



# Genome characterization and complete sequence of a new badnavirus from *Pandanus amaryllifolius*

Robert A. Alvarez-Quinto<sup>1</sup> · Samuel Grinstead<sup>2</sup> · Philippe Rott<sup>3,4</sup> · Dimitre Mollov<sup>5</sup>

Received: 1 March 2022 / Accepted: 10 April 2022 / Published online: 24 May 2022

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022

## Abstract

A new badnavirus was sequenced from fragrant pandan grass (*Pandanus amaryllifolius*) displaying mosaic and chlorosis on the leaves. The complete genome sequence was determined by high-throughput sequencing. The new badnavirus was tentatively named "pandan mosaic associated virus" (PMAV). Similar to those of other members of the genus *Badnavirus*, the genome of PMAV consists of a circular DNA molecule of 7,481 bp with three open reading frames (ORF) potentially coding for three proteins. ORF3 encodes a polyprotein with conserved protein domains including zinc finger, trimeric dUTPase, aspartic protease, reverse transcriptase (RT), and RNase H domains. Pairwise comparisons of the highly conserved RT + RNase H region revealed the highest nucleotide (nt) sequence identity (70.71%) to taro bacilliform CH virus-Et17 (MG017324). In addition to PMAV, viral sequences corresponding to orchid fleck dichorhavirus (OFV) were detected in the same plant sample. The complete sequence of the OFV coding region shared >98% nt sequence identity with other isolates of OFV available in the GenBank database. Disease symptoms could not be attributed exclusively to PMAV or OFV, as both viruses were present in the pandan grass exhibiting mosaic and chlorosis.

Pandan (*Pandanus amaryllifolius* Roxb) (family Pandanaceae) is a tropical monocot plant that is widely cultivated in Southeast Asia. Pandan leaves are used for flavoring and as a natural green colorant for several foods. A young pandan plant has a grass-like appearance, but it grows into a palm-like tree with abundant aerial roots when it becomes mature. Pandan is vegetatively propagated using suckers and cuttings [1, 2]. To date, no reports are available regarding viral diseases occurring in *P. amaryllifolius*, but cucumber mosaic virus has been found infecting the related species *P. tectorius* in China [3].

In November 2019, virus-like symptoms, including mosaic and chlorosis, were observed on several offshoots of fragrant pandan grass at Mounts Botanical Garden in West Palm Beach, FL (Fig. 1A). Symptomatic leaves from these plants were collected in April 2021 and prepared for RNA sequencing (RNA-Seq) for further investigation.

Total RNA was extracted using an RNeasy Plant Mini Kit (QIAGEN) according to the manufacturer's instructions. RNA was treated with DNase to remove host genomic DNA, and ribosomal RNA was depleted. Complementary DNA was generated using a Truseq RNA Library Prep Kit and sequenced using a NextSeq 500 Illumina platform (75-bp single-end reads). A total of ~27 million raw single-ended reads were obtained, and reads were processed using CLC Genomics Workbench 11 (QIAGEN). *De novo* assembly generated a total of 61,968 contigs, and viral contigs were identified by BLAST analysis. Three viral contigs sharing low amino acid (aa) sequence identity (<60%) with members of the genus *Badnavirus* were identified. Badnavirus contigs comprising a partial genome sequence were investigated further. Additionally, three contigs showed  $\geq 98\%$  aa sequence identity to homologous proteins of orchid fleck virus (OFV). Subsequent analysis generated a nearly complete genomic OFV sequence, including the complete open reading frames of OFV RNA1 and RNA2. The RNA1 and RNA2 fragments

Handling Editor: Elvira Fiallo-Olivé.

✉ Dimitre Mollov  
dimitre.mollov@usda.gov

<sup>1</sup> Department of Botany and Plant Pathology, 4575 Research Way Lab, Corvallis, OR 97333, USA

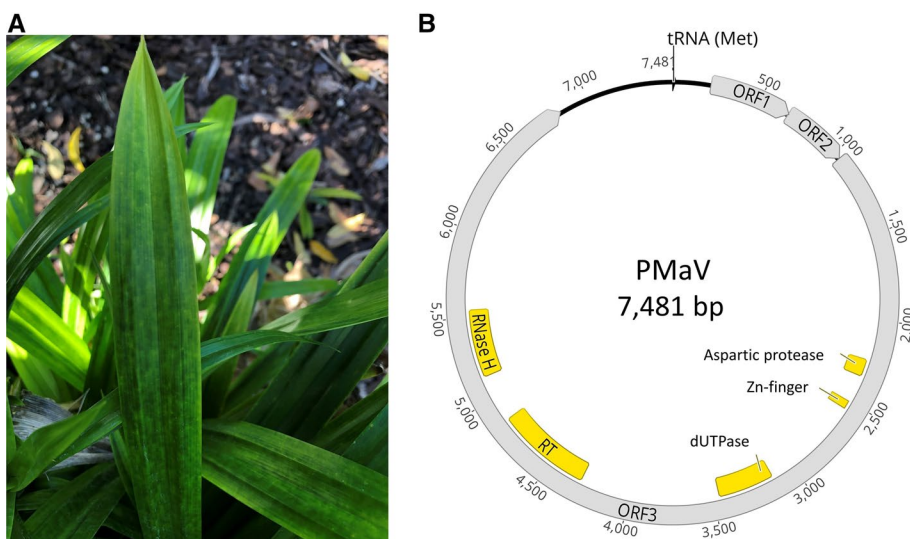
<sup>2</sup> USDA-ARS National Germplasm Resources Laboratory, USDA-ARS, Beltsville, MD, USA

<sup>3</sup> CIRAD, UMR PHIM, 34398 Montpellier, France

<sup>4</sup> PHIM Plant Health Institute, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France

<sup>5</sup> USDA-ARS Horticultural Crops Disease and Pest Management Research Unit, Corvallis, OR 97330, USA

**Fig. 1** (A) Pandan leaves exhibiting symptoms of mosaic and chlorosis. (B) Diagram depicting the genome organization of pandanus mosaic associated virus. ORFs are represented by grey boxes, and putative conserved domains are represented by yellow boxes.



were 6,380 and 5,977 nucleotides (nt) long, respectively. A total of 4,795 raw data reads mapped to RNA1 (coverage of 57X per nt), and 20,656 reads mapped to RNA2 (coverage of 262X per nt). OFV is a bisegmented negative-sense RNA virus of the genus *Dichorhavirus*. The nearly complete genomic sequence including all of the ORFs of the pandan isolate of OFV was submitted to the GenBank database under the accession numbers OK624601 and OK624602 for RNA-1 and RNA-2, respectively. This is the first report of OFV infecting *Pandanus* sp. No other viral contigs were identified.

DNA sequencing (DNA-Seq) was used to generate data for the new badnavirus genome assembly.

Total DNA from symptomatic pandan leaves was extracted using a DNeasy Plant Kit (QIAGEN) and used for sequencing on a NovaSeq Illumina platform. Sequencing generated approximately 6.75M paired reads, which were assembled into 344,342 contigs by *de novo* assembly as described previously. A BLAST search resulted in identification of nine contigs with similarity to sequences from members of the genus *Badnavirus*. Further assembly resulted in a circular DNA sequence of 7,481 bp with 39.8% GC content (Fig. 1B) and an average assembly coverage of 37X per nt position. The full genome sequence obtained from the DNA-Seq analysis completely matched the contigs obtained from the RNA-Seq experiment. The new pandan badnavirus was therefore detected by both RNA- and DNA-based HTS analysis.

The new badnavirus was tentatively named "pandan mosaic associated virus" (PMaV), and the complete genome sequence was deposited in the NCBI GenBank database under accession number OK624603.

Sequence analysis revealed a motif complementary to the 3' end of methionine tRNA (5'- TGG TAT CAG AGC GAG GTT-3'), and this sequence was set as the starting

point of the genomic sequence. Three putative open reading frames (ORFs) were predicted using ORFfinder (NCBI). The sequence of each putative protein was searched for conserved protein domains using the CD-Search tool (NCBI). ORF 1 (nt position 213–674) codes for a putative protein of 17.7 kDa containing a domain of unknown function (DUF1319), which has been found only in members of the genus *Badnavirus*. Pairwise alignments showed that the ORF1-encoded protein shared the highest aa sequence identity (44.2%) with taro bacilliform virus-Tz17 (AWK49020). ORF2 (nt position 671–1,039) encodes a putative protein of 13.6 kDa of unknown function. The ORF2-encoded protein shared the highest aa sequence identity (45.3%) with taro bacilliform virus-Ke52 (AWK49017). ORF3 (nt position 1,039–6,837) codes for a putative polyprotein of 220.4 kDa with conserved protein domains including zinc finger, trimeric dUTPase, aspartic protease (AP), reverse transcriptase (RT), and RNase H domains. The most similar ORF3-encoded polyprotein sequences were those of taro bacilliform CH virus (MG017324) and taro bacilliform virus-Aus7 (MG017318), with 59% and 48% identity, respectively. Pairwise alignment of publicly available sequences of the highly conserved RT + RNase H region (approximately 1 kb) from members of the genus *Badnavirus* showed the highest nt sequence identity (70.7% and 69.4%) to taro bacilliform CH virus-Et17 (MG017324) and jujube mosaic-associated virus-BJ (MN274946), respectively. According to the current ICTV species demarcation criteria for members of the genus *Badnavirus* (nt differences greater than 20% in the RT + RNase H region) [4], PMaV is proposed to be a new badnavirus.

The phylogenetic relationship of PMaV to other members of the genus *Badnavirus* was examined using genetic information for 43 badnaviruses corresponding to both ICTV-approved and tentative members of the genus available in the GenBank database [4, 5]. Multiple sequence alignments



◀ **Fig. 2** Phylogenetic tree constructed using RT-RH1 nucleotide sequences of pandanus mosaic associated virus (PMAV) and 43 members of the genus *Badnavirus*. The tree was constructed by the maximum-likelihood method, using the general time-reversible model. Bootstrap values (1000 replicates) are indicated at each node. The tree is drawn to scale, with branch lengths corresponding to the number of substitutions per site. PMAV is indicated by a star, and rice tungro bacilliform virus (genus *Tungrovirus*) was used as an outgroup.

were performed using Clustal W. Maximum-likelihood (ML) phylogeny was inferred with MEGA 11, and the general time-reversible model [6, 7] was chosen as the best model for our dataset. A phylogenetic tree was constructed based on the nucleotide sequences of the RT-RH1 domain (Fig. 2). In the ML-inferred phylogram, PMAV and taro bacilliform virus (TaBV) were located in sister taxa.

Because members of two different virus species were detected in the pandan sample analyzed in this study, it remains unclear whether the foliar virus-like symptoms observed (Fig. 1A) were associated with PMAV, OFV, or both. Further studies are needed to determine the role of each virus, both as single and mixed infections, in the observed symptomatology.

**Acknowledgements** We thank Joel Crippen and Joshua Spall at Mounts Botanical Garden (West Palm Beach, FL) for their help in sampling fragrant pandan grass (*Pandanus amaryllifolius*). Robert A. Alvarez-Quinto is supported by the Republic of Ecuador through the Ecuadorean Science and Technology Secretariat (SENESCYT).

**Author contributions** All authors have read and agreed to the published version of the manuscript. Conceptualization: Dimitre Mollov, Philippe Rott. Methodology: Robert Alvarez, Samuel C. Grinstead, Dimitre Mollov. Formal analysis and investigation: Robert Alvarez, Samuel C. Grinstead. Writing – original draft preparation: Robert Alvarez, Dimitre Mollov. Writing – review and editing: Robert Alvarez, Samuel C. Grinstead, Philippe Rott, Dimitre Mollov. Funding acquisition: Dimitre Mollov. Resources: Dimitre Mollov, Philippe Rott. Supervision: Dimitre Mollov

**Funding** The authors have not disclosed any funding.

**Declarations**

**Conflict of interest** All authors declare no conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

**References**

1. Wakte KV, Zanan RL, Saini A et al (2012) Genetic diversity assessment in *Pandanus amaryllifolius* Roxb. populations of India. *Genet Resour Crop Evol* 59:1583–1595. <https://doi.org/10.1007/s10722-012-9882-y>

2. Liew P (2013) Pandanus amaryllifolius: the only Pandanus with fragrant leaves. Tropical Biodiversity - University of Reading Weblogs. <https://blogs.reading.ac.uk/tropical-biodiversity/2013/01/pandanus-amaryllifolius/>. Accessed 11 January 2021
3. Kaili W, Jie Z, Kuo W et al (2019) Identification of Cucumber mosaic virus infecting *Pandanus tectorius*. Acta Horti Sin 46:2055–2060. <https://doi.org/10.16420/j.issn.0513-353x.2018-1003>
4. Teycheney P-Y, Geering ADW, Dasgupta I et al (2020) ICTV virus taxonomy profile: Caulimoviridae. J Gen Virol 101:1025–1026. <https://doi.org/10.1099/jgv.0.001497>
5. Geering ADW (2021) Badnaviruses (Caulimoviridae). In: Bamford DH, Zuckerman M (eds) Encyclopedia of virology. Academic Press, Elsevier, Oxford, pp 158–168
6. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York
7. Tamura K, Stecher G, Kumar S (2021) MEGA 11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. <https://doi.org/10.1093/molbev/msab120>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.