Virus Research 238 (2017) 29–34

Contents lists available at ScienceDirect

Virus Research

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

Short communication

Characterization of badnaviruses infecting *Dioscorea* spp. in the Pacific reveals two putative novel species and the first report of dioscorea bacilliform RT virus 2

Amit Sukal, Dawit Kidanemariam, James Dale, Anthony James, Robert Harding*

Centre for Tropical Crops and Biocommodities, Queensland University of Technology (QUT), 2 George St, Brisbane, 4001, Australia

ARTICLE INFO

Keywords:
Yam
Episomal badnavirus
Rolling circle amplification
Dioscorea
Dioscorea bacilliform ES virus
Dioscorea bacilliform AL virus 2

ABSTRACT

The complete genome sequences of three new badnaviruses associated with yam (*Dioscorea* spp.) originating from Fiji, Papua New Guinea and Samoa were determined following rolling circle amplification of the virus genomes. The full-length genomes consisted of a single molecule of circular double-stranded DNA of 8106 bp for isolate FJ14, 7871 bp for isolate PNG10 and 7426 bp for isolate SAM01. FJ14 and PNG10 contained three open reading frames while SAM01 had an additional open reading frame which partially overlapped the 3′ end of ORF 3. Amino acid sequence analysis of ORF 3 from the three isolates confirmed the presence of conserved motifs typical of other badnaviruses. Phylogenetic analysis revealed the sequences to be closely related to other *Dioscorea*–infecting badnaviruses. FJ14 and PNG10 appear to be new species, which we have tentatively named dioscorea bacilliform ES virus (DBESV) and dioscorea bacilliform AL virus 2 (DBALV2), respectively, while SAM01 represents a Pacific isolate of the recently published dioscorea bacilliform RT virus 2 and is described as dioscorea bacilliform RT virus 2-[4RT] (DBRTV2-[4RT]).

Members of the genus *Badnavirus* (family *Caulimoviridae*) have non-enveloped, bacilliform-shaped virions with an approximate diameter of 30 nm and length ranging from 120 to 150 nm (King et al., 2012). The genome consists of a single molecule of circular, double-stranded DNA of 7.2–9.2 kb, typically encoding three open reading frames (ORFs) all on the (+) strand (Geering, 2014). Replication occurs via reverse transcription of a greater-than-genome length RNA which subsequently serves as template for both the translation of viral proteins and reverse transcription for replication of the genome (King et al., 2012). Badnavirus ORF 1 encodes a small protein with an unknown function, while ORF 2 encodes a protein referred to as VAP (virion-associated protein) which possesses a conserved coiled-coil motif (Stavolone et al., 2001). ORF 3 encodes a large polyprotein that is cleaved into several mature proteins, including a movement protein (MP), coat protein (CP), aspartic protease (AP), reverse transcriptase (RT) and ribonuclease H (RNaseH) (Geering, 2014).

Badnaviruses are transmitted through vegetative propagation, mealybug vectors and in some cases through seed (Bhat et al., 2016). They are serologically and genetically heterogeneous. Further, genomic DNA of several species, such as banana streak viruses (BSV) (Gayral et al., 2008), dracaena mottle virus (DrMV) (Su et al., 2007), fig badnavirus 1 (FBV-1) (Laney et al., 2012), and dioscorea bacilliform virus (DBV) (Seal et al., 2014), is integrated into the host genome, which hinders the development of diagnostic protocols (Kenyon et al., 2008; Seal et al., 2014). This difficulty in diagnosis presents challenges to the safe exchange of germplasm.

Yams (*Dioscorea* spp.) are economically important, annual or perennial tuber-bearing, dioecious, climbing, tropical monocots classified in the family *Dioscoreaceae* (Mignouna et al., 2008). Cultivated yams are ranked as the fourth most important root crop by production after potato, cassava and sweet potato (FAOSTAT, 2014). They provide a staple food source for millions of people in Africa, South America, Asia and the Pacific, and wild yams provide a valuable food source in times of famine. Yam production is highest in West Africa, which accounts for 95% of the world's total production (Mignouna et al., 2008). Although most of the production occurs in the African region, predominated by *Dioscorea rotundata-cayenensis*, yam is of importance in the South Pacific where *D. alata* and *D. esculenta* are the dominant species (Kenyon et al., 2008). Complete genomes of five *Dioscorea*-infecting badnavirus species have been published, namely dioscorea bacilliform alata virus (DBALV), dioscorea bacilliform sansibarensis virus (DBSNV), dioscorea bacilliform rotundata virus 1 (DBRTV1), dioscorea bacilliform rotundata virus 2 (DBRTV2) and dioscorea bacilliform trifida virus (DBTRV), however, none of these reports are from the Pacific region

* Corresponding author.
E-mail address: r.harding@qut.edu.au (R. Harding).

http://dx.doi.org/10.1016/j.virusres.2017.05.027
Received 7 April 2017; Received in revised form 30 May 2017; Accepted 31 May 2017
Available online 04 June 2017

(Briddon et al., 1999; Seal and Muller, 2007; Bömer et al., 2016, Umber et al., 2016).

Yam production and improvement in the Pacific region is hindered by a lack of genetic diversity. Germplasm exchange within the Pacific region and between Pacific and Africa has been difficult due to a lack of reliable virus diagnostic protocols, especially for badnaviruses. To address this problem, a project was initiated in 2014 by the Secretariat of the Pacific Community (SPC), Fiji, to characterize the diversity of badnaviruses infecting yams in the Pacific region.

SPC maintains an *in vitro* collection of yams (278 cultivars) which is comprised of seven different species, namely, *D. alata*, *D. rotundata, D. esculenta, D. bulbifera, D. nummularia, D. transversa* and *D. trifida*. A subset of this collection (50 cultivars including *D. alata* [28], *D. rotundata* [1], *D. esculenta* [15], *D. bulbifera* [2], *D. nummularia* [2], *D. transversa* [1] and *D. trifida* [1]) was initially screened using an immunocapture polymerase chain reaction (IC-PCR) protocol with a general badnavirus polyclonal antiserum (BenL) kindly provided by Prof. Ben Lockhart (University of Minnesota, USA) and the degenerate badnavirus primers BadnaFP/RP (Yang et al., 2003). In extracts from four of the 50 accessions, including two *D. alata* types from Papua New Guinea (PNG) (DA-PNG03 and DA-PNG10), one *D. esculenta* type from Fiji (DE-FJ14) and one *D. rotundata* type from Samoa (DR-SAM01), the expected 579 bp product was amplified. To validate that the amplification was derived from episomal badnavirus DNA and not integrated badnavirus sequences, total nucleic acid (TNA) was extracted from leaf tissue (Kleinow et al., 2009) and subjected to rolling circle amplification (RCA) using the TempliPhi 100 Amplification Kit (GE Healthcare).

Briefly, 1 µl of TNA (adjusted to 500 ng/µl with sterile water) was mixed with 4 µl of the kit sample buffer and the mixture was denatured for 3 min at 95 °C and snap cooled on ice. Reaction buffer (5 µl) was mixed with 0.2 µl of phi29 DNA polymerase, added to each denatured TNA sample mixture, and incubated at 30 °C for 18 h. The reaction mixture was then incubated at 65 °C for 10 min to inactivate the phi29 enzyme. Based on *in-silico* restriction analysis of published full-length DBSNV (GenBank accession no. DQ822073 and DQ822074) sequences, the RCA products were digested with the restriction enzymes *Bam*HI and *Sal*I for which the published DBSNV sequences contained only one or two recognition sites.

Digestion of the RCA-amplified DNA using *Sal*I resulted in a single fragment of approximately 7.5 kb for all four samples, while digestion using *Bam*HI yielded a single fragment of approximately 7.5 kb from sample SAM01, two fragments of approximately 4 and 3 kb from sample FJ14 and three fragments of approximately 3, 2.5 and 2 kb in samples PNG03 and PNG10. The restriction fragments were excised and purified using Freeze 'N Squeeze™ DNA Gel Extraction Spin Columns (Bio-Rad) and subsequently ligated into appropriately cut and dephosphorylated pUC19 and sequenced as described previously (James et al., 2011). BadnaFP/RP primers were used to sequence the RT/RNaseH–coding region.

Pairwise sequence comparison of the 529 bp RT/RNaseH–coding region delimited by the BadnaFP/RP primers of samples PNG03 and PNG10 revealed that they were identical. When the PNG sequences were compared with the sequences from FJ14 and SAM01 there was 64-69% nucleotide similarity between the three groups. When analyzed using BLASTn, FJ14 was found to be most similar to DBALV (accession KX008571) with 73.2% nucleotide similarity, while PNG03/PNG10 was most similar to DBSNV (accession DQ822074) with 71% nucleotide similarity and SAM01 was most similar to DBRTV2 (accession KX008577) with 95% nucleotide similarity.

Since full genome sequences of the three putative badnaviruses were not available in GenBank at the time of the original analysis, three independent full-length clones of FJ14, PNG10 (as a representative of the PNG isolates) and SAM01, generated from the *Sal*I-digested RCA products, were completely sequenced in both directions by primer walking. To confirm the sequences spanning the *Sal*I restriction sites, PCR was carried out using sequence-specific primers flanking this region. Briefly, PCR master mix consisted of 10 µl of 2X GoTaq Green Master Mix (Promega), 10 ρmol of each sequence-specific primer and 1 µl of TNA extract (diluted to 30–50 ng/µl) in a final volume of 20 µl. PCR cycling conditions were as follows: initial denaturation at 94 °C for 2 min followed by 35 cycles of 94 °C for 20 s, 50 °C for 2 min, and 72 °C for 2 min, with a final extension at 72 °C for 10 min. The amplified products were cloned into pGEM-T Easy (Promega) and sequenced as described previously. Complete genome sequences were then assembled using Geneious v9.0.2 (http://www.geneious.com; Kearse et al., 2012).

The assembled full-length genome sequences of FJ14, PNG10 and SAM01 comprised 8106 bp, 7871 bp and 7426 bp, respectively. The intergenic regions (IR) of FJ14 and PNG10 comprised 1280 nt and 1210 nt, respectively, while the IR of SAM01 was considerably smaller at 751 nt. The IR of all isolates contained several conserved nucleic acid motifs previously described for plant dsDNA viruses (Benfey and Chua, 1990; Medberry and Olszewski, 1993). A putative $tRNA^{met}$-binding site was identified in all sequences ($FJ14_{1\text{-}18}$ and $PNG10_{1\text{-}18}$–TGGTATCAGAGCTTGGTT, $SAM01_{1\text{-}18}$–TGGTATCAGAGCTCGGTT; underlined nucleotides are mismatches) with 94% and 89% nucleotide identity to the plant $tRNA^{met}$ consensus sequence (3' ACCAUAGUCUCGGUCCAA 5'), which has been previously described as the priming site for reverse transcription (Medberry et al., 1990). The putative $tRNA^{met}$-binding site was designated as the origin of the circular genome, consistent with the convention currently used for badnaviruses. Transcriptional promoter elements including putative TATA boxes and polyadenylation signals, analogous to the 35S promotor of cauliflower mosaic virus, were also identified in the region upstream of the $tRNA^{met}$-binding site (Table S1).

SnapGene® software (www.snapgene.com; GSL Biotech) was used to predict the presence of putative ORFs on the plus strand of the three full-length sequences. FJ14 and PNG10 were predicted to have three ORFs, while SAM01 was predicted to have four ORFs, with the size and arrangement consistent with other published badnavirus sequences (Fig. 1). For FJ14, PNG10 and SAM01, ORF 1 was predicted to encode a putative protein of 142 amino acids (aa) with a calculated *Mr* of 16.3 kDa, 16.5 kDa and 16.6 kDa, respectively. A conserved domain was identified within the ORF 1 protein (pfam07028: DUF1319) (Finn et al., 2016), which is a c106184 superfamily that appears to be restricted to badnaviruses (Marchler-Bauer et al., 2015). ORF 2 of FJ14, PNG10 and SAM01 was predicted to encode a putative protein of 128 aa (*Mr* = 13.8 kDa), 131 aa (*Mr* = 14.5 kDa) and 121 aa (*Mr* = 13.6 kDa), respectively. No conserved domains were identified in ORF 2 of any of the sequences. ORF3 of FJ14, PNG10 and SAM01 was predicted to encode a putative polyprotein of 2005 aa (*Mr* = 226.6 kDa), 1946 aa (*Mr* = 221.7 kDa) and 1892 aa (*Mr* = 213.8 kDa), respectively, with conserved domains for a movement protein, aspartic protease, reverse transcriptase, ribonuclease H and RNA-binding zinc finger-like domain (CXCX2CX4HX4C) predicted from the amino acid sequences.

In addition to the three typical ORFs found in badnaviruses, isolate SAM01 was predicted to have an additional ORF 4 of 417 nt (position 6656–7072), which partially overlaps ORF 3 and encodes a 138 aa putative protein of calculated *Mr* of 15.5 kDa. The size and genome position of this ORF is similar to a putative small ORF present in several other badnavirus genomes, including piper yellow mottle virus (PYMoV) (Hany et al., 2014), pagoda yellow mosaic associated virus (PYMAV) (Wang et al., 2014), yacon necrotic mottle virus (Lee et al., 2015), grapevine roditis leaf discoloration-associated virus (GRLDaV) (Maliogka et al., 2015) and rubus yellow net virus (Kalishuck et al., 2013). These small ORFs have little (5–20%) aa sequence homology and no conserved domains.

To determine the taxonomic position of FJ14, PNG10 and SAM01, phylogenetic analysis and sequence comparison to published yam badnavirus sequences was carried out using the 529 bp RT/RNaseH–coding region delimited by the BadnaFP/RP primer binding