CrossMark

Characterization of a new badnavirus from *Wisteria sinensis*

Yongqiang Li[1] · Congliang Deng[2] · Yan Qiao[3] · Xiaoli Zhao[2] · Qi Zhou[4]

Abstract The complete genome sequence of a previously undescribed badnavirus isolated from a wisteria plant exhibiting mosaic and crinkle symptoms in Beijing, China, was determined. The circular double-stranded DNA genome of this virus was 7362 bp in size with four open reading frames (ORFs 1 to 4) on the plus strand. Sequence analysis showed that this virus shared the highest (69%) nucleotide (nt) sequence identity with pagoda yellow mosaic associated virus (PYMAV). In the RT-RNase H region of the ORF-3 encoded polyprotein, this virus shared 74% nt sequence identity with PYMAV. Phylogenetic analysis provided further evidence that the virus identified in this study is a member of a new species in the genus *Badnavirus*. The name wisteria badnavirus 1 (WBV1) is proposed for this new virus.

Wisteria sinensis, also known as Chinese wisteria, is a long-lived perennial shrub-like climbing vine in the family Fabaceae [13]. It is commonly distributed in natural forests, riparian zones and ruderal areas due to the characteristic of ease of establishment, hardiness, fast growth and long lifespan, but in city parks and home gardens, it is highly valued for its foliage and the blue or white pea-like flowers that adorn walls and pergolas as ornamental plants. As with other flowering plants, *W. sinensis* is affected by different biotic agents. Wisteria powdery mildew, which is caused by various members of the genus *Erysiphe* [7]; bacterial leaf spot, which is caused by *Pseudomonas syringae* pv. syringae [25]; and wisteria mosaic disease, which is caused by wisteria vein mosaic virus (WVMV) [4, 6, 10, 17] and cucumber mosaic virus [20], are important diseases of *W. sinensis* that reduce their quality and ornamental value. Due to vegetative propagation by cuttings, grafting, and air layering, viruses are easily transmitted to propagules and can be dispersed worldwide in the *W. sinensis* plant trade.

RNA silencing is a conserved antiviral defense mechanism in plants. Once a plant is infected by a virus, antiviral silencing is triggered by cleaving viral double-stranded RNA and highly structured single-stranded RNA (ssRNA) resulting from base pairing between plus- and minus-strand viral RNAs or imperfect folding of self-complementary sequences within viral ssRNA and processing these cleavage products into small interfering RNAs (siRNAs) of different length. The sequences of these virus-derived siRNAs overlap and can be assembled into long contigs and even complete virus genome sequences by using bioinformatic tools, providing a rapid and unbiased method for virus identification. This method has contributed tremendously to unraveling the nature of microbial agents associated with disease and characterization of a number of novel or known plant viruses with RNA or DNA genomes [1–3, 19, 27].

✉ Yongqiang Li
   lyq@bua.edu.cn

✉ Qi Zhou
   zhouq@bjciq.gov.cn

[1] College of Plant Science and Technology, Beijing University of Agriculture, Beijing 102206, China

[2] Plant Laboratory of Beijing Entry-Exit Inspection and Quarantine Bureau, Beijing 100026, China

[3] Beijing Plant Protection Station, Beijing 100029, China

[4] Chinese Society of Inspection and Quarantine, Beijing 101312, China

Badnaviruses (family *Caulimoviridae*) are important agricultural and horticultural pathogens that cause serious crop losses [8]. In recent years, new badnaviruses have been identified from different plants including taro [14], yacon [16], pagoda tree [26] and grapevine [18]. The badnavirus genome consists of a 7.2- to 9.2-kb circular double-stranded DNA molecule that is encapsidated in a nonenveloped bacilliform virion and typically encodes three open reading frames (ORFs) that are tandemly arranged on the plus strand [8]. ORF1 encodes a small protein of unknown function, while the protein encoded by ORF2 is virion-associated. ORF3 encodes a large polyprotein that is post-translationally cleaved into the movement protein (MP), coat protein (CP), viral aspartic protease (AP), reverse transcriptase (RT) and ribonuclease H (RNase H) [8]. Additional ORFs have been reported in taro bacilliform virus [28], piper yellow mottle virus [11], cacao swollen shoot virus [9], PYMAV [26], citrus yellow mosaic virus [12] and dracaena mottle virus [22].

In June of 2015, a *W. sinensis* plant with mosaic and crinkle symptoms in the leaves was observed in Beijing, China. To identify the causal agent(s), the symptomatic leaves were collected and quickly frozen in liquid nitrogen, and total RNA was extracted from the samples using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's recommended procedures. High-quality RNA assessed using a NanoDrop ND-1000 spectrophotometer and Agilent 2100 Bioanalyzer was then used for sRNA library construction following the recommendations of New England Biolabs. Briefly, 16- to 28-nt small RNAs were gel-purified from a 15% PAGE gel, adaptors were added to both ends of the purified RNAs, and the RNAs were amplified by RT-PCR using adaptor-specific primers. The PCR products were isolated and gel-purified.

Sequencing was performed on a HiSeq 2000 platform (Illumina). Raw Illumina sRNA reads were trimmed and cleaned by removing sequences smaller than 16 nucleotides or longer than 30 nucleotides, low-quality tags, and polyA or N tags, using an in-house Perl script. The resulting reads were assembled into contigs using Velvet software [29] with a *k*-mer value of 17. These contigs were then further analyzed by BLAST searches against the GenBank nr nucleotide and protein databases at the NCBI.

After quality trimming of the raw sequence data, a total of 29,116,531 clean reads were obtained and 14,303 contigs were assembled using the Velvet program, with lengths ranging from 33 to 355 nt. A BLASTn and BLASTx analysis of assembled contigs against the NCBI database, using a highly homology sequence search, identified 238 and 162 contigs showing high sequence similarity to WVMV (genus *Potyvirus*, family *Potyviridae*) and pagoda yellow mosaic associated virus (PYMAV) (genus *Badnavirus*, family *Caulimoviridae*), respectively, while most of the remaining sequences were of host origin. The presence of the two viruses was confirmed by RT-PCR/PCR using primers that were designed based on the assembled contigs (Supplemental Table 1).

To obtain the complete genome sequence of the candidate badnavirus, total DNA was extracted from leaf tissues using a TaKaRa MiniBEST Plant Genomic DNA Extraction Kit, and PCR was conducted with LA *Taq* DNA polymerase using primer pairs targeting overlapping fragments designed based on the assembled contigs. The PCR products were ligated into the vector pMD19-T (TaKaRa, Dalian, China), and eight independent clones from each sample were sequenced. The reconstructed circular genome of the new badnavirus isolate, designated as "wisteria badnavirus 1" (WBV1), was 7326 nt in length, which is

Fig. 1 Schematic representation of the genome organization of wisteria badnavirus 1 showing open reading frames 1-4 (ORF1-ORF4), as well as the tRNA primer binding site and conserved functional motifs in the ORF3 polyprotein
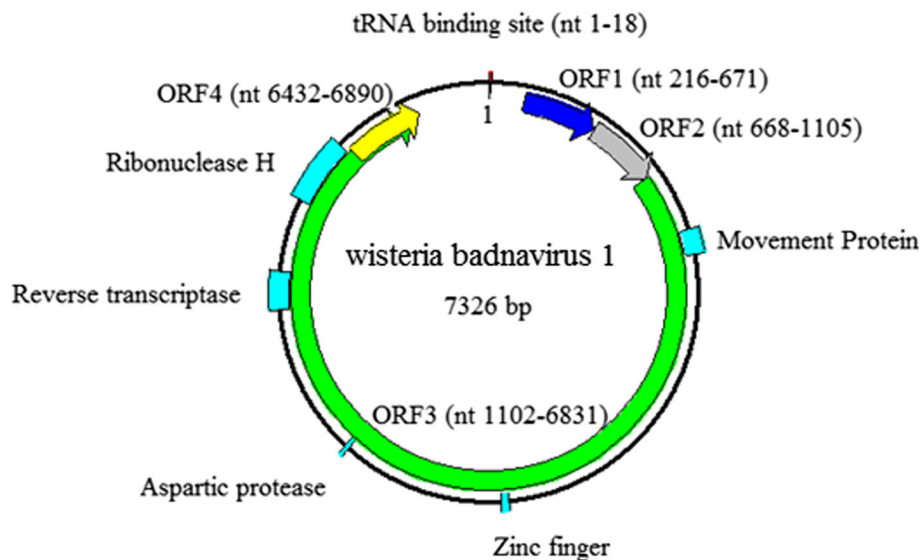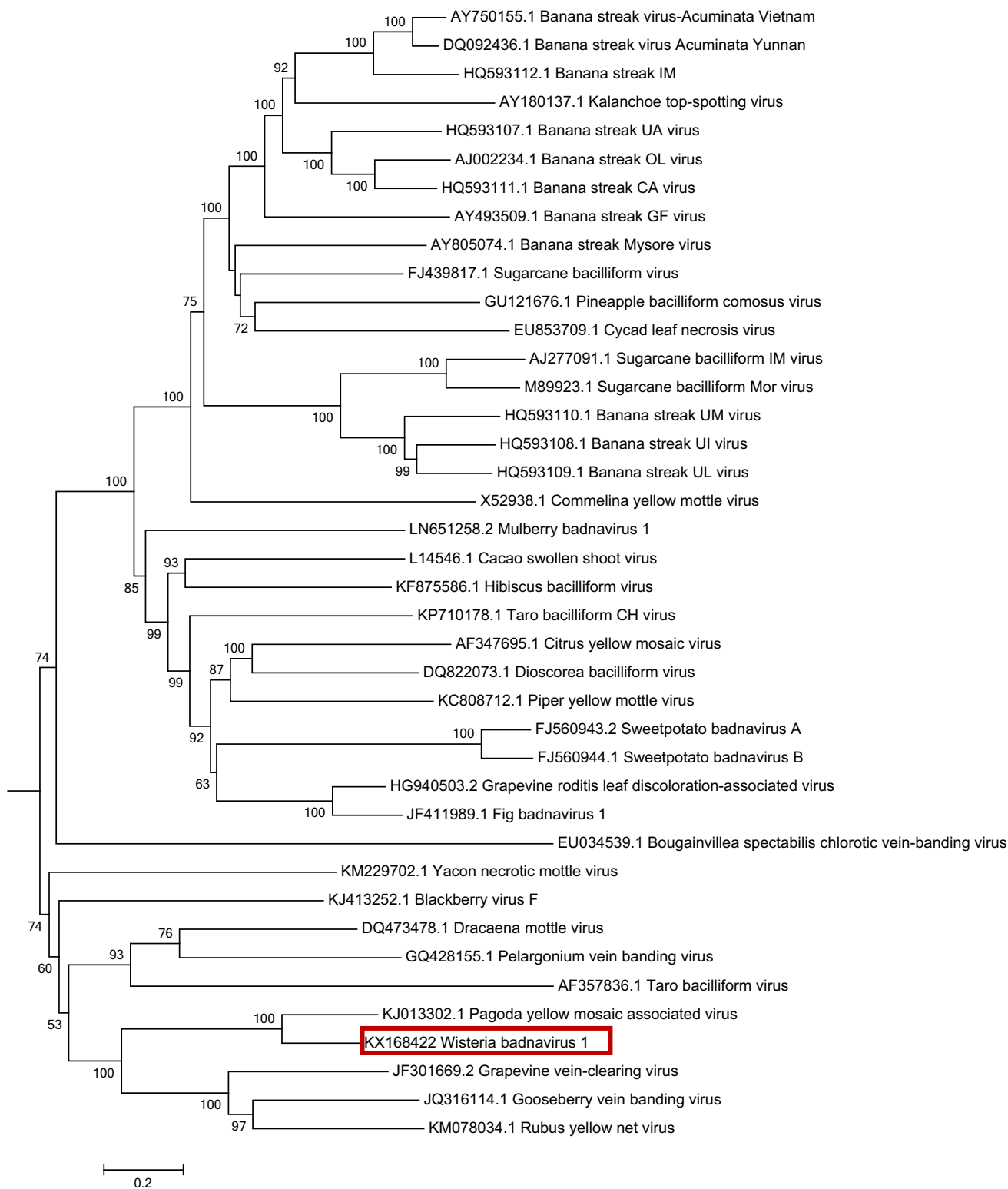
Fig. 2 Neighbor-joining phylogenetic tree obtained from alignment of the available complete genome sequences of badnaviruses, with 1000 bootstrap replicates. Bootstrap values are shown at the relevant nodes, and only bootstrap values ≥50% are shown. The scale bar represents genetic distance (substitutions per nucleotide)

within the range of badnavirus genomes, and deposited in the GenBank database with the accession number KX168422. To confirm the existence of this virus, rolling-circle amplification (RCA) was conducted using an Illustra TempliPhi 500 Amplification Kit (GE Healthcare). The RCA product was digested with StuI or SphI, and the expected band was observed in an agarose gel (Supplemental Figure 1). Sequence comparison of WBV1 with other reported badnaviruses showed that it shared the highest (69%) nt sequence identity with PYMAV. As with other badnaviruses, the genome starts at the conserved tRNA^{met} primer binding site (5′-TGGTATCA-GAGCTCAACA-3′) [5], which lies before a 215-bp-long 5′ non-coding region, followed by three conserved ORFs, which have overlapping start and stop codons (5′-ATGA-3′) (Fig. 1). A fourth potential ORF, ORF4, which is not typical of badnaviruses but has also been observed in the genomes of fig badnavirus 1, cacao swollen shoot virus (CSSV) and PYMAV [15, 21, 26], was found overlapping the end of ORF3 (Fig. 1). ORF1 (nt 216–671) of WBV1 potentially encodes a 151-aa protein with a predicted molecular weight (MW) of 17.8 kDa. BLAST analysis showed the ORF1 shares 72% nt and 73% aa sequence identity with ORF1 of PYMAV, and no conserved domains were detected. ORF2 (nt 668–1,105) encodes a 145-aa protein with a predicted MW of 16.3 kDa. It shares 75% nt and 79% aa sequence identity with ORF2 of PYMAV. ORF3 (nt 1,102–6,831) encodes a 1,909-aa polyprotein with a predicted MW of 215.6 kDa that shares 72% nt and 78% aa sequence identity with PYMAV ORF3 and contains domains homologous to those of MP, AP, RT and RNase H and a highly conserved zinc-finger-like RNA binding domain (CXCX2CX4HX4C) in the polyproteins of badnaviruses. In the RT-RNase H region of the ORF3, WBV1 shares the highest (74%) nt sequence identity with PYMAV. This value is below the 80% identity threshold used as the species demarcation criterion in the genus *Badnavirus* according to the International Committee on Taxonomy of Viruses [8], indicating that WBV1 represents a new badnavirus. ORF4 partially overlaps ORF3 (nt 6432-6890) and encodes a peptide of 152 aa. This ORF shares 74% nt and 75% aa sequence identity with PYMAV ORF4. To examine the taxonomic position of WBV1, all of the available badnaviruses in the GenBank database were downloaded and aligned using the Clustal X program [24]. A phylogenetic tree was constructed with the aligned sequences using the neighbor-joining (NJ) method implemented in the MEGA6.06 program [23] with the best-fit models, and phylogenetic analysis showed that WBV1 clustered together with PYMAV (Fig. 2). To the best of our knowledge, this is the first report of a DNA virus infecting wisteria. In our survey of viral diseases in wisteria, mosaic and crinkle symptoms were commonly observed, and the

association of WBV1 with any of the observed symptoms should be investigated. Furthermore, studies focused on the transmission pattern, host range, epidemiology and population structure of this virus should be conducted to form the basis for virus control.

References

1. Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, Navalinskiene M, Samuitiene M, Boonham N (2009) Next generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Mol Plant Pathol 10:537–545
2. Al Rwahnih M, Daubert S, Golino D, Rowhani A (2009) Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. Virology 387:395–401
3. Al Rwahnih M, Dave A, Anderson MM, Rowhani A, Uyemoto JK, Sudarshana MR (2013) Association of a DNA virus with grapevines affected by red blotch disease in California. Phytopathology 103:1069–1076
4. Bos L (1970) The identification of three new viruses isolated from *Wisteria* and *Pisum* in the Netherlands, and the problem of variation within the potato virus Y group. Netherlands J Plant Pathol 76:8–46
5. Bouhida M, Lockhart BE, Olszewski NE (1993) An analysis of the complete sequence of a sugarcane bacilliform virus genome infectious to banana and rice. J Gen Virol 74:15–22
6. Clover GRG, Denton JO, Denton GJ (2015) First report of *Wisteria vein mosaic virus* on *Wisteria* spp. in the United Kingdom. New Dis Rep 31:1
7. Cook RTA, Denton JO, Denton G (2015) Pathology of oak-wisteria powdery mildew. Fungal Biol 119:657–671
8. Geering ADW, Hull R (2012) Family Caulimoviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier, London
9. Hagen LS, Jacquemond M, Lepingle A, Lot H, Tepfer M (1993) Nucleotide sequence and genomic organization of cacao swollen shoot virus. Virology 196:619–628
10. Halliwell RS, Porterfield MR (1984) First report of a potyvirus infecting *Wisteria floribunda* in the United States. Plant Dis 68:450
11. Hany U, Adams I, Glover R, Bhat A, Boonham N (2014) The complete genome sequence of Piper yellow mottle virus (PYMoV). Arch Virol 159:385–388
12. 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

13. Jiang YF, Chen XL, Lin H, Wang F, Chen F (2011) Floral Scent in Wisteria: Chemical composition, emission pattern, and regulation. J Am Soc Hortic Sci 136:307–314

14. 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

15. Laney AG, Hassan M, Tzanetakis IE (2012) An Integrated badnavirus is prevalent in Fig germplasm. Phytopathology 102:1182–1189

16. 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 160:1139–1142

17. Liang WX, Song LM, Tian GZ, Li HF, Fan ZF (2006) The genomic sequence of *Wisteria vein mosaic virus* and its similarities with other potyviruses. Arch. Virol 151:2311–2319

18. Maliogka VI, Olmos A, Pappi PG, Lotos L, Efthimiou K, Grammatikaki G, Candresse T, Katis NI, Avgelis AD (2015) A novel grapevine badnavirus is associated with the Roditis leaf discoloration disease. Virus Res 203:47–55

19. Marais A, Faure C, Mustafayev E, Barone M, Alioto D, Candresse T (2015) Characterization by deep sequencing of Prunus virus T, a novel Tepovirus infecting Prunus species. Phytopathology 105:135–140

20. Milojević K, Radovic N, Djekic IM, Vucurovic A, Nikolic D, Bulajic A, Krstic B (2016) First report of Cucumber mosaic virus infecting *Wisteria* sinensis in Serbia. Plant Dis. doi:10.1094/PDIS-01-16-0096-PDN

21. Muller E, Sackey S (2005) Molecular variability analysis of five new complete cacao swollen shoot virus genomic sequences. Arch Virol 150:53–66

22. Su L, Gao S, Huang YW, Ji CQ, Wang D, Ma Y, Fang RX, Chen XY (2007) Complete genomic sequence of Dracaena mottle virus, a distinct badnavirus. Virus Genes 35:423–429

23. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

24. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

25. Varvaro L (1987) A bacterial leaf spot of wisteria in Italy. EPPO Bull 17:287–290

26. Wang YL, Cheng XF, Wu XX, Wang AM, Wu XY (2014) Characterization of complete genome and small RNA profile of pagoda yellow mosaic associated virus, a novel badnavirus in China. Virus Res 188:103–108

27. Wu QF, Ding SW, Zhang YJ, Zhu SF (2015) Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. Ann Rev Phytopathol 53:425–444

28. Yang I, Hafner G, Dale J, Harding R (2003) Genomic characterisation of taro bacilliform virus. Arch Virol 148:937–949

29. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829