Genomic Characterization and Population Structure of a Badnavirus Infecting Blackberry

Muhammad Shafiq Shahid, Department of Plant Pathology, Division of Agriculture, University of Arkansas, Fayetteville 72701; Nina Aboughanem-Sabanadzovic, Institute for Genomics, Bioinformatics and Biotechnology, and Sead Sabanadzovic, Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State 39762; and Ioannis E. Tzanetakis, Department of Plant Pathology, Division of Agriculture, University of Arkansas

Blackberry (Rubus spp.) is cultivated worldwide for its high nutritional value, with acreage and production increasing significantly in the last 20 years. Diseases pose major production constraints across the United States and the world (Ellis et al. 1984; Garrett 1978; Martin et al. 2013; Williamson et al. 1998). In particular, viral diseases have emerged as a major concern for producers in the southeastern United States, with more than a dozen new viruses identified in the last decade (Martin et al. 2013; Thekke- Veetil et al. 2013) (N. Aboughanem-Sabanadzovic, S. Sabanadzovic, and I. E. Tzanetakis, unpublished). The vast majority of these viruses are associated with a disorder referred to as blackberry yellow vein disease (BYVD) (Martin et al. 2013).

BYVD is caused by virus complexes, with severity generally increasing with the number of viruses infecting plants (Martin et al. 2013). Because of the number of viruses and vectors involved, genetic resistance has not been pursued as a control strategy. Vector control is a challenging undertaking given that their identity and distribution is largely yet unknown. Planting virus-tested material, certified to be free of viruses of concern, is currently the better management practice; even though clean plants do not eliminate disease, they prolong the productive lifespan of plants and produce profitability (Gergerich et al. 2015).

All viruses associated with BYVD identified to date—namely, blackberry yellow vein associated virus (BYVaV), tobacco ringspot virus (TRSV), blackberry chlorotic ringspot virus (BCRV), and blackberry virus Y (BYY), among others (Martin et al. 2013)—have an RNA genome; therefore, the virus could be eliminated from infected germplasm using thermo-, chemo-, or cryotherapy combined with in vitro shoot tip culture (Cielinska 2002, 2003; Cielinska and Zawadzka 1999; Wang and Valkonen 2009; Wang et al. 2008). Because of the complex etiology of BYVD, the presence of a single virus in the germplasm predisposes plants to the disease development as additional viruses accumulate (Martin et al. 2013). In such cases, a badnavirus infection poses the worst-case scenario.

Members of some genera of the family Caulimoviridae, with badnaviruses being the prime example, could integrate to the host genome (Daniells et al. 2001; Geering et al. 2001; Ndowora et al. 1999; Pahalawatta et al. 2008). Whereas the integrated form does not cause disease, it could transcribe, initiating virus replication and infection (episomal form). Genome integration has major implications for all affected crops, especially those propagated by asexual means such as blackberry, because there is no method available for virus elimination.

This study investigated viruses associated with BYVD. One of the plants screened during a BYVD survey was infected by a putative new member of the genus Badnavirus, tentatively named blackberry virus F (BVF). After characterizing the genome of BVF, we evaluated several aspects of its epidemiology, including distribution, population structure, host range, and association with BYVD, as well as the virus ability to integrate into the host genome.

Materials and Methods

Virus characterization. The isolate used for virus characterization was obtained from a blackberry plant showing BYVD symptoms (MS-469). Nucleic acids were extracted as previously described (Laney et al. 2012; Sabanadzovic and Abou Ghanem-Sabanadzovic 2009) and used for degenerate oligonucleotide-primed reverse-transcription polymerase chain reaction (DOP-RT-PCR) (Telenius et al. 1992), cloning, and sequencing. Several virus and virus-like sequences were found in MS-469, including BYVaV, TRSV, blackberry vein banding-associated virus, and a putative new badnavirus. After acquiring sequence data from DOP-RT-PCR-generated clones, the genomic sequence of the badnavirus was completed by overlapping PCR amplicons, as described by Laney et al. (2012). Products were ligated into TOPO 2.1 (Life Technologies) and transformed into One Shot TOP10 Escherichia coli cells (Life Technologies) and transformed into One Shot TOP10 Escherichia coli cells (Life Technologies) according to the manufacturer’s instructions. Plasmids were sequenced to obtain at least 3× coverage of the genome, with contigs assembled using Lasergene (DNASTar Inc.).

Phylogenetic and recombination analyses. Nucleotide and amino acid sequence alignments between BVF and other badnaviruses were performed using Lasergene. Open reading frames (ORF) were predicted using ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/). Phylogenetic analysis was performed using both maximum-likelihood and neighbor-joining methods, with the latter presented here. The program used was ClustalX (Thompson et al. 1994) with Kimura’s correction (Kimura 1980). Bootstrap values were calculated based on 1,000 pseudoreplicates and the tree was visualized with Treeview (Page 1996). Possible recombination events were evaluated using RD4 (Martin et al. 2015) and considered trustworthy when detected by the majority of the algorithms used.
Host range. Leaf tissue from MS-469 was used to mechanically inoculate at least 8 plants from 20 plant species at the three- to four-leaf stage (Table 1). Briefly, plants were dusted with 600-mesh carborundum and inoculated with BVF-infected tissue ground (1:10, vol/vol) in chilled phosphate buffer (0.1 M, pH 7.2) containing 2% nicotine. After inoculation, plants were rinsed and kept in an insect-proof greenhouse with a 14-h photoperiod. Plants were observed for symptom development and tested for possible latent infections by performing PCR on nucleic acids extracted from apical leaves collected a month postinoculation. Amplicons were sequenced directly to verify their identity.

Distribution, population structure, and detection. In total, 403 BYVD-affected blackberry samples were collected from Arkansas (n = 228), Florida (n = 26) Georgia (n = 61), Mississippi (n = 13), North Carolina (n = 53), and South Carolina (n = 22) from 2008 to 2012. The population structure study of the virus was based on a segment between nucleotides 683 and 2,032 (unpublished data), with PCR amplicons cloned and sequenced for at least 3× coverage. This was done in order to minimize false-negative results due to isoforms. Primer set 1562F (5'-AGTTGAGCTGTTGACCC-3')/2029R (5'-GCTTCCCTAC TAGAGCTGTG-3') was designed on genomic regions that showed 100% nucleotide identities in all studied isolates and provided the best amplification among sets tested. A duplex PCR using 1562F/2029R together with NADHF (5'-AGTAGATGCTATACGAAGGATC-3')/NADHR (5'-AGTAGATGCTATACGACATATAA-3'), amplifying the NADH dehydrogenase ND2 subunit mRNA (Tzanzetakis et al. 2007), was applied to all 403 BYVD-affected samples. PCR cycling conditions included an initial denaturation for 3 min at 94°C; followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 20 s, and extension at 72°C for 30 s; and a final extension step at 72°C for 10 min. The PCR products were analyzed by electrophoresis in 1.5% Tris-acetate-EDTA agarose gels after staining with GelRed and visualized with a UV transilluminator.

Virus integration. To differentiate between the integrated virus genome and episomal forms, RT-PCR, rolling circle amplification (RCA), and Southern hybridization were employed. RT-PCR on all BVF-infected material was used to determine whether the virus sequences detected were actively transcribed in a given plant. Total nucleic acids from PCR-positive plants were digested with 1 μL of DNase I (Fermentas) for 1 h at 37°C to eliminate genomic and putative integrated copies of the virus. Nucleic acids were purified and RT-PCR performed as described before (Poudel et al. 2013).

Eight herbaceous plant species sustaining BVF replication upon artificial inoculation and eight naturally infected blackberry plants were used in RCA. Nucleic extracts from 100 mg of foliar tissue were treated with 2 U of ribonuclease A (Thermo Scientific) and 0.02 U of ribonuclease T1 (Sigma-Aldrich) for 30 min at 37°C followed by a 30-min incubation with 4 U of proteinase K (Merck Millipore). Material was EthOH precipitated and 100 to 200 ng of template was used to amplify circular DNA molecules. After RCA, performed essentially as described by Laney et al. (2012), concatamers were digested overnight with 10 to 20 U of KpnI or SacI (one recognition site in the virus genome) and SphI, (two recognition sites in the virus genome). Identity of the RCA products was verified by hybridization using digoxigenin-labeled probes prepared using 1562F/2029R amplicons and following a previously described procedure (Laney et al. 2012).

Field-grown blackberry plants that tested BVF PCR/RT-PCR positive, PCR positive/RT-PCR negative, and PCR negative, and BVF-infected Nicotiana rustica, were used in Southern blot analysis. Genomic DNA was isolated as described earlier and treated with 6 U of ribonuclease A and 0.08 U of T1 RNase for 60 min at 37°C, followed by proteinolytic digestion with 10 U of proteinase K at 37°C overnight and phenol-chloroform extraction. DNA (10 to 15 μg) was digested overnight at 37°C with either KpnI or SphI endonucleases with one or two recognition sites, respectively, in the BVF genome. Probe labeling and hybridization were performed as described above.

Results

Characterization and phylogenetic and recombination analyses. The full-length nucleotide sequence of the BVF genome is deposited in GenBank under accession number KJ413252. The genome consists of 7,663 bp and codes for four ORF. A putative trNA<sup>BET</sup> binding-site sequence (TTGGATACAGCTTTATTACAA1-19), characteristic for all known badnaviruses, was identified within the intergenic region, as was the TATA box (TTATTAA1-550s-751s) (Fig. 1). ORF1 (ATG183 to TAG1346) encodes a putative protein (p1) of 184 amino acids (aa) with molecular mass of 21 kDa. The p1 function is unknown, as is the case of the proteins encoded by ORF in the same genomic location of other badnaviruses. BVF p1 shares 33% aa identity to its citrus yellow mosaic virus (CYMV) counterpart. ORF2 (ATG183 to TAG1346) encodes a putative DNA binding protein (p2) of 157 aa in length with a molecular mass of 15 kDa that is most similar to the fig badnavirus-1 (FBV-1) ortholog with 33% aa identities. ORF3 (ATG1346 to TAG1959) encodes a polyprotein (p3) of 1,948 aa with a molecular mass of 223 kDa that

| Table 1. List of the herbaceous hosts mechanically inoculated with blackberry virus F (BVF) isolate MS-469 |
|---|---|---|---|---|---|
| Species, cultivar | Exp. 1 | Exp. 2 | Exp. 3 | BVF PCR + | Infection (%) |
| Beta vulgaris 'Ruby Queen' | 3 | 6 | 3 | 0/12 | 0 |
| B. vulgaris, Swiss chard 'H-32' | 4 | 4 | 4 | 5/12 | 42 |
| Brassica rapa subs. pekinensis | 2 | 4 | 5 | 0/12 | 0 |
| Chenopodium quinoa | 4 | 4 | 8 | 7/16 | 44 |
| Cowpea 'Elegance 801' | 5 | 5 | 4 | 2/14 | 14 |
| Cucumis melo 'Edisto 47' cantaloupe | 4 | 6 | 3 | 3/13 | 23 |
| C. sativus 'National Pickling' | 5 | 6 | 6 | 6/17 | 35 |
| Glycine max 'Willam 82' | 4 | 2 | 6 | 2/12 | 17 |
| Gomphrena globosa | 4 | 5 | 4 | 7/13 | 54 |
| Nicotiana accidentalis | 4 | 6 | 4 | 0/14 | 0 |
| N. benthamiana | 6 | 5 | 4 | 0/15 | 0 |
| N. rustica | 6 | 4 | 6 | 3/16 | 19 |
| N. tabacum | 6 | 5 | 8 | 5/20 | 25 |
| Phaseolus vulgaris var. 'Black Valentine' | 4 | 2 | 3 | 0/9 | 0 |
| Pisum sativum 'Wando' | 5 | 4 | 7 | 6/17 | 37 |
| Pumpkin; Cucurbita pepo 'Connecticut field' | 4 | 4 | 2 | 3/10 | 30 |
| Pumpkin; C. pepo 'National Pickling' | 5 | 6 | 4 | 8/15 | 53 |
| Solanum lycopersicum 'Bradly' | 4 | 4 | 7 | 6/15 | 40 |
| Spinach 'Meerkat' | 4 | 4 | 5 | 3/13 | 23 |
| Vigna unguiculata 'Monarch' | 5 | 3 | 4 | 0/12 | 0 |

*PCR = polymerase chain reaction. All positive amplicons were confirmed by direct sequencing of the product.
contains key badnavirus motifs, including those involved in movement (Arg128 to Pro189), the coat protein zinc-finger domain (Cys867 to Pro901), the pepsin-like aspartate protease (Val1151 to Gly1229), reverse transcription (Leu1331 to Gly1550), and RNase H (Thr1652 to Leu1773).

P3 is most similar to FBV-1, CIVMV, and cacao swollen shoot virus (CSSV) orthologs with 47, 45, and 44% aa identities, respectively. ORF4 (ATG6694 to TGA7125) overlaps ORF3 and encodes a putative protein (p4) of unknown function of 143 aa with a molecular mass of 16 kDa.

The relationship of BVF and other badnaviruses was determined based on the complete nucleotide sequence of their genomes (unpublished data). BVF clusters with Commelina yellow mottle virus, Dioscorea bacilliform virus (DBV), CSSV, FBV-1, and CIVMV (Fig. 2) and shows less than 80% nucleotide identity to any member of the genus in the RT/RNase H domain, the demarcation criteria for badnavirus species (Geering and Hull 2012), suggesting that BVF is a putative new member of the genus.

Recombination analysis carried out with RDP4 identified a recombination event involving a CSSV-like virus as the major parent and a DBV-like virus as the minor parent, with the recombination breakpoints mapping at nucleotides 1,726 and 2,174 of the BVF genome (Table 2).

Host range. In all, 14 of the 20 plant species used in mechanical transmission trials sustained virus replication, as revealed by PCR performed on nucleic acids purified from apical leaves 1 month postinoculation, as well as additional RCA and Southern hybridization experiments (Table 1; Fig. 3). None of the infected plants displayed visible symptoms during the month-and-a-half-long postinoculation observation period.

Distribution, population structure and detection. Primer set 1562F/2029R gave consistent results, providing a useful, efficient tool for virus detection and allowing for multiplexing with an internal control in either PCR or RT-PCR (Fig. 4). Extensive screening of 403 BYVD-infected blackberry revealed the presence of BVF in samples collected from several states: Arkansas (n = 5), Georgia (n = 6), Mississippi (n = 9), North Carolina (n = 2), and South Carolina (n = 1). Those samples were evaluated as part of the population structure study to identify potential patterns that would point to the geographic

---

Fig. 1. Schematic representation of the blackberry virus F (BVF) genome showing relative positions of RNA binding site TATA box (denoted by box); open reading frame (ORF) 1; ORF2; ORF3, with movement protein (MP), capsid protein zinc-finger domain (CP), pepsin-like aspartate protease (Pro), reverse transcription (RT), and RNase H (RNase H) motifs; and ORF4.

Fig. 2. Phylogenetic analysis based on the complete nucleotide sequences of blackberry virus F (BVF) and other badnaviruses. Numbers at nodes represent percent (%) bootstrap confidence score. Virus names and acronyms: Taro bacilliform virus (TaBV); Bougainvillea spectabilis chlorotic vein banding virus (BSCVBaV); Commelina yellow mottle virus (CoYMV); Dioscorea bacilliform virus (DBV); cacao swollen shoot virus (CSSV); fig badnavirus-1 (FBV-1); citrus yellow mosaic virus (CIVMV); sugarcane bacilliform IM virus (BSBIMV); sugarcane bacilliform MO virus (BSMOV); banana streak Ul virus (BSULV); banana streak UL virus (BSULV); banana streak MY virus (BSMYV); pineapple bacilliform comosus virus (PBCoV); banana streak GF virus (BSGFV); Kalanchee top spotting virus (KTSV); banana streak AV virus (BSAVY); banana streak VN virus (BSVNV); banana streak UL virus (BSULV); goosberry vein banding virus (GVBaV), and banana streak UA virus (BSUAO). Cauliflower mosaic virus (CaMV), a member of the genus Caulimovirus, was used as the outgroup. The bar represents 0.1 nucleotide substitution/site.
Virus integration. Virus characterization and phylogenetic analysis placed BVF in the genus Badnavirus, suggesting that the virus may integrate into the host genome. In order to verify possible integration into a host genome, RT-PCR, RCA, and Southern blotting experiments were employed.

The majority of the PCR-positive samples failed to amplify viral RNA after DNase digestion, suggesting that those plants only have the inactive, integrated form of the virus. On the other hand, it was evident that BVF is also found in the active, episomal form in some studied plants because the virus was readily transmissible onto herbaceous plants from the sap extracted from blackberry MS-469 (Table 1; Fig. 3).

Although RCA was successfully performed on nucleic acids extracted from experimental host plants, it was never successful on blackberry in either of our two laboratories using different extraction protocols and several dilution of the template, including blackberry nucleic acids spiked with a plasmid control (data not shown). We speculate that this is because of an unidentified inhibitor of the RCA reaction. Southern hybridization was employed as an alternative (Fig. 5). Because the BVF genome has a single KpnI restriction site, signals corresponding to the full genomic size (approximately 7.6 kb) were visible in a PCR/RT-PCR-positive blackberry plant as well as an indicator species (N. rustica). The PCR/RT-PCR-positive blackberry displayed additional signals, possibly part of an integrated form of the virus. The PCR-positive/RT-PCR-negative blackberry displayed multiple bands that collectively exceeded the genome size of BVF (Fig. 5). The hybridization performed on SphI-digested samples essentially confirmed the results obtained for KpnI (Fig. 5). The possibility that isolate variation and mixed infection with variants or defective molecules cannot be disregarded but, given the accumulated evidence, it is realistic to assume that the virus integrates in the blackberry genome.

Discussion

In this work, we identified and characterized a new virus provisionally named BVF. BVF is a virus with genome size and structure similar to that of badnaviruses (Geering and Hull 2012). Sequence analysis of both integrated and episomal isolates collected from distant geographical areas revealed a relatively homogenous virus population, consistent with a recent emergence or introduction of the virus and a low error-rate replication, or due to the fact that few variants are circulating through the nursery system. The virus population in the southeastern United States does not display significant genetic drift, one of the primary indicators of multiple host species (Bedhomme et al. 2012; Rico et al. 2006). In the geographic areas sampled, BVF likely moves passively via infected propagation material or actively via vectors, a hypothesis that will be tested after the identification of such carriers.

A recombination event was identified (Table 2) which, in combination with the population structure of the virus in the southeastern United States, indicates that BVF evolution is possibly shaped through recombination, the major source of variation and evolution for members of the Caulimoviridae family (Hull et al. 2000). The chance of recombination is rather elevated in a perennial crop such as blackberry, where mixed virus infections are very common, as observed for several other viruses infecting this host (Poudel et al. 2012; Thekke-Veetil et al. 2013).

Badnaviruses are important pathogens worldwide, not only due to the disease they may cause in an active, episomal form but also because of their integration into the host genome (Daniells et al. 2001; Geering et al. 2001; Ndowora et al. 1999). This study provides evidence that BVF actively replicates in some of the blackberry plants used in this study (i.e., MS-469), as well as in several additional plant species upon artificial inoculation with the virus. The virus infected less than 5% of samples tested, suggesting, at best, a marginal role in BYVD etiology. Further studies are needed to understand its effects in mixed infections with other viruses. It is difficult to predict what effect BVF will have in the blackberry germplasm. None of the infected plants were solely infected by BVF (data not shown), making the evaluation of the virus effects in a single infection impossible at this point. However, its occurrence may aggravate symptoms, either directly by causing new diseases or indirectly by interacting with viruses already present in the plants, leading to plant decline, as has been the case with other viruses which are part of the BYVD complex (Martin et al. 2013). Nevertheless, despite relatively low overall incidence of BVF in tested blackberry samples, it is important to point out that the virus was found in several states, indicating a wide geographic distribution. Whether this is due to an active transmission by an as-yet-identified vector or to the dissemination of infected propagation material is still to be determined. However, the most important finding of this study is the fact that BVF integrates in the blackberry genome, with origin of the virus and use conserved sequences to develop reliable detection tests. Diversity was minimal because isolates shared more than 97% identities at the nucleotide level (GenBank accessions KT698919 to -41; unpublished data).

Table 2. Detection of a recombination event in blackberry virus F genome supported by the majority of the RDP4 implemented methodsa

<table>
<thead>
<tr>
<th>Major parent</th>
<th>Minor parent</th>
<th>RDP</th>
<th>Bootscan</th>
<th>Maxchi</th>
<th>Chimera</th>
<th>SiScan</th>
<th>PhylPro</th>
<th>3Seq</th>
<th>GENECONV</th>
<th>LARD</th>
<th>Beginb</th>
<th>Endb</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSSV</td>
<td>DBV</td>
<td>2.88E-09</td>
<td>4.16E-05</td>
<td>1.43E-03</td>
<td>4.43E-04</td>
<td>5.98E-15</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>1,726</td>
<td>2,174</td>
</tr>
<tr>
<td>(NC_001574)</td>
<td>(NC_009010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a NS = nonsignificant.

*b Break points.

Fig. 3. A, Rolling circle amplification (RCA) digested with KpnI (single restriction site in the blackberry virus F [BVF] genome) and B, hybridization on the RCA products of BVF-infected indicator plants. RCA products were digested with SacI (single restriction site in the BVF genome) or SphI (two restriction sites in the BVF genome). Lanes 1 to 8 represent a RCA-positive plant from each of the species presented in A; lanes 9 and 10 are BVF-free Nicotiana rustica and N. tabacum, respectively.
practical implications for the blackberry industry in the United States and globally.

Virus integration makes its eradication or elimination from infected plants unfeasible. It may be possible to eliminate the BVF episomal form through thermo- and chemotherapeutic techniques, based on previous experience with Rubus RNA viruses (Wang et al. 2008), but there is the possibility of virus activation given the experience with other badnaviruses and related pararetroviruses (Dallot et al. 2001; Lockhart et al. 2000; Ndowora et al. 1999; Pahalawatta et al. 2008; Richert-Poggeler et al. 2003). It is important to stress that routine in vitro culture propagation, which is the basic technique for virus-free planting material, greatly influences activation of endogenous sequences of banana streak virus (BSV) in Musa spp. (Dallot et al. 2001). Due to possible genome activation (as in the case of Musa spp. and BSV), it is of utmost importance to carefully select BVF- and possibly virus-free material, as drafted in the United States National Pathogen-Tested Certification Program for Caneberry Nursery Stock Production (www.ncpberries.org). It is important to be aware of peculiar problems associated with BVF infections and eliminate infected material from the propagation pipelines. The detection protocol developed in this study is sensitive and universal, allowing for detection of both BVF forms in field or nursery settings, a crucial step in identifying BVF-free material for sanitation.

In summary, we have identified and characterized a novel virus from blackberry, a prototype of a putative new species of the genus Badnavirus, broadening knowledge on the array of viruses infecting the crop. In addition to its mere scientific value, this discovery will have a great impact on production of virus-free planting material for the berry industry in the United States and elsewhere.

Acknowledgments

We thank T. Ho for assistance in the initial bioinformatics analysis. This work was supported, at least in part, by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture-Specialty Crops Research Initiative (2009-51181-06022), USDA National Clean Plant Network (14-8130-0420-CA), and Hatch projects 1002361 and MBS-112010.

Literature Cited


