Virus Research 243 (2018) 19–24

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Short communication

Molecular characterization of two badnavirus genomes associated with *Canna yellow mottle* disease

Dulanjani Wijayasekara[a], Peter Hoyt[b], Austin Gimondo[c], Bruce Dunn[c], Aastha Thapa[a], Hannah Jones[a], Jeanmarie Verchot[a,*]

[a] Department of Entomology & Plant Pathology, 127 Noble Research Center Oklahoma State University, Stillwater, OK 74078, United States
[b] Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK 74078, United States
[c] Department of Horticulture & Landscape Architecture, Oklahoma State University, Stillwater, OK 74078, United States

ARTICLE INFO

Keywords:
Badnavirus
Potyvirus
Subviral DNAs
Satellite DNA
Plant virus
Cannaceae
Musa virology
Badnavirus
Bacilliform DNA containing plant viruses
Ornamental gingers
Banana streak virus
Canna virus
Caulimoviridae
Tropical plant virus

ABSTRACT

Members of the genus *Badnavirus* have a single non-covalently closed circular double-stranded DNA genome of 7.2–9.2 kb. The genome encodes three open reading frames (ORFs) on the positive DNA strand. *Canna yellow mottle virus* (CaYMV) is a badnavirus that has been described as the etiological cause of yellow mottle disease in canna, although only a 565 bp fragment of the genome has been previously reported from cannas. In this report, concentrated virions were recovered from infected canna plants and nucleic acids were extracted. Two full-length sequences represent two badnavirus genomes were recovered and were determined to be 6966 bp and 7385 bp in length. These DNAs represent a virus strain belonging to *Canna yellow mottle virus* and a novel species tentatively termed Canna yellow mottle associated virus. Phylogenetic analysis indicates that these two viruses are closely related to sugarcane bacilliform GD virus, pineapple bacilliform comosus virus, banana streak MY virus, and cycad leaf necrosis virus. We also showed naturally grown canna plants to be frequently co-infected by these two badnaviruses along with a potyvirus, *Canna yellow streak virus*.

Canna species are native to Central and South America and have been extensively hybridized for floriculture production (Cooke, 2001). Hybrid cannas are traded globally and are subject to serious virus diseases. Accumulation of viruses in vegetatively propagated rhizomes has hindered international trade of germplasm. The three most common viruses reported in canna plants are *Bean yellow mosaic virus* (BYMV), *Canna yellow streak virus* (CaYSV) and *Canna yellow mottle virus* (CaYMV) (Rajakaruna et al., 2013). While researchers have reported the complete genome sequences for several isolates of BYMV and CaYSV, only a 565 bp portion of the CaYMV genome has been described from infected canna plants. This fragment was used to develop diagnostic PCR primers known as CaYMV-3 and CaYMV-4 (Momol et al., 2004). These primers have since been widely employed for diagnostic detection of the virus associated with canna yellow mottle disease. Based on the 565 bp fragment of CaYMV, researchers reported this virus belongs to the *Badnavirus* genus. The genome of a typical member of the genus *Badnavirus*, consists of single circular double stranded DNA molecules ranging from 7.2 to 9.2 kb with three open reading frames (ORF) (Borah et al., 2013; King et al., 2011). The ORF3 encodes a long polyprotein that contains the aspartic acid protease, reverse transcriptase and RNase H domains. Recently, Zhang et al. (2017) reported the complete genome sequence which they proposed to be a variant of CaYMV identified in *Alpinia purpurata* (Vieill.) K. Schum. in Hawaii. The complete genome of canna yellow mottle virus in *A. purpurata* (CaYMV-Ap) was 7, 120 bp (Zhang et al., 2017).

To facilitate virus indexing of vegetatively propagated canna plants, we maintain more than 1000 plants per year, of several varieties, for the past seven years in a greenhouse. We developed a workflow for screening large sample populations to identify and segregate infected and healthy plants. We optimized a method for extracting total nucleic acids from leaves of 8-week old plants and developed a reliable two-step multiplex reverse transcription, polymerase chain reaction (RT-PCR) to simultaneously detect BYMV, CaYSV, and CaYMV. This method employs primer concentrations to ensure the reliable and sensitive detection of as little as 0.1 pg of virus nucleic acids (Chauhan et al., 2015). The results are managed in a database that contains information

Abbreviations: CaYMV, *Canna yellow mottle virus*; CaYSV, *Canna yellow streak virus*; NGS, next generation sequencing; ORFs, open reading frames

* Corresponding author. Present Permanent address: Texas A & M Agrilife Research, Dallas Center, 17360 Coit Road, Dallas, TX 75252, United States.
E-mail address: jm.verchot@ag.tamu.edu (J. Verchot).

http://dx.doi.org/10.1016/j.virusres.2017.10.001
Received 7 August 2017; Received in revised form 29 September 2017; Accepted 2 October 2017
Available online 05 October 2017

about plant varieties, individual plant symptoms monitored over the life time of the plant, and the diagnostic results for every individual. This workflow enabled us to determine which plants and varieties were best to use for this study.

This research was undertaken to recover complete badnavirus genome sequences from greenhouse grown ornamental canna plants. We collected 80–100 g of infected leaves (cultivars 'Red Futurity' and 'Striped Beauty') that tested positive by RT-PCR (using the CaYMV-3 and -4 primer) for CaYMV for virus purification. We employed two published methods involving extraction and differential centrifugation for the recovery of bacilliform virions that was developed for isolation of members of the genus *Caulimovirus* (Covey et al., 1998; Zhuang et al., 2011). DNA was extracted from virions and resuspended in deionized water using conventional procedures.

The pooled DNA preparations were sequenced using a Roche 454-Junior™ Genomic Sequencer. There were 375,342 raw reads with median read lengths of 441 bp were cleaned using the FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) on the iPlant/CyVerse website (https://user.iplantcollaborative.org/) to eliminate primer and adapter sequences, and to identify and restrict sequencing reads to those with phred scores ≥ 30. These were automatically assembled by the Roche Genome Assembly software (v. 2.8) and then reassembled using Newbler v2.7 resulting in 13,253 cleaned contigs. These sequences were used for similarity BLASTn searches in the NCBI GenBank databases using viruses (Taxid: 10239) as a limiting organismal name. While three contigs were identified with high similarity ($e < 1 \times 10-30$) to CaYMV isolate V17 ORF3 gene (EF189148.1), and *Sugarcane bacilliform virus* isolate BataviaD, (FJ439817.1), only one of these, contig0028 (6966 bp) represented a full-length virus genome (Genbank Accession: NC030462.1). We ascribed a provisional name to this genome of canna yellow mottle associated virus 1 (CaYMAV-1).

As a second approach to recovering badnavirus nucleic acids, RNA was extracted from infected (variety "Red Futurity") leaves that tested positive by RT-PCR for CaYMV using the CaYMV-3/4 primer set. Total RNA was isolated using the Roche® SeqCap RNA system. Oligo$(dT)_{12-18}$ primers were used for reverse transcription to prepare cDNA and then random hexamers were used to synthesize the second DNA strand. The dsDNA was ligated to sequencing adapters. Using the Roche 454-Junior™ Genomic Sequencer, we obtained 163,870 reads with an average read length of 359 bp. Sequence reads were assembled into 10,745 contigs with an average length of 687 bp. These contigs were submitted to BLASTn suite and the resulting output was imported to MEGAN6 software (Huson et al., 2016) which automatically assigns output reads using a Lowest Common Ancestor algorithm to their respective taxons. Of the 10,745 contigs, 6249 were assigned to taxons and we identified a subset ascribed to virus taxons. Within this subset of data, we identified and assembled four contigs to obtain a second full-length circular viral genome that was 7385 bp in length. The phylogenetic data presented in the following paragraphs suggest that this virus is a variant of the recently reported CaYMV-Ap. In this report, this genome was provisionally identified as variant 1 of the canna yellow mottle virus in *Alpinia purpurata* (CaYMV-Ap01) (Genbank Accession: MF074075) (Fig. 1A).

To clone the full-length genomes designated as CaYMAV-1 and CaYMV-Ap01, overlapping PCR products were generated using DNA extracted from infected leaves and primers listed in Table S1. Individual PCR products were introduced into pGEM-T Easy vectors (Promega, USA). Several colonies of each transformant were sequenced (Fig. 1B) and the resulting transformant sequences were aligned using SDSC Biology Work Bench and MEGA 6.0 software (Hall, 2013; Sohpal et al., 2010; Subramaniam, 1998). The full-length of the genomes recovered by direct cloning where then aligned with the CaYMAV-1 and CaYMV-Ap01 genomes obtained by *de novo* sequencing (Edgar, 2004).

The genomes of CaYMAV-1 and CaYMV-Ap01 each show an organization that is consistent with species belonging to the genus *Badnavirus*. There are two small ORFs encoding the P1 and P2 proteins. P1 has an unknown function and P2 has nucleic acid binding properties and is reported to be associated with virions. The larger ORF3 encodes a P3 polyprotein (Fig. 1A) that is suggested to contain the movement protein, coat protein, aspartic acid protease, and RT/RNAase H1 domains (King et al., 2011). The CaYMAV-1 ORF1 (nucleotide positions 564–1085) produces a protein of 173 amino acids. The CaYMAV-1 ORF2 (nucleotide position 1085–1486) encodes a protein of 133 amino acids. The translation stop codon for ORF1 and start codon for ORF2 reside in a common TAATGA element within the CaYMAV-1 genome, and a similar combination of ORF1 stop and ORF2 start codons is seen in the genome of banana streak virus from *Musa acuminata Colla* (BSAcYNV). The CaYMAV-1 ORF3 (nucleotide position 1491–6545) encodes a polyprotein of 1682 amino acids (Fig. 1B) (Zhuang et al., 2011). For CaYMV-Ap01, ORF1 (nucleotide positions 544–1060) produces a protein of 172 amino acids and ORF2 (nucleotide position 1059–1463) encodes a protein of 134 amino acids. The translation stop codon for ORF1 and start codon for ORF2 resides in a common TGATGA element (Fig. 1B). The CaYMV-Ap01 ORF3 coding region (nucleotide position 1463–6848) encodes a 1795 amino acid polyprotein.

The numbering of CaYMAV-1 and CaYMV-Ap01 genomes starts with the putative $tRNA^{met}$-binding site located in the intergenic region which is a defining feature of badnavirus genomes (Fig. 1A). The CaYMAV-1 putative $tRNA^{met}$-binding site has the highly conserved 18 nucleotide consensus sequence TGGTATCAGAGCGAGGTT. A similar consensus occurs in the CaYMV-Ap01 intergenic region, with three mismatches TGGTATCAGAGCTGAGTT (the mismatches are underlined). The putative poly-adenylation signal (AATAAA) in the CaYMAV-1 genome is located near nucleotide 6873, and in the CaYMV-Ap01 genome is near position 7291 (Fig. 1A) (Borah et al., 2013; King et al., 2011).

To better confirm the existence of these two virus genomes in infected cannas, RNA was isolated from a subset of seven symptomatic plants. RT-PCR using primers that differentially detect both genomes. We recently reported that PCR detection of viral cDNAs is highly sensitive and can detect as little as 0.1 pg of viral RNA sequences (Chauhan et al., 2015). Since badnaviruses replicate through an RNA intermediate and produce transcripts as templates for translation, there is a greater abundance of viral RNAs than DNAs in infected cells (Periasamy et al., 2006). To test for both CaYMAV-1 and CaYMV-Ap01, we used three sets of PCR primers (Table S1). The primer set named CaYMV-3/4 detects a 565 bp fragment of both CaYMAV-1 and CaYMV-Ap01genomes and all seven samples tested positive for badnavirus infection using these primers (Fig. 1A and C) (Chauhan et al., 2015). The 0036F/R primer pair hybridizes to unique sequences in the CaYMV-Ap01 genome and produce a 1288 bp PCR product. Four samples produced PCR products with the 0036 F/R primer set indicating the presence of CaYMV-Ap01 (Fig. 4A and C and data not shown). The CaYMV-7-F/R primer pair hybridizes to unique sequences in the CaYMAV-1 genome and produces a 978 bp fragment of the CaYMAV-1 genome (Fig. 4A and C and data not shown). Interestingly, two samples tested positive for both CaYMAV-1 and CaYMV-Ap01 (Fig. 1C). These data indicate that both badnaviruses can simultaneously infect canna plants. Controls for PCR cycling included cDNA synthesized from healthy control host plants or plasmids containing 978-bp CaYMAV-1 or 1288-bp CaYMV-Ap01 genome fragments (Fig. 1C).

A maximum likelihood (ML) tree was constructed using the full-length genomes of 39 badnavirus species and included the genomes of CaYMAV-1 and CaYMV-Ap01 to learn about their relatedness to other badnavirus species (Fig. 2 and Table S2) (Hung and Weng, 2016). The GenBank accession numbers for the sequences are in Table S2. The ML tree contains three major clades. Clade 1 contains 19 species including CaYMV-Ap, CaYMV-Ap01 and CaYMAV-1. The genome sequences of CaYMV-Ap and CaYMV-Ap01 share 94% identity which is well above the 80% threshold for demarcation of species, established by the International Committee on Virus Taxonomy (ICTV) (King et al., 2011).