



Complete genome sequence of a new member of the genus *Badnavirus* from red pitaya (*Hylocereus polyrhizus*)

Qianming Zheng¹ · Xiaohe Wang¹ · Junliang Zhou¹ · Yuhua Ma¹

Received: 1 September 2019 / Accepted: 25 November 2019 / Published online: 7 February 2020
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

Here, we report a circular double-stranded DNA virus from red pitaya (*Hylocereus polyrhizus*). The complete genome sequence is 7,837 nt in length and shares 98.7% nucleotide sequence identity with epiphyllum mottle-associated virus (EpMoaV) and 40.4–54.6% with other members of the genus *Badnavirus*. It has four open reading frames (ORFs), encoding putative proteins of 19.9, 14.8, 225.7 and 14.2 kDa, respectively. The reverse transcriptase (RT)-ribonuclease H (RNase H) region exhibits less than 70.5% nucleotide sequence identity to RT-RNase H of other badnaviruses, and 99.7% to that of EpMoaV. Phylogenetic analysis revealed that the virus from this study and EpMoaV form a single group. Consequently, we propose this virus as a new member of the genus *Badnavirus* in the family *Caulimoviridae* and have named it “pitaya badnavirus 1” (PiBV1). PiBV1 and EpMoaV should be considered two isolates of a badnavirus that infects members of the family *Cactaceae*.

Pitaya is a perennial tropical fruit crop of the genus *Hylocereus* in the family *Cactaceae*, which includes red pitaya (*H. polyrhizus*) and white pitaya (*H. undatus*). Pitaya fruits are rich in numerous nutrients such as vitamin C, polysaccharides, flavonoid, polyphenol and betalain, which are beneficial to human health [1, 2]. In the past years, as an economically important fruit crop, pitaya has been widely cultivated in southern and southwestern China, where its planting area has reached 40,000 hectares.

Usually, pitaya plants are vegetatively propagated using mature stem tissues, which results in the accumulation of multiple viruses. A series of virus-like symptoms, such as mottling, yellowing and chlorosis, have been observed previously on the surface of stems of pitaya plants [3]. In the past decades, a number of viruses belonging to the genus *Potexvirus* in the family *Alphaflexiviridae*, including cactus virus X (CVX), zygocactus virus X (ZyVX), schlumbergera virus X (SchVX), and pitaya virus X (PiVX), have been shown to infect *Hylocereus* plants, and have been reported in Taiwan, Hainan and Guizhou in China [3–6] and in other countries, including Brazil, Korea and the USA [7–9]. Until now, there has been no evidence of any other new viruses infecting pitaya.

In September 2017, we conducted field surveys of virus diseases in pitaya orchards in Guizhou, China. Stem tissues showing viral symptoms such as chlorosis, mottling and malformation were collected and quickly frozen in liquid nitrogen. Total RNA was isolated from the symptomatic tissues using TRIzol Reagent (Invitrogen, CA, USA), and the ribosomal RNA was then removed using a Ribo-Zero Plant Kit (Illumina, CA, USA), following the manufacturer’s protocol. Finally, cDNA libraries were constructed and sequenced with paired-end reads (2 × 150 bp) on an Illumina HiSeq 4000 platform. For the data analysis procedure, low-quality sequences were removed from raw reads and *de novo* assembly of clean reads was done using Trinity

Handling Editor: T. K. Frey.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00705-019-04503-7>) contains supplementary material, which is available to authorized users.

✉ Yuhua Ma
m_yh79@163.com

Qianming Zheng
zqm851015@163.com

Xiaohe Wang
wangxiaoke9191@163.com

Junliang Zhou
gsszjl2008@163.com

¹ Institute of Pomology Science, Guizhou Provincial Academy of Agricultural Sciences, Huaxi District, Jinzhu Town, Guiyang 550006, Guizhou, China

software [10]. The resulting contigs were used to search the NCBI GenBank database using the online tools BLASTx and BLASTn.

BLASTn analysis identified six contigs with lengths from 2,502 nt to 7,978 nt that showed 98–99% nucleotide sequence identity to with EpMoaV [11] and 70% identity to jujube mosaic-associated virus (JuMaV) [12]. The six contigs were assembled from 4,711–34,087 reads. BLASTx analysis revealed that the encoded polyprotein shared 34–37% amino acid sequence identity with those of blackberry virus F (BVF, KJ413252), citrus yellow mosaic virus (CYMV) [13], fig badnavirus 1 (FBV-1) [14] and yacon necrotic mottle virus (YNMoV) [15]. After reassembly and removal of overlapping sequences, a circular DNA sequence of 7,837 nt was generated.

To confirm the presence of pitaya badnavirus-like sequences, total DNA was isolated from symptomatic samples using by CTAB method and used as template for PCR amplification. Then, three specific primer sets covering the whole circular genome were designed based on the assembled circular DNA sequence (Fig. 1). These primers are shown in Supplementary Table 1. The PCR reaction was performed using *Ex Taq* DNA polymerase (TaKaRa, Dalian, China), and amplicons were sequenced using the Sanger method. After *de novo* assembly and removal of overlapping sequences, the circular DNA genome sequence was reconstructed. This virus was tentatively named “pitaya badnavirus 1” (PiBV1) (Fig. 1), and its genome sequence was deposited in the GenBank database under the accession number MK991812.

The complete genome of PiBV1 is 7,837 bp in length and has a GC content of 50.2%. The full-length genome showed 98.7% nucleotide sequence identity to EpMoaV, 54.6% to BVF and 40.0–48.2% to other members of the genus *Badnavirus* (Supplementary Table 2). Similar to other members of the genus *Badnavirus* [11, 12, 15, 16], a complementary sequence (5′-TGGTATCAGAGCGAGGTT-3′) of a plant tRNA-Met binding site was also identified in the PiBV1 genome (Fig. 1). Therefore, we defined these 18 nucleotides as the start site (TGGTATCAGAGCGAGGTT₁₋₁₈). The PiBV1 genome was also found to contain a TATA box (TATA₇₆₄₇₋₇₆₅₀) and a polyadenylation-like signal (AAT AAA₇₇₉₇₋₇₈₀₂), which are typical motifs found in badnavirus genomes [11, 17].

Four putative ORFs were predicted in the viral genome, using the program ORF Finder (<http://www.ncbi.nlm.nih.gov/projects/gorf>) (Fig. 1). Theoretical molecular masses were calculated using ExPASy (http://web.expasy.org/compute_pi/). ORF1 (nt position 192–731) encoded a 179-aa protein with a molecular mass of 19.9 kDa (Table 1) that shared 99.4% amino acid sequence identity with the protein encoded by ORF1 of EpMoaV, and 17.4–47.5% with those of other members of the genus *Badnavirus* (Supplementary

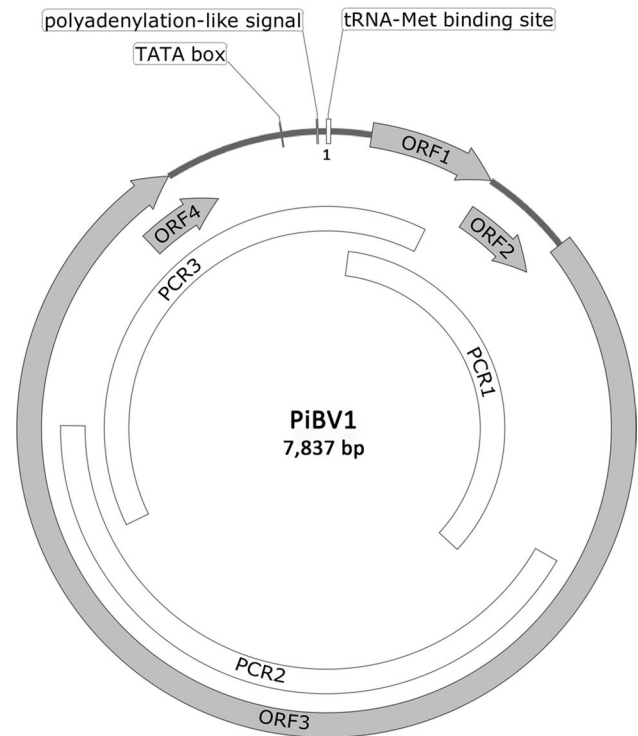


Fig. 1 Schematic representation of the PiBV1 genome. The “tRNA-Met binding site” was set as the start site of the PiBV1 genome. ORF1–4 are the four putative open reading frames (ORFs) of the PiBV1 genome. The three open boxes (PCR1–3) overlapping each other represent the PCR fragments used to construct the PiBV1 genomic DNA sequence

Table 2). Similar to other badnaviruses [18], the ORF1 also contained a protein domain of unknown function (DUF1319). ORF2 (nt position 728–1,132) encoded a 134-aa protein (14.8 kDa), overlapping with the 3′ end of the ORF1. Pairwise comparisons suggested that the protein encoded by ORF2 shared 94.8% amino acid sequence identity with the ORF2 protein of EpMoaV and 14.4–36.5% with those of other members of the genus *Badnavirus*. ORF3 (nt position 1,129–7,140) encoded a 2,003-aa polyprotein precursor (225.7 kDa), and it shared a very high level of amino acid sequence identity (98.9%) with the ORF3 of EpMoaV and 34.7–46.0% identity with that of other members of the genus *Badnavirus*. ORF3 was found to contain a number of highly conserved regions encoding the movement protein, coat protein, viral aspartic protease, zinc knuckle finger, reverse transcriptase, and RNase H domains, which are homologous to those of other badnaviruses [15, 18–20]. ORF4 (nt position 6,888–7,289) encoded a putative 133-aa protein (14.2 kDa) overlapping with the 3′ end of ORF3 by 253 nt. The ORF4 of PiBV1 shared 97.7% amino acid sequence identity with that of EpMoaV.

To examine the taxonomic position of PiBV1, phylogenetic trees were constructed using the neighbour-joining

Table 1 Putative open reading frames (ORFs) in the PiBV1 genome sequence

ORF	Genome position (5'-3')	Protein		Function	Top two viruses with the highest aa sequence identity*
		aa	kDa		
1	192-731	179	19.9	Unknown	99.4%, ORF1 of epiphyllum mottle-associated virus 42.5%, ORF1 of sweet potato badnavirus B
2	728-1,132	134	14.8	Unknown	94.8%, ORF2 of epiphyllum mottle-associated virus 36.5%, ORF2 of dioscorea bacilliform SN virus
3	1,129-7,140	2,003	225.7	Polyprotein	98.9%, ORF3 of epiphyllum mottle-associated virus 46.0%, ORF3 of blackberry virus F
4	6,888-7,289	133	14.2	Unknown	97.7%, ORF4 of epiphyllum mottle-associated virus -

*The GenBank accession numbers of these viruses are listed in Supplementary Table 2

method in MEGA 7.0 [21]. ClustalX was used for multiple sequence alignment. The full-length genome nucleotide sequence and the deduced amino acid sequence of ORF3 were used for constructing phylogenetic trees with 23 members covering the two major clades of badnaviruses [11]. One thousand bootstrap replicates were used. The phylogenetic tree based on full-length genome nucleotide sequences, revealed that PiBV1 and EpMoaV formed a single clade, and were closely related to BVF, dracaena mottle virus (DrMV), lucky bamboo bacilliform virus (LBBV), YNMoV and JuMaV (Fig. 2). A phylogenetic tree based on the deduced amino acid sequence of the ORF3 polyprotein also showed similar topology (data not shown).

For the genus *Badnavirus*, the species demarcation criterion proposed by the ICTV is less than 80% nucleotide sequence identity in the reverse transcriptase and RNase H (RT-RNase H) regions [22]. In pairwise sequence comparisons, the RT-RNase H region of PiBV1 shared 99.4% nucleotide sequence identity with that of EpMoaV. In April 2019, EpMoaV was isolated from a member of the genus *Epiphyllum* of the family *Cactaceae* [11]. Given the high level of identity of the PiBV1 and EpMoaV genome sequences (98.7%), we supposed that PiBV1 and EpMoaV were two isolates of a badnavirus that infects members of the family *Cactaceae*. However, since the nucleotide sequences of the RT-RNase H regions of PiBV1 and other badnaviruses were less than 70.5% identical, we propose that PiBV1 can be considered a new member of the genus *Badnavirus*.

To summarize, we determined the complete genome sequence of PiBV1, a new badnavirus infecting red pitaya in Guizhou, China. Until now, all viruses identified in red pitaya have belonged to the genus *Potexvirus* of the family *Alphaflexiviridae*. To our knowledge, this is the first report of a badnavirus infecting red pitaya. Large-scale field investigations and tests are required for understanding the prevalence and genetic variation of PiBV1.

Acknowledgements This research was financially supported by the Youth Fund Program of Guizhou Academy of Agricultural

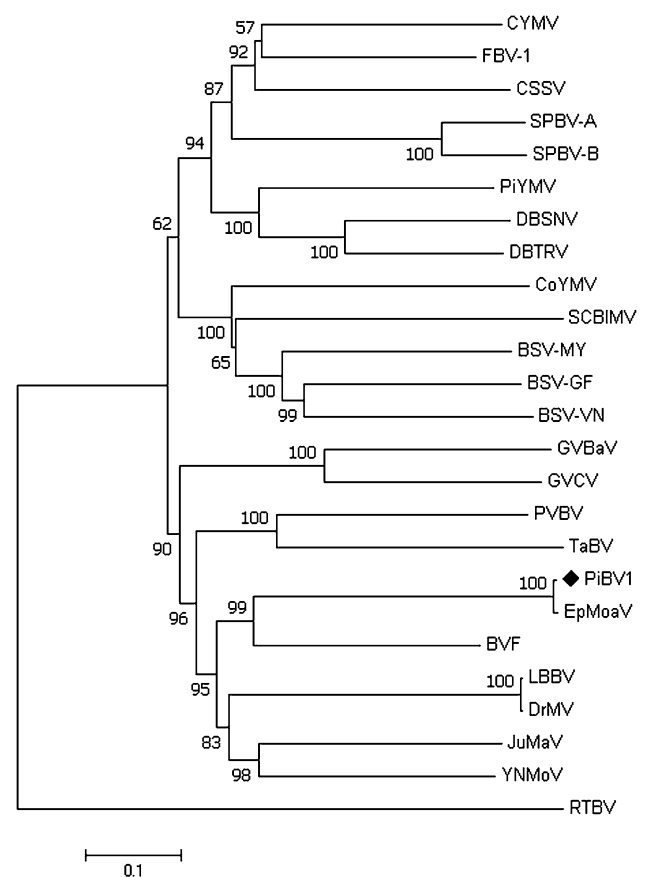


Fig. 2 Phylogenetic analysis of PiBV1 and other members of the genus *Badnavirus*. The phylogenetic tree was constructed using the neighbour-joining method based on full-length nucleotide sequences. Numbers at the branch points are percent bootstrap values (1000 replications). The GenBank accession numbers for other members of the genus *Badnavirus* are listed in Supplementary Table 2. RTBV (rice tungro bacilliform virus, NC_001914), belonging to the genus *Tungrovirus* in the family *Caulimoviridae*, was used as an out-group

Sciences ([2017]11), the National Natural Science Foundation of China (31560547), and the Guizhou Science and Technology Fund Plan of Guizhou Provincial Department of Science and Technology ([2016]1146).

Compliance with ethical standards

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

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

References

- Hua QZ, Chen CB, Tel-Zur N, Wang HC, Wu JY, Chen JY, Zhang ZK, Zhao JT, Hu GB, Qin YH (2018) Metabolomic characterization of pitaya fruit from three red-skinned cultivars with different pulp colors. *Plant Physiol Biochem* 126:117–125
- Wu YW, Xu J, He YZ, Shi MY, Han XM, Li WY, Zhang XW, Wen XP (2019) Metabolic profiling of Pitaya (*Hylocereus polyrhizus*) during fruit development and maturation. *Molecules* 24:1114
- Liou MR, Hung CL, Liou RF (2001) First report of Cactus virus X on *Hylocereus undatus* (Cactaceae) in Taiwan. *Plant Dis* 85(2):229
- Liou MR, Chen YR, Liou RF (2004) Complete nucleotide sequence and genome organization of a Cactus virus X strain from *Hylocereus undatus*. *Arch Virol* 149(5):1037–1043
- Peng C, Yu NT, Liu ZX (2016) Molecular identification of Cactus virus X infecting *Hylocereus polyrhizus* (Cactaceae) in Hainan island, China. *Plant Dis* 100(9):1956
- Zheng QM, Wang Z, Cai YQ (2017) High-throughput sequencing analysis of viruses in pitaya from Guizhou Province, China. *Acta Phytopathologica Sinica* 47(3):364–369
- Duarte L, Alexandre M, Galletti S (2008) Potexvirus diversity in Cactaceae from Sao Paulo state in Brazil. *J. Plant Pathol* 90(3):545–551
- Kim JS, Park CY, Lee SH (2016) First report of Cactus virus X infecting *Hylocereus undatus* in Korea. *Plant Dis* 100(12):2544
- Gazis R, Poude B, Dey K, Zhang S, Palmateer AJ, Capoverde V, Baker C, Adkins S (2018) First report of cactus virus X in *Hylocereus undatus* (dragon fruit) in Florida. *Plant Dis* 100(12):2544
- Grabherr MG, Haas BJ, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech* 29(7):644–652
- Lan PX, Tian TY, Pu LL, Rao WL, Li F, Li RH (2019) Characterization and detection of a new badnavirus infecting *Epiphyllum* spp. *Arch. Virol* 164(7):1837–1841
- Du KT, Liu SJ, Chen ZR, Fan ZF, Wang H, Tian GH, Zhou T (2017) Full genome sequence of jujube mosaic-associated virus, a new member of the family *Caulimoviridae*. *Arch Virol* 162(10):3221–3224
- Huang Q, Hartung JS (2001) Cloning and sequence analysis of an infectious clone of Citrus yellow mosaic virus that can infect sweet orange via Agrobacterium-mediated inoculation. *J Gen Virol* 82:2549–2558
- Laney AG, Hassan M, Tzanetakis IE (2012) An integrated badnavirus is prevalent in Fig germplasm. *Phytopathology* 102:1182–1189
- Lee YJ, Kwak HR, Lee YK, Kim MK, Choi HS, Seo JK (2015) Complete genome sequence of yacon necrotic mottle virus, a novel putative member of the genus Badnavirus. *Arch Virol* 160(4):1139–1142
- Lim S, Park JM, Kwon SK, Cho HS, Kim HS, Lee SH, Moon JS (2019) Complete genome sequence of a tentative new member of the genus Badnavirus identified in *Codonopsis lanceolate*. *Arch Virol* 164(6):1733–1737
- Chingandu N, Zia-ur-rehman M, Sreenivasan TN, Surujdeo-Maharaj S, Umaharan P, Gutierrez OA, Brown JK (2017) Molecular characterization of previously elusive badnaviruses associated with symptomatic cacao in the New World. *Arch Virol* 162(5):1363–1371
- Kazmi SA, Yang Z, Hong N, Wang G, Wang Y (2015) Characterization by small RNA sequencing of Taro Bacilliform CH Virus (TaBCHV), a novel badnavirus. *PLoS One* 10:e0134147
- Li YQ, Deng CL, Qiao Y, Zhao XL, Zhou Q (2017) Characterization of a new badnavirus from *Wisteria sinensis*. *Arch Virol* 162(7):2125–2129
- Iskra-Caruana ML, Duroy PO, Chabannes M, Muller E (2014) The common evolutionary history of badnaviruses and banana. *Infect Genet Evol* 21:83–89
- Kumar S, Stecher G, Tamura K (2016) Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870
- King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (2012) Badnavirus. In: *Virus taxonomy. Ninth report of the international committee on taxonomy of viruses*. Elsevier Academic Press, London

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