ANNOTATED SEQUENCE RECORD

Complete genome sequence of a previously undescribed badnavirus occurring in *Polyscias fruticosa* L. (Ming aralia)

R. A. Alvarez-Quinto[1] · B. E. L. Lockhart[1] · N. Olszewski[2]

Abstract

A previously undescribed badnavirus was identified in plants of *Polyscias fruticosa* (Ming aralia) showing symptoms of mild mosaic and leaf senescence. Characteristic bacilliform virions of the *Polyscias* badnavirus averaging 30×120 nm in size were observed by transmission electron microscopy in partially purified leaf tissue extracts from symptomatic but not asymptomatic plants collected in the USA and Nigeria. The isolate from the USA was complete sequenced. The genome is 7592 bp in length and contains three open reading frames with an arrangement similar to that of other members of the genus *Badnavirus*. The largest open reading frame (ORF3) encodes a putative polyprotein, with predicted domains including zinc finger, aspartic protease, reverse transcriptase (RT) and RNase H, in that order. The USA and Nigeria isolates of the virus had a high level (98%) of nucleotide sequence identity in the RT+RNase H region. Within the genus *Badnavirus*, these viruses were most closely related to schefflera ringspot virus (SRV), sharing 63% identity at the nucleotide level. Based on the ICTV species demarcation criteria for the genus *Badnavirus* (more than 20% nucleotide sequence divergence in the RT+RNase H region), the *Polyscias* virus is proposed to be a new member of the genus, and the name polyscias mosaic virus (PoMV) is proposed. The complete genome sequence was deposited in the NCBI GenBank database under accession no. MH475918.

Introduction

Ming aralia (*Polyscias fruticosa* L. Harms) is a perennial shrub (also considered a dwarf tree) native to tropical Asia that is commonly used as either an indoor or landscape ornamental. *P. fruticosa* belongs to the family Araliaceae, which includes several important foliage crops, including aralias and scheffleras [1]. In early studies, the presence of bacilliform viral particles was reported to occur widely and frequently among several members of this family, including schleffleras (*Schefflera actinophylla* and *S. arboricola*)

✉ R. A. Alvarez-Quinto
alvar419@umn.edu

1 Department of Plant Pathology, University of Minnesota, 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA

2 Department of Plant Biology, University of Minnesota, St. Paul, MN 55108, USA

and aralias (*Polyscias balfouriana*, *P. balfouriana* 'marginata', *P. fruticosa and P. guilfoylei*) [2]. In the same study, a mealybug-transmitted badnavirus, schefflera ringspot virus (SRV), was characterized and partially sequenced from schefflera plants. Bacilliform viral particles observed in the other members of the family Araliaceae were suggested to be isolates of SRV. However, PCR-amplified fragments of the isolates from schefflera and aralia did not cross-hybridize, indicating genetic heterogeneity among these isolates [2]. In 2017, virus-like symptoms were observed in *P. fruticosa* plants, viral particles were characterized, and a genomic sequence was obtained. This communication reports the complete sequence and genome organization of a new bacilliform virus from *P. fruticosa* and its relationship to other members of the genus *Badnavirus*.

Virus source and isolation

Ming aralia plants (*P. fruticosa*) showing virus-like symptoms were obtained from nurseries in the USA (Florida and California) and landscape plants in Nigeria. Leaf samples showed a mild mosaic differing from symptoms previously reported in this species (Fig. 1A). Virus particles were

Fig. 1 A. Leaf showing common symptoms associated with the presence of PoMV. **B.** Leaf from an asymptomatic plant. **C.** Viral particles observed using TEM. **D.** Graphic representation of the genome of polyscias mosaic virus (PoMV). Orange boxes indicate the predicted open reading frames (ORFs), and conserved domains are indicated by overlaying blue boxes
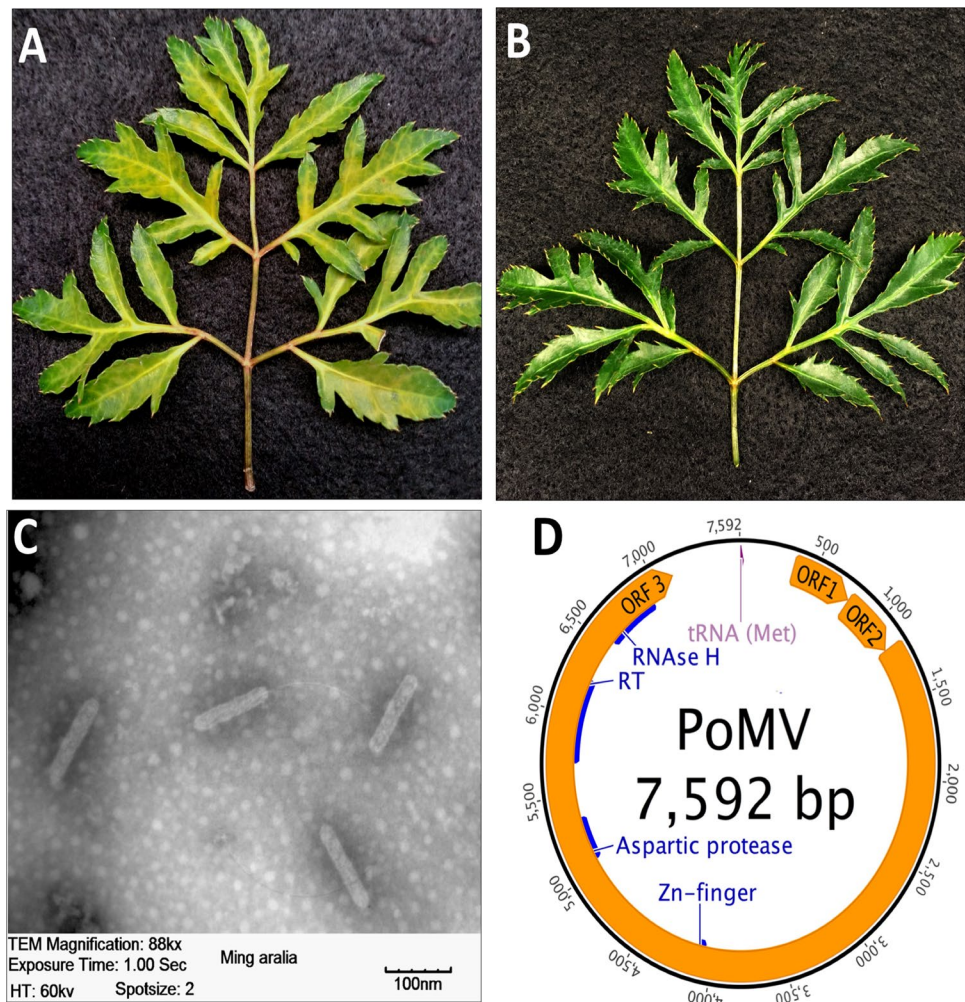
purified from symptomatic tissue as described previously for piper yellow mottle virus (PYMoV) and commelina yellow mottle virus (CoYMV) [3, 4]. Purified suspensions were examined by transmission electron microscopy (TEM) using samples mounted on carbon-coated formvar grids negatively stained with 2% phosphotungstic acid, pH 7.0 (PTA). Bacilliform virions characteristic of badnaviruses were observed by TEM in negatively stained purified leaf extracts from symptomatic but not asymptomatic plants (Fig. 1B). The particle size was, on average, 30 nm width and 120 in length, determined by measuring 100 individual particles (Fig. 1C).

Genome sequencing

Preliminary sequence analysis using the degenerate primers Badna T+2 revealed low sequence similarity in the highly conserved RT+RNase region to previously reported badnaviruses, including SRV. However, a high level of similarity was detected among samples form the USA and Nigeria

(98% identity within the RT-RNase H region). The complete genomic sequence was obtained using a sample from the USA. Briefly, genomic DNA was extracted from purified viral suspensions as described previously [3, 4]. Subsequently, the virus suspension was incubated with DNase I (4 units) and RNase A (25 g/mL) at 37 °C for 30 min, followed by digestion with proteinase K (200 μg/mL) in the presence of 1% SDS at 50 °C for 30 min. The mixture was subjected to phenol:chloroform extraction, and the nucleic acid was precipitated using ethanol. DNA extracted from virions was used as a template for genome sequencing. Initially, the RT+RNase H region was amplified and sequenced with the degenerate primer pair BADNA-T (5'-MYM WNG CTC TGA TAC CA-3') and BADNA-2 (5'-TAY ATH GAY GAY ATH YT-3') [5]. The RT+RNase H sequence was used to generate a pair of outward-facing specific primers (AR2-L: 5'-CTC CTT GTG GAC TTG TCT TTG CAT AGA GG-3' and AR2-R: 5'-GGA TGA ATG ATC AAG ACT GGA AGA TAG TGC-3') designed to amplify the complete circular genome using Long Amp polymerase (NEB). PCR products were cloned using the pGEM-T Easy Vector System (Promega). The complete sequence was obtained from three

clones by Sanger sequencing and primer walking. Sequences were assembled with Geneious R11 [6].

Sequence analysis and genome organization

The viral genome consists of a circular DNA molecule with a length of 7,592 bp and 39% G+C content. The sequence was deposited in the NCBI GenBank database under accession no. MH475918. The tRNA methylation site (5'-TGG TAT CAG AGC ATG GTT-3'), a highly conserved motif required for replication, was identified and set as the starting point of the sequence. Open reading frames (ORFs) were predicted using ORFfinder (NCBI) and Geneious R11 [6]. The genome contains three overlapping ORFs organized in a manner that is typical for badnaviruses (Fig. 1C), with overlapping stop/start codons between ORF1 and ORF2 (TTGA) and between ORF2 and ORF3 (ATGA). ORF1 (position 353 to 772 nt) encodes a putative protein of approximately 16.20 kDa. A domain of unknown function (DUF1319) restricted to members of the genus *Badnavirus* was identified, and the highest identity in ORF1 is 49% at the amino acid level with cacao mild mosaic virus (NC_033738).
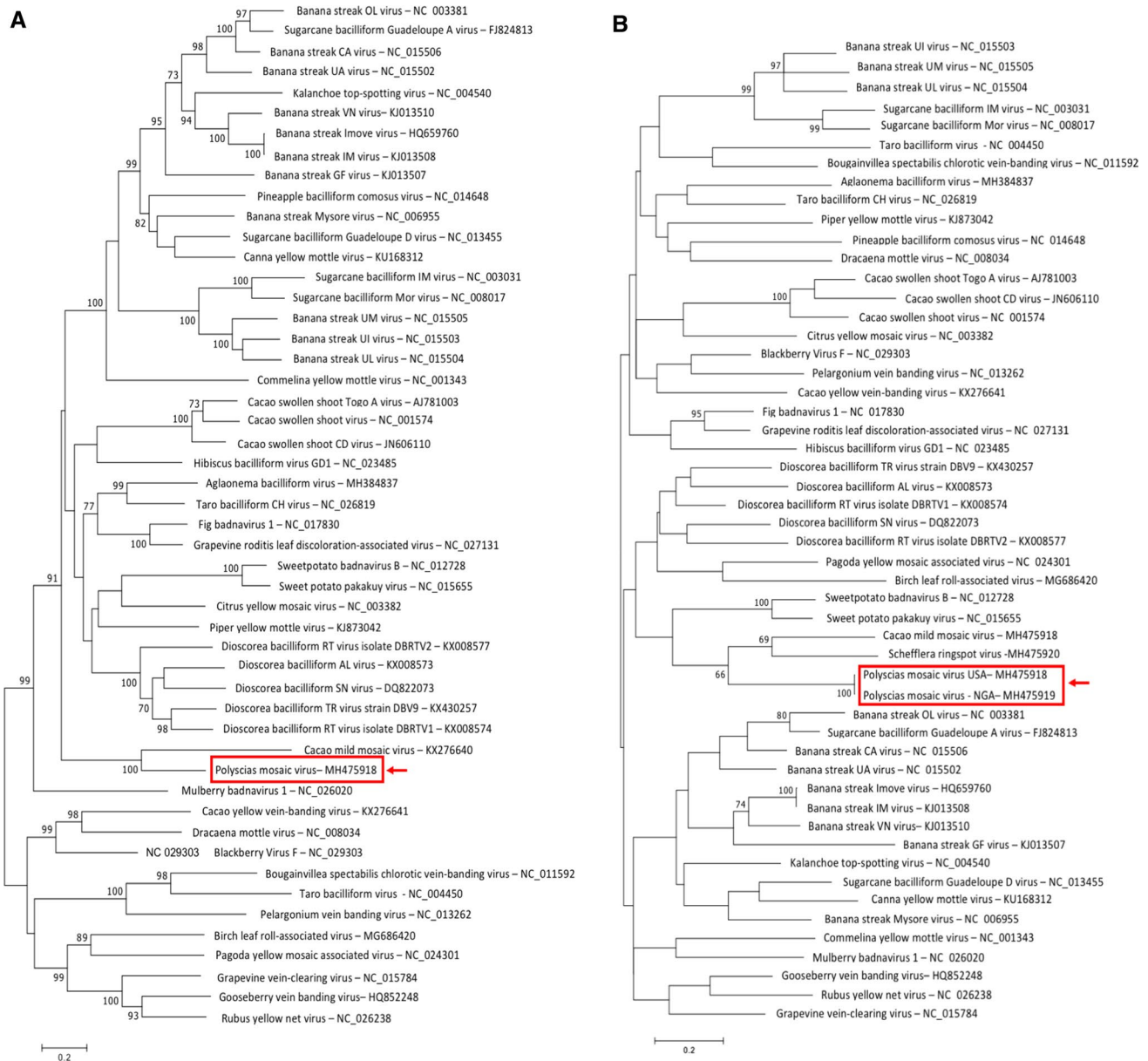

Fig. 2 Maximum-likelihood phylogenetic trees of members of the genus *Badnavirus*. **A.** Tree based on the complete genome sequence. **B.** Tree based on the RT+RNase H region. Boxes and arrows indicate the position of PoMV. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Bootstrap values (1000 replicates) are shown at each node

ORF2 (nt position 769 to 1,164) is predicted to encode a 14.19-kDa protein showing highest similarity (34% identity at the amino acid level) to its counterpart in cacao mild mosaic virus. ORF3 (positions 1,161 to 7,136) encodes a putative polyprotein of 228.46 kDa, with conserved domains identified including zinc finger, aspartic protease, reverse transcriptase (RT) and RNase H. The highest level of identity in ORF3 is 45% at the amino acid level with citrus yellow mosaic virus (EU708317). Pairwise alignments using only the RT+RNase H region showed the highest identity (63%) at the nucleotide level with schefflera ringspot virus (SRV). Nucleotide sequence divergence higher than 20% within the RT+RNAse H region is one of the criteria for species demarcation in the genus *Badnavirus*. Based on RT+RNase H sequence alignments between the *P. fruticosa* badnavirus and other members of the genus, this new virus should belong to a new species in the genus *Badnavirus*, and we propose the name "polyscias mosaic virus" (PoMV) for this virus. Analysis of the partial sequence obtained from the Ming aralia sample from Nigeria revealed a high level of identity to PoMV-USA (98%) at the nucleotide level, and therefore it was considered an isolate of PoMV (PoMV-NGA). The partial sequence of PoMV-NGA has been submitted to the GenBank database (MH475919).

Phylogenetic analysis

Phylogenetic relationships between PoMV and members of the genus *Badnavirus* were assessed using genetic information from 50 members of this genus. Trees were constructed using the complete genome sequences (Fig. 2A) and the RT+RNAse H region (Fig. 2B). Since the complete sequence of SRV is not available, it was not included in the complete genome analysis. Alignment of complete genome sequences was done using the MATFF algorithm in the Galaxy Platform [7], and maximum-likelihood (ML) phylogenetic analysis was carried out using the software MEGA 8.0 [8]. ML trees were inferred based on the general time-reversible model and gamma distribution with invariable sites [9]. The ML tree inferred from the complete genome sequences (Fig. 2A) locates PoMV in a single sister clade with cacao mild mosaic virus (KX276640), a recently reported badnavirus isolated from cacao. The ML tree inferred from the highly conserved RT+RNAse H region (Fig. 2B) resulted in a tree with low overall bootstrap support; however, it clusters both isolates of PoMV in a clade with schefflera ringspot virus (SRV), cacao mild mosaic virus, sweet potato badnavirus B (NC_012728), and sweet potato pakakuy virus (NC_015655).

SRV was first described and partially characterized from *Schefflera actinophylla* (sin. *Brassaia actinophylla*) [2]. In the same study, bacilliform viral particles were found infecting different members of the genus *Polyscias* (*P. balfouriana*, *P. marginata* and *P. fruticosa).* However, a genomic DNA-DNA hybridization assay revealed heterogeneity within the RT+RNase H region. The molecular characterization of PoMV supports previous observations about badnavirus diversity in the botanic family Araliaceae.

Compliance with ethical standards

References

1. Plunkett GM, Lowry PP, Burke MK (2001) The phylogenetic status of polyscias (Araliaceae) based on nuclear its sequence data. Ann Missouri Bot Gard 88:213–230. https://doi.org/10.2307/2666225
2. Lockhart BEL, Olszewski NE (1996) Schefflera ringspot virus, a widely distributed mealybug-transmitted badnavirus occurring in *Shcefflera* and *Aralia*. Acta Hortic. International Society for Horticultural Science (ISHS), Leuven, pp 196–203
3. Lockhart BEL (1990) Evidence for a double-stranded circular DNA genome in a second group of plant viruses. Phytopathology 80:127–131
4. Lockhart BEL, Kiratiya-Angul K, Jones P et al (1997) Identification of Piper yellow mottle virus, a mealybug-transmitted badnavirus infecting *Piper* spp. in Southeast Asia. Eur J Plant Pathol 103:303–311. https://doi.org/10.1023/A:1008699414536
5. Lockhart BEL, Olszewski N (1993) Serological and genomic heterogeneity of banana streak badnavirus: implications for virus detection in Musa. In: Breed Banan Plantain Resist to Dis Pests, Ganry, J. CIRAD-INIBAP, Montpellier, pp 102–113
6. Kearse M, Moir R, Wilson A et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199
7. Afgan E, Baker D, Batut B et al (2018) The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544
8. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7. 0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054
9. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University, Oxford