The complete genome sequence of andraeanum bacilliform virus, a novel badnavirus encoding the polyprotein through two open reading frames

Siqi Zhou
South China Agricultural University
Xiaoyun Zhou
South China Agricultural University
Yiwen He
South China Agricultural University
Guohui Zhou
South China Agricultural University
Xin Yang (yangxin@scau.edu.cn)
South China Agricultural University https://orcid.org/0000-0001-9499-4170

Research Article

Keywords: andraeanum bacilliform virus (AnBV), badnavirus, Anthurium andraeanum, viral genome sequence

Posted Date: January 25th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2487079/v1

Abstract

A novel badnavirus, tentatively named as "andraeanum bacilliform virus" (AnBV), was identified in *Anthurium andraeanum* plants displaying leaf deformation and darkening. The viral complete genome sequence was determined by high-throughput and PCR sequencing. Similar to those of other Badnavirus, the genome of AnBV consists of a circular DNA molecule of 7212 bp with six open reading frames (ORFs) potentially coding for six proteins. Interestingly, ORF3 is split into two parts with a 55-nt intergenic region between them. Similar genome structure was reported in sweet potato pakakuy virus (SPPV) and jujube mosaic-associated virus (JuMaV). Pairwise comparisons of the highly conserved RT/RNase H region revealed the highest amino acid sequence identity (78.6%) to Dioscorea bacilliform AL virus 2 (DBALV-2). Phylogenetic analysis revealed that AnBV and DBALV-2 form a separate clade within the genus Badnavirus, suggesting that AnBV is a novel badnavirus infecting *Anthurium andraeanum*.

Full Text

Badnaviruses (Family: *Caulimoviridae*; Genus: *Badnavirus*) are one group important plant pathogenic viruses, causing diseases on agricultural or horticultural crops and ornamental plants over the world, especially on tropical and subtropical crops produced via vegetable propagation, such as banana, sugarcane and sweet potato et al. They are non-enveloped bacilliform DNA viruses with 120–150 nm in length and 30 nm in width and transmitted naturally by aphids or mealybugs and artificially by vegetative propagation. Up to 2022, 68 species have been recognized by the International Committee on Taxonomy of Viruses (ICTV, https://ictv.global/taxonomy) in the genus *Badnavirus* (Supplementary Table 1). The genome of badnavirus consists of a 7.2–9.2 kb double-stranded open circular DNA, which encodes 3 to 8 open reading frames (ORFs). Among these ORFs, ORF1 to ORF3 are conserved in all badnaviruses. The function of ORF1 and ORF2 are undetermined. Generally, ORF3, the largest ORF, ranging in length from 5.1 to 6.0 kb, encodes a large polyprotein which is self-cleaved into four or five viral functional proteins, including movement protein (MP), coat protein (CP), aspartate protease (AP), reverse transcriptase (RT) and ribonuclease H (RNaese H). It is not known whether the additional ORFs are expressed, nor what their function could be. In addition to infective episomal forms, some badnaviruses exist as endogenous viruses integrated into their host genomes [1].

Lineages in *Anthurium andraeanum* Linden are ones of the most popular ornamental plants being enjoyable both foliage and flower in potting or gardening. They are propagated in large-scale via tissue-culture under controlled conditions worldwide. No virus infecting *Anthurium* plants has been reported so far. In 2020, plants of *A. andraeanum* cv. Xiaojiao showing virus-like symptoms were observed in a commercial greenhouse in South China Botanical Garden, Guangzhou, Guangdong, China. The symptoms included deformation and darkening of leaves, stunting of plants, and miniaturization of flowers (Supplementary Fig. 1). The symptomatic leaves were collected and a novel badnavirus, tentatively named as andraeanum bacilliform virus (AnBV), was identified by small RNA sequencing. The complete genome of AnBV was determined and analyzed.

To scan virus in the sample, the total RNA of leaf samples was extracted using a RNeasy Plant Mini Kit (Qiagen, Germany). The small RNA fragments were isolated on a 12% polyacrylamide gels, followed by sequencing on an Illumina Hiseq2500 sequencer performed by Sangon Biotech Co. (Shanghai Biotech Co., China). The sequencing data were analyzed with SPAdes software (Sangon, China) to obtain contigs. Out of 1050 contigs obtained, 3 contigs (named contig 1, 2 and 3) with length in 97, 146 and 122 nt matched badnavirus genome by searches in the GenBank Database (http://www.ncbi.nlm.nih.gov) using local BLASTn programs. The reads of small RNAs forming these 3 contigs were in low abundance (15 reads per million) and ranged 21 to 24 nt in size. No contig was found matching to any other viral genome. The three contigs showed the highest nucleotide identities with banana streak CA virus (86%), citrus yellow mosaic virus (79%) and blackberry virus F (73%), respectively. The contigs were mapped to expected typical badanvirus circle genome according to the position of their matched regions on above viruses (Supplementary Fig. 2). The primers (Supplementary Table 1) based on contig sequences and locations were designed and overlapping (nested) polymerase chain reaction (PCR) were conducted. The amplicons were sequenced with Sanger's method to fill the gaps. After assembly, a circle DNA sequence with the length of 7212 bp and GC content of 41.9% was obtained. A badnaviral characteristic methionine tRNA biding site ($tggtatcagagcgaggtt_{1-18}$), a TATA-box ($tataaata_{7031-7038}$) and conserved ORFs were identified in this sequence. Therefore, it was regarded as the complete genome sequence of AnBV and deposited in the NCBI GenBank database under accession number OQ095368. To verify the episomal virus form, electron microscopy observation and rolling circle amplification (RCA) were conducted, but the positive results were not obtained. This might perhaps be due to the limited sensitivity of these two techniques and low titter of the virus, as corroborated by low abundance of virus-derived RNA reads in the small RNA sequencing test.

The genome structure of AnBV (Supplementary Fig. 2) is atypical compared to most badnaviruses. Six ORFs are found with ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder). ORF1 and ORF2 with a 4 nucleotide-overlap (ATGA) are conserved in all badnaviruses. Whereas ORF3 is split into two parts, ORF3a and ORF3b, with a 55-nt intergenic region between them. ORF3a is 1179 nt in length and 4 nt overlapped with ORF2. ORF3b, the largest ORF, consisted of 4092 nt. Two small ORFs, ORF4 and ORF5, are followed behind ORF3b. Similar genome structure was reported in sweet potato pakakuy virus (SPPV, FJ560943) [2] and jujube mosaic-associated virus (JuMaV, KX852476) [3]. The position, the length of the split ORFs and intergenic region are different between these three viruses (Fig. 1). The most obvious difference is the split site. It is located more rear in SPPV than in AnBV and JuMaV. As a result, SPPV ORF3a includes MP and CP coding regions, while AnBV and JuMaV ORF3a only includes MP coding region.

The sequence identities (MegAlign by the Clustal W program) and conserved domain analyses (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) indicated that AnBV is a novel species in *Badnavirus*. The whole genome sequence nucleotide (nt) identities of AnBV with recognized badnaviruses range from 39.8–54.7%. The all badnavirus conserved domains (MP, CP, AP RT and RNase H) are found in the putative products encoded by ORFs of AnBV, despite the MP is separately located on the ORF3a product (Supplementary Fig. 2). The identities of AnBV in the conserved RT/RNase H coding

region with recognized badnaviruses are less than 72.2% (nt) and 78.6% (amino acid, aa) (Table 1). Both values are less than the guidelines (80% nt and 89% aa) for demarcation of species within the genus *Badnavirus* setting by ICTV [4].

Phylogeny analysis further confirmed that AnBV is a unique species in the genus Badnavirus. The all recognized badanviruses (Supplementary Table 2) were used to reconstructed phylogenetic trees by the Maximum-likelihood method in MEGA 11 [5]. The phylogenetic trees based on amino acid sequences of conserved RT/RNase H region (Fig. 2) revealed that badnaviruses clustered into six groups. The three viruses with split ORF3 were scattered in different groups. AnBV was closely related to Dioscorea bacilliform AL virus 2 (DBALV-2), while JuMaV was closed to epiphyllum mottle-associated virus (EpMoaV) and Dracaena mottle virus (DrMV), and SPPV is form a group alone.

Next to SPPV and JuMaV, AnBV is the third badnavirus with split ORF3 being identified up to now. It is expectable that more viruses with similar genome structure will be discovered in future. We suggest they should be recognized as members of the genus *Badnavirus* rather than be grouped to a new genus proposed by Du et al. [3], though the genome organization is one of the main characteristics that distinguish genera in family *Caulimoviridae* [4]. The expression pattern of ORF3a and ORF3b, the infection cycle and the damage of AnBV are need to investigate furtherly.

Declarations

Funding: The study was supported by the Department of Agriculture of Guangdong Province (20210909).

Conflict of interest: The authors have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Bhat, A., Hohn, T., and Selvarajan, R. Badnaviruses: the current global scenario. Viruses 2016, 8:177. doi: 10.3390/v8060177
2. Kreuze, J.F.; Perez, A.; Gargurevich, M.G. and Cuellar, W.J. Badnaviruses of Sweet Potato: Symptomless Coinhabitants on a Global Scale. Front. Plant Sci. 2020, 11:313. doi: 10.3389/fpls.2020.00313
3. Du, K.; Liu, S.; Chen, Z.; Fan, Z.; Wang, H.; Tian, G.; Zhou, T. Full genome sequence of jujube mosaic-associated virus, a new member of the family Caulimoviridae. Arch. Virol. 2017, 162, 3221–3224. Doi: 10.1007/s00705-017-3438-6
4. King, A.M.Q.; Adams, M.J.; Carstens, E.B.; Lefkowitz, E.J. Virus taxonomy: Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, London. 2012.

5. Tamura, K., Stecher, G., Kumar, S. MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 2021, 38, 3022–3027. https://doi.org/10.1093/molbev/msab120.

Tables

Table 1
Identities (%) of andraeanum bacilliform virus (AnBV) genome and conserved RT/ RNase H sequence with selected badnaviruses.

Viruses*	Complete genome	RT/ RNase H	
		nucleotide	amino acid
Dioscorea bacilliform AL virus 2 (MH404164)	47.6	72.2	78.6
Grapevine badnavirus 1 (NC055481)	47.6	69.7	75.2
Dioscorea bacilliform RT virus 1 (NC038986)	46.6	68.8	76.1
Jujube mosaic-associated virus (KX852476).	44.9	64.7	68.8
Sweet potato pupakuy virus (FJ560943)	43.0	61.7	63.7

* Dioscorea bacilliform AL virus 2 and grapevine badnavirus 1 are two selected badnaviruses share highest genomic sequence nt identities with AnBV. Dioscorea bacilliform AL virus 2 and dioscorea bacilliform RT virus 1 are two selected badnaviruses share highest RT/RNase H sequence aa identities with AnBV. Jujube mosaic-associated virus and sweet potato pupakuy virus are two b adnavirus having a similar genome structure with AnBV.

Figures

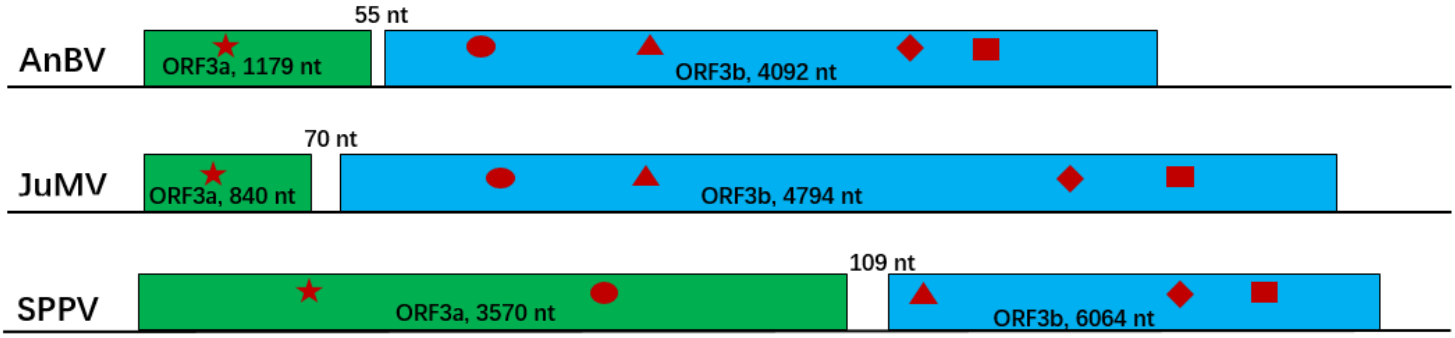

Figure 1

Comparison of ORF3a and ORF3b between three badnaviruses with similar genome structure. Three badnaviruses with split ORF3: AnBV, andraeanum bacilliform virus (OQ095368); SPPV, sweet potato pupakuy virus (FJ560943); JuMaV, jujube mosaic-associated virus (KX852476). Putative function of the

ORF encoding protein predicted through BLASTp search and conserved domain search (https://www.ncbi.nlm.nih.gov). Asterisk, circle dot, triangle, diamond and rectangle indicate the conserved domain of movement protein, coat protein, aspartic protease, reverse transcriptase and RNase H, respectively.

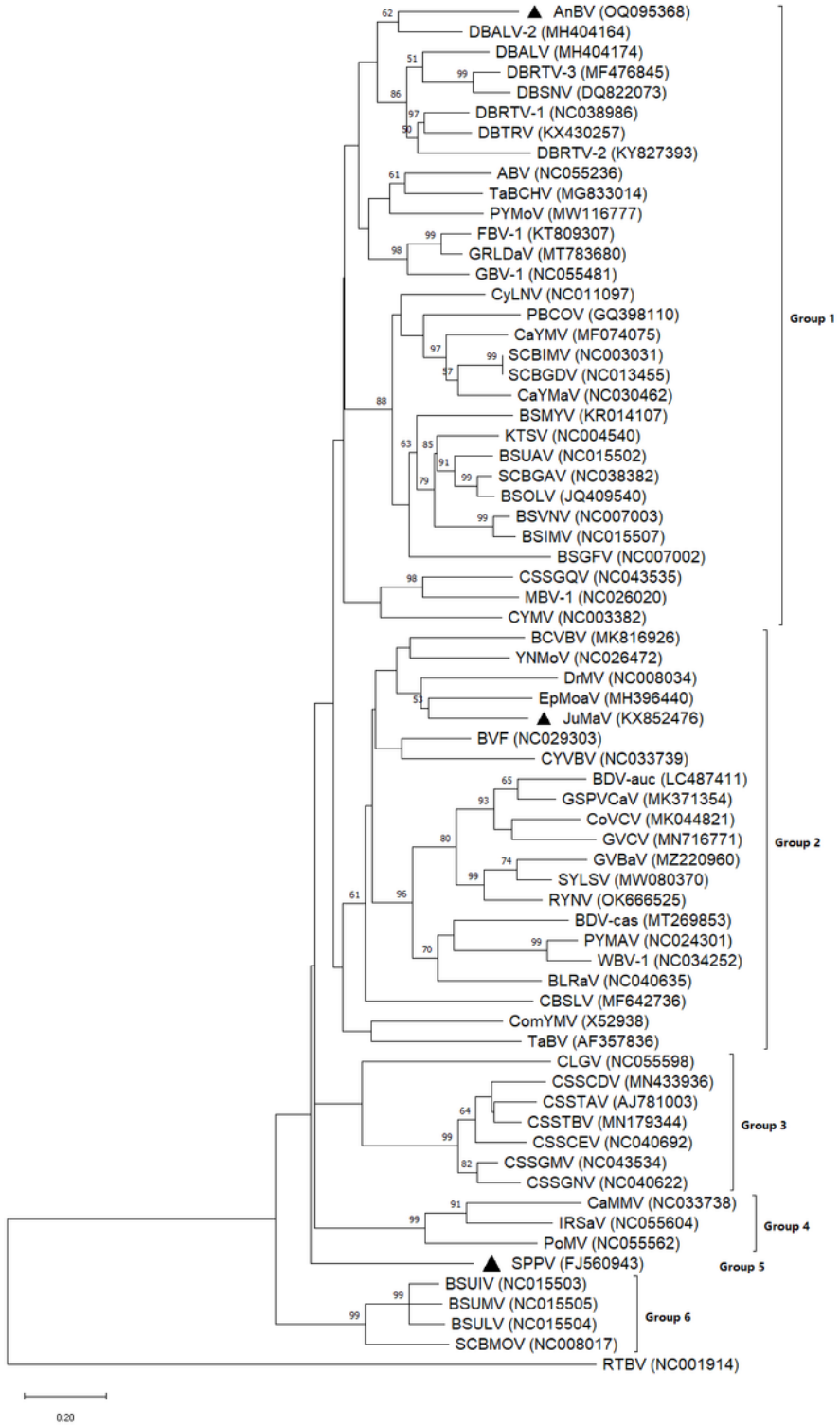

Figure 2

Phylogenetic trees of badnaviruses based on conserved RT and RNase H amino acid sequences.
Phylogenetic trees were generated using Maximum-likelihood method with 1000 bootstrap replicates. The triangles indicate the position of three viruses with split ORF3. RTBV (rice tungro bacilliform virus, NC001914), belonging to the genus Tungrovirus in the family Caulimoviridae, was used as an out-group.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfile.docx