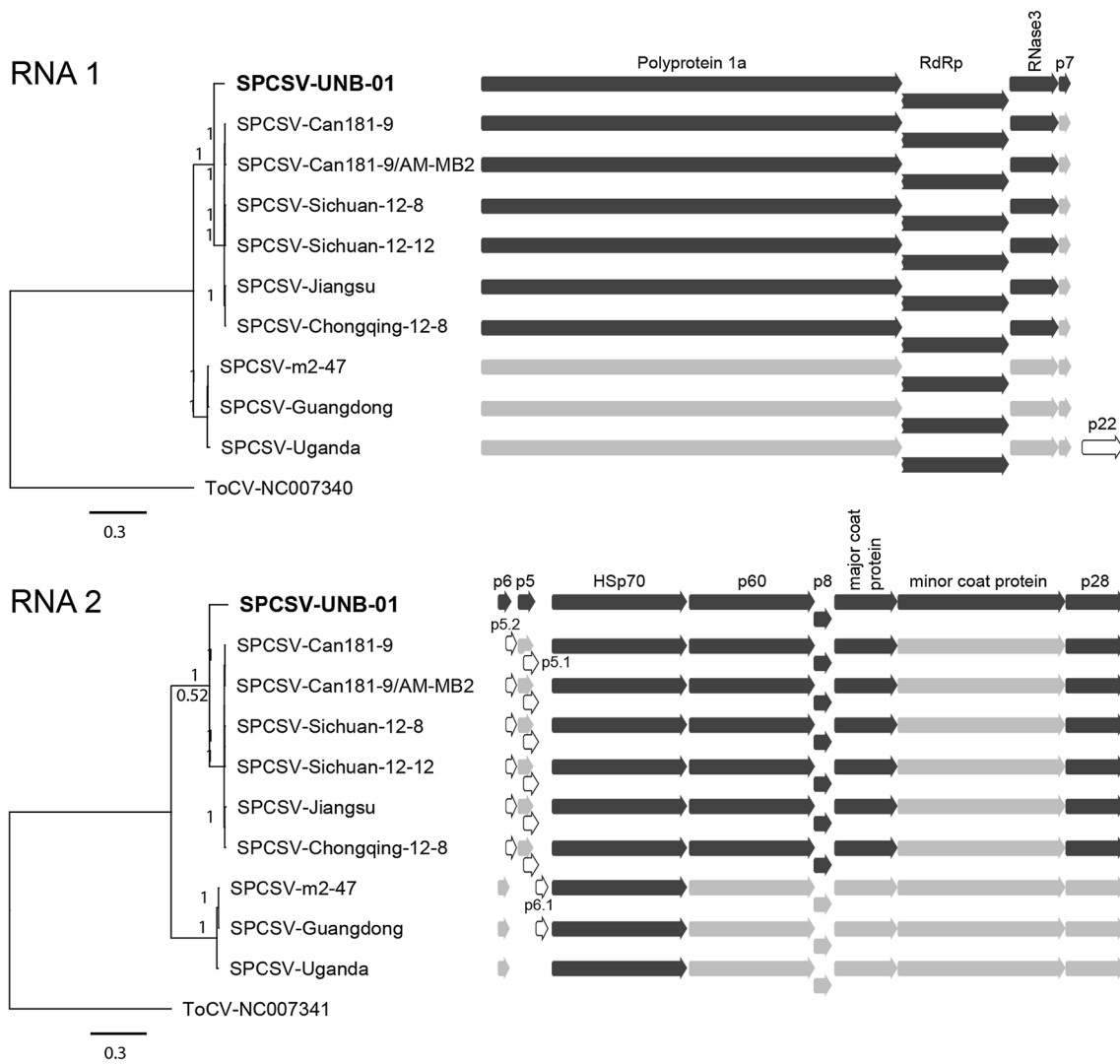
Fig. 1 Phylogenetic trees RNA1 (a) and RNA2 (b) of SPCSV-UNB-01 and 10 other crinivirus isolates, constructed by Bayesian inference in the MrBayes plugin in Geneious R8 with the GTR+G+I model, 1 million generations, and 25% burn-in, followed by the genomic organization of RNA1 and RNA2 of sweet potato chlorotic stunt virus (SPCSV) isolate SPCSV-UNB-01 compared to other nine SPCSV genomes with sequences available in the GenBank database: RNA1 accession numbers:, FJ807784 (Can181-9), KC888964 (Sichuan-12-8), KC888965 (Sichuan-12-12), KC888966 (Chongqing-12-8), KU511273 (Can181-9/AM-MB2), KC146842 (Guangdong) NC004123 (Uganda), HQ291259 (m2-47) and KC146840 (Jiangsu). RNA2 accession numbers: FJ807785 (Can181-9), KC888961 (Sichuan-12-8), KC888962 (Sichuan-12-12), KC888963 (Chongqing-12-8), KU511274 (Can181-9/AM-MB2), KC146843 (Guangdong), NC004124 (Uganda), HQ291260 (m2-47), KC146841 (Jiangsu). Each arrow indicates an ORF. Black arrows indicate >90% amino acid sequence identity to SPCSV-UNB-01, and grey arrows indicate <90% identity. White arrows represent ORFs that are not present in SPCSV-UNB-01.

were not detected and therefore could not explain the origin of this distinct SPCSV isolate.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Adikini S, Mukasa SB, Mwanga RO, Gibson RW (2016) Effects of Sweet potato feathery mottle virus and Sweet potato chlorotic stunt virus on the yield of sweet potato in Uganda. J Phytopathol 164(4):242–254

2. Dolja VV, Kreuze JF, Valkonen JP (2006) Comparative and functional genomics of closteroviruses. Virus Res 117(1):38–51

3. Qin Y, Wang L, Zhang Z, Qiao Q, Zhang D, Tian Y, Wang S, Wang Y, Yan Z (2014) Complete genomic sequence and comparative analysis of the genome segments of Sweet potato chlorotic stunt virus in China. PLoS One 9(8):e106323

4. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17(8):754–755

5. Cañizares MC, Navas-Castillo J, Moriones E (2008) Multiple suppressors of RNA silencing encoded by both genomic RNAs of the crinivirus, Tomato chlorosis virus. Virology 379(1):168–174

6. Geer LY, Domrachev M, Lipman DJ, Bryant SH (2002) CDART: protein homology by domain architecture. Genome Res 12(10):1619–1623

7. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1(1):1–5