Characterization of the promoter of *Grapevine vein clearing virus*

Y. Zhang,1 C. A. Angel,1 S. Valdes,1 W. Qiu2 and J. E. Schoelz1

1Division of Plant Sciences, University of Missouri-Columbia, Columbia, MO 65211, USA
2Center for Grapevine Biotechnology, William H. Darr School of Agriculture, Missouri State University, Mountain Grove, MO 65711, USA

*Grapevine vein clearing virus* (GVCV) is a recently discovered DNA virus in grapevine that is closely associated with the grapevine vein clearing syndrome observed in vineyards in Missouri and surrounding states. The genome sequence of GVCV indicates that it belongs to the genus *Badnavirus* in the family *Caulimoviridae*. To identify the GVCV promoter, we cloned portions of the GVCV large intergenic region in front of a GFP gene present in an *Agrobacterium tumefaciens* binary vector. GFP expression was assessed by ELISA 3 days after agroinfiltration of *Nicotiana benthamiana* leaves. We found that the GVCV DNA segment between nts 7332 and 7672 directed expression of GFP and this expression was stronger than expression using the *Cauliflower mosaic virus* 35S promoter. It was revealed by 5’ and 3’ RACE that transcription was initiated predominantly at nt 7571 and terminated at nt 7676.

Members of the genus *Badnavirus* are plant viruses that replicate by reverse transcription; their genome is composed of circular, ds DNA encapsidated into a bacilliform virion with a diameter of 30 nm and a length of 120–150 nm (Tidona & Darai, 2011). The first badnavirus to be identified was *Commelina yellow mottle virus* (CoYMV) (Medberry et al., 1990; Migliori & Lastra, 1978). Since that first discovery, many new badnavirus genomes have been characterized at the nucleotide sequence level, and much information has been inferred from their nucleotide sequences. For example, it is generally accepted that all badnaviruses have a single promoter that drives transcription of a terminally redundant pregenomic RNA (Tidona & Darai, 2011). However, less information is available concerning the precise map locations of the badnavirus promoters and the precise transcription start sites. In particular, both the 5’ and 3’ ends of transcripts have been mapped only for CoYMV and *Rice tungro bacilliform virus*, whereas the 5’ end has been mapped for several variants of the *Banana streak virus* (BSV) (Medberry et al., 1990; Qu et al., 1991; Schenk et al., 2001). Information on promoters and transcripts is an important next step in the characterization of any badnavirus beyond the level of the nucleotide sequence. Furthermore, given the examples of how infectious clones of badnaviruses can be developed and inoculated, an understanding of the terminally redundant transcript of badnaviruses is important, if not essential. Generally, infectious clones for badnaviruses contain a longer than full-length genome copy, with duplicated promoter regions on both ends of the insert (Dasgupta et al., 1991; Huang & Hartung, 2001; Medberry et al., 1990).

Recently a badnavirus has been discovered in grapevines grown in the Midwest USA that is consistently associated with a new disease. The disease was first observed in *Vitis vinifera* in Missouri vineyards in 2004 and has since been found in Indiana, Illinois and Arkansas (Guo et al., 2014; Lunden, 2010; Qiu & Lunden, 2007; Zhang et al., 2011). Typical symptoms of the disease include chlorosis in major or minor veins, especially in young shoots that emerge in early spring. The affected vines eventually become dwarfed, bear smaller clusters with fewer fruits and, in the most severe cases, the infection kills the vines (Guo et al., 2014; Qiu & Lunden, 2007; Zhang et al., 2011). The virus, designated *Grapevine vein clearing virus* (GVCV), was identified through deep sequencing of small interfering RNAs (siRNAs) and it was the first DNA virus reported in grapevines (Zhang et al., 2011). The complete nucleotide sequence of GVCV was assembled from clones amplified through PCR, revealing a putative badnavirus genome with a size of 7753 bp that contains three ORFs (Fig. 1a). Within the *Badnavirus* genus, GVCV is most closely related to *Taro bacilliform virus* (Zhang et al., 2011). Although a final decisive proof is still lacking, GVCV is considered the most likely causal agent of grapevine vein clearing syndrome.

In this study, we used an agroinfiltration assay in *Nicotiana benthamiana* to identify the GVCV promoter and transcript initiation site. This information was subsequently used to confirm the 5’ and 3’ ends of the GVCV transcript in infected grapevine leaf tissue. Knowledge of the promoter and transcript initiation sites will be valuable...
To identify the putative GVCV promoter region, we amplified a GVCV segment that included both TATA1 and TATA2 and the start codon of sORF A (Fig. 1b). The template for PCR was pGVCV 6192-1935, one of the clones used for determining the GVCV nucleotide sequence (Zhang et al., 2011). PCR primers used for amplification of all GVCV sequences are listed in Table S1. The PCR product was cut with the restriction enzymes EcoRI and HindIII and the DNA segment replaced the 35S promoter present in p35S-GFP (Angel et al., 2011), as the 35S promoter is delimited by EcoRI and HindIII sites, obtaining pGVCV-GFP. This cloning step also fused the GVCV sORF A coding region upstream and in-frame with the GFP coding sequence (Fig. 1b).

To determine whether the large intergenic region of GVCV contained promoter activity, we performed transient expression assays in N. benthamiana and compared GFP expression in pGVCV-GFP, p35S-GFP and a binary vector lacking any promoter sequence (pA35S-GFP). To construct pA35S-GFP, p35S-GFP was digested by EcoRI and HindIII and then recircularized by ligation, removing the 35S promoter. The plasmids pGVCV-GFP, p35S-GFP and pA35S-GFP were transformed into the Agrobacterium tumefaciens strain AGL1 and agroinfiltrated into young leaves of N. benthamiana following procedures described in Angel et al. (2011). To reduce leaf-to-leaf variation in response to agroinfiltration, all constructs were agroinfiltrated into leaf panels on a single leaf and at least three leaves were included in each test. Leaves were collected at 3 days after infiltration (a.i.) and the expression level of GFP was assessed by illumination with a UV lamp and quantified by a GFP-specific ELISA. For GFP quantification, N. benthamiana leaf tissue agroinfiltrated with GVCV promoter constructs and controls was collected at 3 days a.i., homogenized with 0.05 M sodium phosphate buffer (pH 7) (10 μl buffer per 10 mg tissue) and the concentration of total protein assessed by a Bradford assay (Bio-Rad). ELISA was performed using a GFP ELISA kit (Cell Biolabs) following the manufacturer’s instructions. GFP values for the GVCV promoter were normalized against the GFP expression level obtained with the 35S-GFP control present in the same leaf.

At 3 days a.i., illumination of the leaves with the UV lamp indicated no visual differences between expression of GFP from pGVCV-GFP and p35S-GFP infiltrated leaves (Fig. 2a). However, ELISA data showed that the strength of the GVCV full-length promoter expression was significantly higher than with the 35S promoter (Fig. 2b). The agroinfiltration assay and quantification of GFP by ELISA was repeated in two additional tests and in each case, expression from the GVCV promoter was significantly stronger than the 35S promoter (data not shown).

To further define the GVCV promoter, a second recombinant plasmid was made that contained only the TATA1 box (p341-GFP in Fig. 1b) and its expression was compared to pGVCV-GFP expression. The primers used for amplification for the development of an infectious clone and ultimately for completion of Koch’s postulates.

An analysis of the GVCV nucleotide sequence revealed two potential TATA boxes that could be used to initiate transcription of the GVCV pregenomic transcript: TATA1 (beginning at nt 7539) and TATA2 (beginning at nt 7131) (Fig. 1b). TATA1 is located within the large intergenic region, between ORF 3 and the tRNA binding site, whereas TATA2 is found within GVCV ORF 3 near its 3′ end (Fig. S1, available in the online Supplementary Material). Furthermore, the large intergenic region contains four short ORFs (sORFs) (Figs 1b and S1). Multiple sORFs are found within the leader sequence of the pregenomic transcript of all caulimoviruses (Baughman & Howell, 1988; Fütterer et al., 1988; Pooggin et al., 2006).
Fig. 2. Expression of GFP from a putative GVCV promoter element. Sections of the GVCV large intergenic region were cloned in front of the GFP gene present in an Agrobacterium tumefaciens binary vector. GFP expression was assessed by ELISA three days after agroinfiltration of Nicotiana benthamiana. A representative picture taken 3 days after infiltration shows GFP fluorescence under UV light. The Empty Vector leaf panel refers to pKYLX7, the Agrobacterium binary vector used for all constructs in this study (Angel et al., 2011) (a). GFP ELISA was used to compare the strength of GVCV full-length promoter and the Cauliflower mosaic virus (CaMV) 35S promoter; expression was normalized against p35S-GFP (n = 3, P <0.01) (b); the strength of the GVCV full-length promoter and was compared with expression from the GVCV 341 promoter region, and normalized against pGVCV-GFP (n = 3, P <0.01) (c). of the GVCV sequences in p341-GFP are listed in Table S1 and, as with pGVCV-GFP, the amplified DNA was cut with EcoRI and HindIII for replacement of the 35S promoter in p35S-GFP. The strength of the p341-GFP promoter was evaluated in an agroinfiltration assay, along with the negative control pΔ35S-GFP and the full-length promoter present in pGVCV-GFP. Each construct was agroinfiltrated into separate leaf panels on the same leaf and multiple leaves were evaluated in each test. This experiment showed that sequences surrounding the TATA1 box were sufficient to drive expression of GFP, and that expression of GFP from p341-GFP was significantly higher than that from pGVCV-GFP (Fig. 2c). This test was repeated three times and the results were confirmed. Given the strength of this promoter element and its position in the GVCV genome sequence relative to the genome sequences of other caulimoviruses, we concluded that the sequences between nts 7332 and 7672 constituted the core promoter of GVCV.

All caulimoviruses produce a terminally redundant RNA that has a dual role for expression of viral proteins and for replication during reverse transcription. To further characterize the transcription and replication strategies of GVCV, we sought to identify the initiation and termination sites of the transcript generated from the GVCV promoter. To identify the 5’ end of the transcript expressed from the GVCV promoter, pGVCV-GFP was infiltrated into N. benthamiana leaves and total RNA was isolated at 3 days a.i. using the RNasy Mini kit (Qiagen), following the manufacturer’s instructions, and used as a template for a 5’ RACE reverse transcription (RT) reaction. The 5’ RACE reverse transcription (RT)-PCR was performed with a reverse primer that corresponded to a sequence within the GFP (primer GFP-Rev 2; Fig. 1 and Table S1) and a forward primer that corresponded to a sequence within the GFP promoter, pGVCV-GFP was infiltrated into N. benthamiana leaves and total RNA was isolated at 3 days a.i. using the RNasy Mini kit (Qiagen), following the manufacturer’s instructions, and used as a template for a 5’ RACE reverse transcription (RT) reaction. The 5’ RACE reverse transcription (RT)-PCR was performed with a reverse primer that corresponded to a sequence within the GFP (primer GFP-Rev 2; Fig. 1 and Table S1) and a forward primer that corresponded to a sequence within the GFP promoter.

To identify the initiation and termination sites of the transcript generated from the GVCV promoter, we sought to identify the initiation and termination sites of the transcript generated from the GVCV promoter. To identify the 5’ end of the transcript expressed from the GVCV promoter, pGVCV-GFP was infiltrated into N. benthamiana leaves and total RNA was isolated at 3 days a.i. using the RNasy Mini kit (Qiagen), following the manufacturer’s instructions, and used as a template for a 5’ RACE reverse transcription (RT) reaction. The 5’ RACE reverse transcription (RT)-PCR was performed with a reverse primer that corresponded to a sequence within the GFP (primer GFP-Rev 2; Fig. 1 and Table S1) and a forward primer that corresponded to a sequence within the GFP promoter.

To investigate whether the same transcript initiation site was used in the GVCV-infected grape tissue as in the transient expression assay in N. benthamiana, a second 5’ RACE reaction was performed on total RNA isolated from GVCV-infected grape leaves. We used the RNA isolation protocol described in Reid et al. (2006), which included cetyl trimethylammonium bromide to counteract the high content of secondary metabolites present in grape leaves. In the RT-PCR we used the reverse primer GVCV-377Rev (Fig. 1b, Table S1) and a single PCR product was generated.
with an approximate size of 600 bp (Fig. 3a). Sequencing of three clones showed that the transcript was initiated at nt 7571 in two and at nt 7574 in one (Fig. S2), confirming that the same transcript initiation sites were used in GVCV-infected grape leaves as in the transient expression assays in *N. benthamiana*.

To identify the 3′ end of the major GVCV present in GVCV-infected grape leaves, 3′ RACE was performed on total RNA isolated from GVCV-infected grapevine using the SMARTer™ RACE cDNA Amplification kit (Clontech). For the 3′ RACE reaction, the forward primer (GVCV-7200Fwd) corresponded to the GVCV sequence at nts 7200–8727 (Table S1) and this reaction generated a band approximately 550 bp in size (Fig. 3b). This PCR band was cloned into pGEM-T Easy and five clones were sequenced to determine the 3′ termination site. Nucleotide sequencing showed that the GVCV transcript was terminated at nt 7676 bp in all four clones (Fig. S2). Taken together, transcript initiation at nt 7571 and termination at 7676 would yield a terminal redundancy of 106 nt in the GVCV mRNA (Fig. 3c).

An inspection of the nucleotide sequence immediately upstream from the transcript terminus indicated that GVCV does not contain a canonical mRNA polyadenylation (poly(A)) signal (5′-AAUAAA-3′) (Fig. S3). The poly(A) signal is highly conserved in mammals, but is much less prevalent in plants. A survey of transcripts of *Arabidopsis* revealed that only approximately 10% of the transcripts contained a poly(A) signal in the predicted location (Loke et al., 2005). A poly(A) signal is much more common in caulimoviruses and a well-defined poly(A) signal can be found in many *Caulimoviridae* sequences, including *Cauliflower mosaic virus* (CaMV) and CoYMV (Medberry et al., 1990; Sanfaçon et al., 1991), the type species for the genera *Caulimovirus* and *Badnavirus*, respectively. Furthermore, the presence of a poly(A) signal has been used to approximate the position of the 3′ end of a number of caulimovirus sequences in the absence of experimental evidence for the 3′ end (Pooggin et al., 1999). However, even with badnaviruses the poly(A) signal cannot always be identified. For example, a poly(A) signal cannot be located in the sequence of the BSV strain GF, even though it is found in the BSV strain OL (Fig. S3, Remans et al., 2005). Therefore, it may not be surprising that GVCV lacks a clear poly(A) signal.

Next generation sequencing techniques involving deep sequencing of siRNAs have greatly facilitated the discovery and characterization of new plant virus genomes and sequences, including the discovery of GVCV (Zhang et al., 2011). Furthermore, in the past few years many new badnavirus genomic sequences have been characterized including *Sweet potato badnavirus* B (Kreuze et al., 2009), *Sweet potato badnavirus* A (Mbanzibwa et al., 2011), *Gooseberry vein banding virus* (Xu et al., 2011), *Fig badnavirus* 1 (Laney et al., 2012), *Rubus yellow net virus* (Kalischuk et al., 2013) and *Piper yellow mottle virus* (Hany et al., 2014), as well as GVCV (Zhang et al., 2011). However, little is known about the genome organization of these viruses beyond the positioning of the ORFs. In the present work, we have delimited the boundaries of the core promoter and identified the 5′ and 3′ ends of the GVCV pregenomic transcript. This information is essential for the development of an infectious GVCV clone and for further characterization of the expression strategy of the three GVCV ORFs.

**Acknowledgements**

This research was supported by the grant (USDA 2009-38901-19962) from the United States Department of Agriculture, the Missouri Agricultural Experiment Station, and the D. F. Millikan Endowment Fund at the University of Missouri.

**References**


