Molecular characterization and population structure of blackberry vein banding associated virus, new ampelovirus associated with yellow vein disease

Thanuja Thekke-Veetil a, Nina Aboughanem-Sabanadzovic b, Karen E. Keller c, Robert R. Martin c, Sead Sabanadzovic d, e, Ioannis E. Tzanetakis a, **

a Department of Plant Pathology, Division of Agriculture, University of Arkansas System, Fayetteville, AR 72701, USA
b Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA
c USDA-ARS, Corvallis, OR 97330, USA
d Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA

ARTICLE INFO

Article history:
Received 22 June 2013
Received in revised form 27 September 2013
Accepted 29 September 2013
Available online 11 October 2013

Keywords:
Blackberry
Ampelovirus
Genome
Diversity
Variation and population structure

ABSTRACT

Blackberry yellow vein disease is the most important viral disease of blackberry in the United States. Experiments were conducted to characterize a new virus identified in symptomatic plants. Molecular analysis revealed a genome organization resembling Grapevine leafroll-associated virus 3, the type species of the genus Ampelovirus in the family Closteroviridae. The genome of the virus, provisionally named blackberry vein banding associated virus (BVBaV), consists of 18,643 nucleotides and contains 10 open reading frames (ORFs). These ORFs encode closterovirid signature replication-associated and quintuple gene block proteins, as well as four additional proteins of unknown function. Phylogenetic analyses of taxonomically relevant products consistently placed BVBaV in the same cluster with GLRaV-3 and other members of the subgroup I of the genus Ampelovirus. The virus population structure in the U.S. was studied using the replication associated polypeptide 1a, heat shock 70 homolog and minor coat proteins of 25 isolates. This study revealed significant intra-species variation without any clustering among isolates based on their geographic origin. Further analyses indicated that these proteins are under stringent purifying selections. High genetic variability and incongruent clustering of isolates suggested the possible involvement of recombination in the evolution of BVBaV.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Blackberry yellow vein disease (BYVD) is the most serious and complex viral disease of blackberry in the United States. The disease occurs in both cultivated and wild blackberries, and has been reported in the Southeastern United States and California (Martin et al., 2013). Symptoms include oak-leaf patterns, irregular chlorosis, line patterns, ringspots, yellowing of veins, and malformed leaves (Susaimuthu et al., 2006, 2007) and results in a rapid decline in productivity and longevity of plants (Tzanetakis, 2012). Several viruses; blackberry yellow vein-associated virus, blackberry virus Y, beet pseudo-yellows virus, blackberry chlorotic ringspot virus, blackberry virus E, blackberry virus S, impatiens necrotic spot virus, and tobacco ringspot virus, have often been found associated with the BYVD complexes (Martin et al., 2013). Symptoms appear only in mixed infections of two or more of these viruses. Plants infected with multiple viruses often exhibit identical symptoms with severity directly associated with the number of viruses in plants (Martin et al., 2013). A new ampelovirus (family Closteroviridae) provisionally named blackberry vein banding associated virus (BVBaV) was first identified in symptomatic plants in Mississippi (Fig. 1; Sabanadzovic et al., 2011) and is the subject of this communication.

Members of Closteroviridae are important pathogens in several horticultural crops. The family comprises viruses with relatively large, single-stranded, positive-sense, RNA genomes that range from 13 to 20 kb in size and are either monopartite or segmented. Viruses of this family belong to four genera; Closterovirus, Ampelovirus, Velarivirus and Crinivirus, based on the differences in genome size, type, organization, biological and epidemiological properties (Martelli and Candresse, 2010; Martelli et al., 2012a, b). Although the genome organization varies among the members of the family, all closterovirids encode replication associated proteins in open reading frames (ORFs) 1a and 1b and have a conserved quintuple gene block involved in movement and encapsidation.

---

* Corresponding author.
** Corresponding author. Tel.: +1 479 575 3180; fax: +1 479 575 7601.
E-mail addresses: SSabanadzovic@entomology.msstate.edu (S. Sabanadzovic), itzaneta@uark.edu, ytzan@yahoo.com (I.E. Tzanetakis).

0168-1702/5 – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.virusres.2013.09.039
In the genus Ampelovirus there are two types of genome organizations: subgroup I, represented by grapevine leafroll-associated virus 3 (GLRaV-3), includes members with larger genomes that contain 12–13 ORFs whereas subgroup II, represented by pine apple mealybug associated virus 1, comprises viruses with smaller genomes of 7 ORFs, lacking the diverged copy of the coat protein. Closteroviruses, ampeloviruses and velarviruses have monopartite genomes. Closteroviruses are transmitted by aphids, ampeloviruses with mealybugs or scale insects whereas there are not known vectors for velarviruses. Most of the criniviruses have bipartite genomes and are transmitted by whiteflies.

Closteroviruses are reported to have highly diverse genetic populations (Dolja et al., 2006; Karasev, 2000). Replication of RNA genomes is an error prone process due to the lack of a proof reading mechanism in the RNA-dependent RNA polymerase (RdRp; Domingo, 1997). This and other factors, such as recombination and selection pressure, play significant roles in evolution and shaping of the population structure of RNA viruses. Knowledge on population structure and the extent of intra-species variation are essential to better understand virus dynamics and develop appropriate detection and control tools. Also, the evolutionary potential of a virus is an important factor to be considered when employing control methods through development of resistant cultivars. Few studies have been reported regarding the genetic variation of ampeloviruses and those were concentrated on the RdRp, major coat protein (CP) and heat shock protein 70 homolog (HSP70h) (Fajardo et al., 2001; Habili et al., 1995; Saldarelli et al., 1998; Turturo et al., 2005; Wang et al., 2011). The replication-associated polyprotein contains conserved domains of papain-like protease (PR), methyltransferase (MTR) and helicase (HEL) whereas the intervening region is highly variable (Dolja et al., 2006). Also, the diverged copy of the coat protein (minor coat protein, Cpm) of GLRaV-1, was reported to be highly variable compared to the CP (Alabi et al., 2011; Little et al., 2001). Thus, a diversity analysis on BVBaV polyprotein and Cpm sequences in addition to HSP70h, the hallmark gene of Closteroviridae, was conducted and the results are reported here. The information obtained in this study contributes to the existing knowledge on blackberry viruses and ampeloviruses in general and gives insight into the forces driving evolution of this diverse group of plant viruses.

Fig. 1. Symptoms observed on the original source of blackberry vein banding associated virus. The plant was also infected with at least two other viruses.

2. Materials and methods

2.1. Molecular characterization

Double-stranded RNA from infected plants was extracted using a protocol involving selective chromatography (Valverde, 1990), and used as a template for virus characterization. Initial virus sequences were obtained by cloning randomly amplified cDNA fragments combined with Illumina next-generation sequencing as previously described (Laney et al., 2011; Sabanadzovic and Valverde, 2011). Gaps between contigs were filled by RT-PCR. The 5’ end of the genome was obtained using 5’ RACE (Life Sciences) according to manufacturer’s instructions. Terminal sequences at the 3’end of the virus genome were determined by RT-PCR, performed on cDNAs generated on artificially polyadenylated dsRNAs as previously described (Abou-Ghanem et al., 1998). Nucleotide sequences of the same region were further verified with the protocol involving ligation of 5’phosphorylated/3’ amino-blocked oligonucleotide to target dsRNAs (Lambden et al., 1992).

Genome sequences were analyzed with ORF finder (Wheeler et al., 2006) and FGENESV0 software (Anonymous, 2007) to identify the ORFs. Conserved domains of the encoded proteins were identified with the conserved domain database (CDD; Marchler-Bauer et al., 2013). The genome sequence was deposited in GenBank under accession no. KC904540.

Phylogenetic analyses were performed on deduced amino acid sequences of the RdRp and HSP70h proteins of BVBaV and 26 recognized closterovirids. Both datasets, initially aligned by MUSCLE (Edgar, 2004), were submitted to phylogenetic estimations with Maximum Likelihood (ML) and Neighbor Joining (NJ) methods. ML-based phylogenies were inferred with PhyML (Guindon and Gascuel, 2003) implemented in SeaView version 4.4 (Gouy et al., 2010) using the LG substitution model (Le and Gascuel, 2008), whereas MEGA v.5 (Tamura et al., 2011) was used to infer phylogenies applying the Neighbor-Joining method with 1000 bootstrap replicates. Amino acid identities of the conserved proteins of the BVBaV was obtained by pairwise comparisons of the polyprotein, RdRp, HSP70h, HSP90h, CP, Cpm and 21 kDa proteins with those of the subgroup I orthologs.
2.2. Population structure study

Twenty-five BVBaV isolates were identified in a survey at the early stages of this project using primer set AmpF (5’-CATGAGTGTGGATGGTAAYAACGG-3’) – AmpR (5’-AACCTTTCCGGGTCGTCGTTA-3’) and used for the population structure study. These plants were collected from several states between 2008 and 2011 (Supplementary Table 1).

Total nucleic acids were extracted from 50 mg tissue using the procedure described for woody plants, and reverse transcribed using random primer as described (Poudel et al., 2013). The cDNAs obtained were diluted twice with water to minimize the effect of potential inhibitors in downstream reactions.

Primers were designed to amplify a 1200 nucleotide (nt) region in the ORFs of the polyprotein (region between MTR and HEL domains (genome position 3941–5140), HSP70h (genome position 11,159–12,358), and CPM (genome position 15,307–16,506) genes using Phire Hot Start II DNA Polymerase (New England Biolabs, MA) by two-step PCR. Due to sequence variability, several sets of primers (Supplementary Table 2) were used to amplify the regions described. The samples were subjected to an initial denaturation at 98 °C for 30 s, followed by 35 cycles of denaturation at 98 °C for 5 s and annealing and extension at 72 °C for 30 s, and a final extension at 72 °C for 1 min. The PCR products were visualized on 1% TBE-agarose gel stained with GelRed® (Biotium Inc., CA).

The amplified products were purified, cloned, and sequenced as described in the Poude1 et al. (2012) study. Sequences were edited using the Sequence Scanner Software v1.0 (Applied Biosystems, CA), assembled by CAP3 (Huang and Madan, 1999) and deposited in GenBank under accession nos. KC912140–KC912189 and KC904515–KC904539.

Nucleotide sequences were aligned by ClustalW and variation in nt and predicted amino acid (aa) sequences of virus isolates were determined using BioEdit (Hall, 1999). The phylogenetic trees were constructed using MEGA v.5 using the Neighbor-Joining method with 1000 bootstrap replicates to determine the robustness of the phylogenetic grouping.

The selection pressure on the studied areas was estimated by calculating the $d_{ns}/d_s$ ratio, which is an indicator of natural selection, where $d_{ns}$ represents the average number of non-synonymous substitution per non-synonymous site and $d_s$ represents the average number of synonymous substitution per synonymous site. The $d_{ns}/d_s$ ratio was estimated using the Synonymous Non-synonymous Analysis Program (SNAP; Korber, 2000) which performs pairwise analysis of the $d_{ns}$ and $d_s$ values and also the average for a particular protein using the Nei–Gojobori method (Nei and Gojobori, 1986). Additionally, sequences were analyzed using the codon based Z-test in MEGA v.5 that also uses the Nei–Gojobori method.

Recombination events were investigated by RDP4 (Martin et al., 2010) by employing the RDP, GENECONV, Bootscans, Maxchi, Chimera, SiScan, and 3Seq algorithms. Analyses were conducted on the polyprotein, HSP70h and CPM regions separately and also after concatenating the sequences of all three regions in one. Only events supported by at least six different RDP-4 implemented methods were considered possible recombinations.

3. Results

3.1. Genome analysis

The BVBaV genome consists of 18,643 nt and resembles the organization of members of subgroup I of the genus Ampelovirus (Martelli et al., 2012a,b). The genome starts with a 781 nt-long 5’ untranslated region (UTR) rich in A+U content (63%). The virus shares 17 of the 25 first nucleotides of 5’ UTR with GLRaV-3, the type member of the genus.

ORF 1a (nt 782–7936), codes for a putative 2384 aa-long replication-associated polyprotein. This protein contains hallmark domains of (in N-to-C direction): PRO with conserved cysteine and histidine residues at positions 343 and 389, similar to GLRaV-3 and PMWaV-2 (Ling et al., 2004; Melzer et al., 2001); MTR (aa residues from 501 to 812), DNA alkali damage repair protein (ALKB; aa residues from 1712 to 1827), and HEL superfamily 1 (aa residues from 2085 to 2344) (Marchler-Bauer et al., 2013). ORF 1b contains conserved motifs of viral RNA-dependent RNA polymerases (RdRp,2 superfamily, pfam00978) which is putatively expressed as a fusion protein with ORF1a via +1 ribosomal frameshift mechanism. The replication-associated protein and RdRp show 53% and 62% identity respectively with their orthologs in GLRaV-3 (Table 1).

The central portion of the genome is characterized by a 1079 nt long intergenic region and the presence of two small ORFs (ORFs 2 and 3) between nt regions 10,646–10,798 and 10,882–11,016 respectively. These ORFs encode 50 aa (p6) and 44 aa (p5) proteins respectively with transmembrane domains between aa residues 19–41 and 5–27 for p6 and p5 respectively, as revealed by in silico predictions (TMHMM; Krogh et al., 2001).

ORF 4 (nt 11,054–12,673) codes for a putative HSP70h protein with an estimated Mr 57.8 kDa followed by ORF 5 (nt 12,683–14,122) encoding a putative, 479 aa-long, CP homolog (Napuli et al., 2003). The two proteins share 54% and 27% identity with the GLRaV-3 orthologs. ORFs 6 and 7 encode putative proteins of ca. 34 kDa and 54 kDa respectively, identified as the virus CP and CPM, as they contain conserved motifs of the closterovirus coat protein superfamily (pfam01785) near the C-termini. The CP of this virus shows 53% identity to the GLRaV-3 ortholog whereas the CPM exhibits 34% identity to its GLRaV-3 counterpart.

The coding region of BVBaV terminates with three ORFs (ORFs 8–10) encoding putative polypeptides with comparable sizes (185, 170 and 184 aa, respectively) and unknown functions. Whereas the product of ORF 8 shares limited homology with the GLRaV-3 21 kDa protein (38%), the other two proteins do not have significant similarity with currently available viral proteins deposited in the GenBank. These ORFs are followed by a 335 nt long 3’ UTR. No small 3’–proximal ORFs, comparable with those reported in GLRaV-3 (i.e. 4 kDa and 7 kDa) were present in the BVBaV genome (Fig. 2).

Independently of the protein used for pairwise comparisons, putative products encoded by the BVBaV genome always shared the highest levels of common aa content with counterparts encoded by GLRaV-3. These identities ranged from 27% in the case of the CP homolog to 62% in case of the RdRp.

Phylogenetic analyses based on Maximum Likelihood or Neighbor-Joining method, performed on RdRp and HSP70h

<table>
<thead>
<tr>
<th>Proteins</th>
<th>GLRaV-1</th>
<th>GLRaV-3</th>
<th>PMWaV-2</th>
<th>LCHV-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyprotein</td>
<td>39</td>
<td>53</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>RdRp</td>
<td>47</td>
<td>62</td>
<td>47</td>
<td>37</td>
</tr>
<tr>
<td>HSP70h</td>
<td>40</td>
<td>54</td>
<td>49</td>
<td>32</td>
</tr>
<tr>
<td>HSP90h</td>
<td>16</td>
<td>27</td>
<td>26</td>
<td>14</td>
</tr>
<tr>
<td>CP</td>
<td>28</td>
<td>53</td>
<td>37</td>
<td>16</td>
</tr>
<tr>
<td>21 kDa</td>
<td>7</td>
<td>38</td>
<td>NA</td>
<td>13</td>
</tr>
</tbody>
</table>

 Identities with two CPM proteins putatively encoded by GLRaV-1 genome.
3.2. Population diversity study

Screening of survey samples collected from southern U.S. indicated the presence of BVaV in five states; Arkansas (AR), Georgia (GA), Mississippi (MS), North Carolina (NC), and South Carolina (SC). Out of the 319 blackberry samples screened, BVaV was detected in 49 samples. Sequence analysis of the coding regions of polyprotein, HSP70h and CPM proteins under study represented approximately 20% of the total genome and these regions revealed significant diversity among the isolates studied. The polyprotein region was most divergent compared to the other two genes, with as much as 23% diversity in both the nt and aa (Supplementary Tables 3 and 4). For the HSP70h, identity among isolates varied between 88–100% in the nt and 95–100% in the predicted aa sequences (Supplementary Tables 5 and 6). The CPM nt sequences had 86–100 nt identities compared to the 91–100% aa identity (Supplementary Tables 7 and 8).

No clustering of isolates with respect to geographical areas was observed in the case of the polyprotein and HSP70h sequences, whereas for the CPM region all GA isolates grouped in to a separate clade except one (GA2; Fig. 4). With respect to the genes under study, the isolates did not exhibit a particular pattern in grouping. Interestingly, isolates from the same geographical location exhibited considerable variation in the nt and aa sequences. For

![Fig. 2. Schematic representation of blackberry vein banding associated virus genome compared with that of Grapevine leafroll-associated virus 3 (GLRaV-3), the type species of the genus Ampelovirus. Boxes show positions of different open reading frames with respective products, whereas lines depict untranslated genomic regions. The same color denotes orthologs in two genomes. Gray boxes denote proteins of unknown function. Black boxes indicate small 3′-co-terminal ORFs present only in GLRaV-3 genome. Abbreviations: PRO, MTR, AlkB, Hel: protease, methyltransferase, AlkB and helicase domains of replicase polyprotein; RdRp: RNA dependent RNA polymerase; p5: 5 kDa protein; p6: 6 kDa protein; HSP70h: heat shock protein 70 homolog; p53: 53 kDa protein; CP: coat protein; CPM: minor CP; p21: 21 kDa protein; p19: 19 kDa protein; p20: 20 kDa protein.](image)

![Fig. 3. Maximum Likelihood-inferred phylogenetic relationships among members of the Closteroviridae based on amino acid sequences of RNA-dependent RNA polymerase (RdRp) and heat shock 70 homologue (HSP70h) (panels A and B, respectively). In both analyzed datasets blackberry vein banding associated virus clusters with members of the genus Ampelovirus. Maximum Likelihood method was implemented with PhyML. Percentage (%) bootstrap support values for the genera in the family Closteroviridae (encircled) and for BVaV (reported in red) are indicated at the branching points. The GenBank accession numbers of viral genomes used to deduce amino acid sequences of RdRp and HSP70h proteins are: beet pseudo-yellows virus (BPYV; NC_005209.2, NC_005210.2), beet yellow stunt virus (BYSV; U51931.1), beet yellows virus (BYV; NC_001598.1), blackberry vein banding-associated virus (BVaV; KC904540), blackberry yellow vein-associated virus (BVaV; NC_006962.2, NC_006963.2), citrus tristeza virus (CTV; NC_001661.1), cordyline virus 1 (CoV-1; HMs88723.1), cucurbit yellow stunt disorder virus (CYSDV; NC_004809.1, NC_004810.1), grapevine leafroll-associated virus 1 (GLRaV-1; NC_010509.1), grapevine leafroll-associated virus 2 (GLRaV-2; NC_007448.1), grapevine leafroll-associated virus 3 (GLRaV-3; NC_004667.1), grapevine leafroll-associated virus 4 (GLRaV-4; NC_016146.1), grapevine leafroll-associated virus 7 (GLRaV-7; NC_016436.1), lettuce chlorosis virus (LCV; NC_012909.1, NC_012910.1), lettuce infectious yellow virus (LIVY; NC_003617.1, NC_003618.1), little cherry virus 1 (LCV-1; NC_001836.1), little cherry virus 2 (LCV-2; NC_005065.1), mint vein banding-associated virus (MVBAV; AY548173.3), mint virus 1 (MV-1; NC_006944.1), pineapple mealybug wilt-associated virus 1 (PMaV-1; NC_010178.1), pineapple mealybug wilt-associated virus 2 (PMaV-2; AF83103.1), pineapple mealybug wilt-associated virus 3 (PMaV-3; DQ339239.2), potato yellow vein virus (PYVV; NC_006863.1), raspberry leaf mottle virus (RLMoV; NC_008585.1), strawberry chlorotic fleck-associated virus (SCFaV; NC_008366.1), tomato chlorosis virus (ToCV; NC_007340.1, NC_007341.1), tomato infectious chlorosis virus (TICV; NC_013258.1, NC_013259.1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)](image)
examples, three isolates from GA, (GA1–3) showed significant variation (22–23%) in the polyprotein when compared to other Georgian isolates (GA 4–7; Supplementary data), whereas in the case of CPM, isolates from NC, NC 9–12 varied significantly in nt (13–14%) and aa sequences (7–8%) from other NC isolates (Supplementary Tables 7 and 8). The clustering of isolates in dendograms derived from the aa sequences showed that few of the GA isolates are significantly different compared to the rest of the isolates in the polyprotein and CPM (GA1–GA3 in polyprotein and GA2 in CPM) regions.

The $d_{ns}$ and $d_s$ were calculated for all the three regions under study to assess the selection pressure operating on these proteins (Table 2). The $d_{ns}/d_s$ ratio was highest for the polyprotein region (0.14) compared to the HSP70h (0.032) and CPM (0.066). No recombination events were detected when nucleotide sequences of the three genes were analyzed with the RDP4 software. Analysis of recombination events in concatenated polyprotein-HSP70h-CPM regions detected three possible recombination events which were supported by significant $p$ values and at least six programs used in RDP4 (Table 3). The putative recombination site for the isolate GA2, a recombinant of GA1-like and GA7-like sequences, was predicted to be between the polyprotein and HSP70h regions. For the region between HSP70h and CPM, isolate GA1 was predicted to be a possible recombinant of a GA2-like and a GA4-like isolates, whereas the North Carolina isolate, NC12, was predicted to be a recombinant of a GA4-like and an isolate of unknown origin.

### 4. Discussion

The BVBaV is a newly identified virus associated with the BYDV complex with the genome size, structure and organization comparable to that of recognized members of the genus *Ampelovirus* (Martelli et al. 2012a,b). BVBaV genome size (18,643 nt) makes it, along with *citrus tristeza virus* (CTV ∼19.3 kb; Karasev et al., 1995) and GLRaV-3 (18.5–18.7 kb; Maree et al., 2013), one of the largest viral genomes made up of a single molecule of single-stranded, positive-sense RNA. BVBaV genome is characterized by a 781 nt

### Table 2

<table>
<thead>
<tr>
<th>Region</th>
<th>$d_{ns}$</th>
<th>Variance ($d_{ns}$)</th>
<th>Standard deviation ($d_{ns}$)</th>
<th>$d_s$</th>
<th>Variance ($d_s$)</th>
<th>Standard deviation ($d_s$)</th>
<th>$d_{ns}/d_s$</th>
<th>Nucleotide identity percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPM</td>
<td>0.019</td>
<td>0.000</td>
<td>0.007</td>
<td>0.287</td>
<td>0.001</td>
<td>0.001</td>
<td>0.034</td>
<td>0.066</td>
</tr>
<tr>
<td>HSP70h</td>
<td>0.002</td>
<td>0.000</td>
<td>0.001</td>
<td>0.062</td>
<td>0.000</td>
<td>0.000</td>
<td>0.007</td>
<td>0.032</td>
</tr>
<tr>
<td>Polyprotein</td>
<td>0.030</td>
<td>0.000</td>
<td>0.009</td>
<td>0.215</td>
<td>0.001</td>
<td>0.027</td>
<td>0.140</td>
<td>77–100</td>
</tr>
</tbody>
</table>

$d_{ns}$ – average number of non-synonymous substitution per non-synonymous site. $d_s$ – average number of synonymous substitution per synonymous site. $d_{ns}$ and $d_s$ and their variation and standard deviations are computed using the Synonymous Non-synonymous Analysis Program (SNAP; Korber, 2000) which compute the $d_{ns}$ and $d_s$ values for a particular protein using Nei–Gojobori method.
long 5' non-coding region (NCR) preceding the first ORF. A similar feature (long 5'NCR) has been previously reported for GLRaV-3 genome (Maree et al., 2008, 2013), but not in other viruses belonging to the family. The organization of first seven ORFs resembles that of GLRaV-3 and ampeloviruses in general: ORF1a/b coding for the conserved replication module (PRO-MTR-AIkB-HEL-POL), followed by the quintuple gene block proteins involved in virus movement and assembly (Dolja et al., 2006). Furthermore, the p21 protein, encoded by ORF 8 in both viruses, is more conserved than CPm. Similar to GLRaV-3 (Ling et al., 2004), BVaV contains two small putative ORFs coding for p5 and p6 proteins downstream from the replication module but unlike its counterpart which codes for two short peptides (p4 and p7) at the 3' end of the genome, the two BVaV 3-terminal proteins are larger (p19 and p20; Fig. 2) and do not show statistically significant similarities with proteins currently deposited in GenBank. Another point of distinction is the location of a large intergenic region: The BVaV intergenic region is located between ORF1b and the two small ORFs (ORF2), whereas in case of GLRaV-3 this region is between ORFs 2 and 3 (Fig. 2).

Similarities with ampeloviruses, in particular with GLRaV-3, were confirmed by phylogenetic analyses: BVaV and GLRaV-3 grouped as sister clades in trees constructed on RdRp and HSP70h amino acid sequences. This grouping was consistent and independent of the dataset used (RdRp or HSP70h) or algorithm (ML or NJ) applied and was supported by high bootstrap values. However, identities with GLRaV-3 and related viruses in all relevant genomic products (RdRp, HSP70h and CP) was far below (Table 1) the species demarcation thresholds (75% aa identity) proposed by the Closteroviridae Study Group of the International Committee on Taxonomy of Viruses (ICTV), indicating that BVaV is a novel species in the taxon. Furthermore, this study also revealed considerable intra-species variations among 25 BVaV isolates, collected from distant geographical regions in the U.S., in the polyprotein, HSP70h and CPm genes. Previous studies have also shown varying levels of nucleotide sequence diversity among closteroviruses depending on the genomic region or genes selected for analysis (Lopez et al., 1998; Wang et al., 2011). The polyprotein region in BVaV was more divergent than the other two genes, with up to 23% diversity in the nt and aa levels. Genetic variation is the basis of virus adaptation and is important for the survival and competitiveness of RNA viruses. The high rate of sequence variability observed in BVaV could indicate high adaptive potential of the virus to diverse geographical areas, vectors or hosts.

Replication of RNA genomes are error prone and closteroviruses have relatively large genomes. Plant RNA viruses especially those infecting clonally propagated perennial crops can accrue considerable mutations over time which would explain the high variation rate observed among the isolates of this virus. Also, BVaV isolates did not exhibit clustering based on geographical origins except for the Georgian isolates for the CPm sequences. Similar to our results, lack of geographical isolation among isolates was reported in GLRaV-3 (Fuchs et al., 2009; Turturo et al., 2005; Wang et al., 2011). This could probably be due to the fact that these isolates may have been disseminated to distant geographical areas through infected blackberry cuttings or because of recombination events that may have taken place in areas of the genome that were not studied here.

Coding areas are under positive or diversifying selection when \( d_{NS}/d_S \) ratio is >1, neutral selection when \( d_{NS}/d_S \) ratio=1 and negative (purifying) selection when \( d_{NS}/d_S \) ratio <1. Although significant genetic variation was observed among isolates in all three genes, the \( d_{NS}/d_S \) ratio, the estimate of selective pressure operating on these genes, was lower than expected if these variations were restricted to synonymous positions and the regions were under strict purifying or negative selection to conserve their amino acids sequences. The strongest constraint was on HSP70h (Table 2). The HSP70h protein encoded by closteroviruses is multifunctional and is involved in virion assembly (Satyanarayana et al., 2004) and cell-to-cell movement of virus particles (Peremyslov et al., 1999) which may explain the pressure on conserving the amino acid sequences of that gene compared to other two regions under study. Relatively high \( d_{NS}/d_S \) ratio of the polyprotein compared to HSP70h and CPm suggested that the region between MTR and HEL domains in the polyprotein ORF can tolerate non-synonymous substitutions and may have a significant role in the adaptive evolution of this gene.

Recombination is an important evolutionary trait of members of the family Closteroviridae (Dolja et al., 2006; Karasev, 2000) and viruses infecting perennial crops have a greater chance for recombination due to the accumulation of different viruses over the years of cultivation. The high genetic variability, the incongruent clustering of isolates for all three genes analyzed, and the recombination events detected in this study, implicate recombination in having a role in evolution and shaping of population structure of this virus.

In conclusion, our results strongly suggest that the virus characterized in this work is a novel species in the genus Ampelovirus for which name blackberry vein banding associated virus is proposed. It was detected originally in a blackberry yellow vein disease-affected blackberry plant and later found to be common in cultivated and wild blackberries and always associated with symp- toms. Moreover, as it has been detected in diseased samples from several states, BVaV appears to be relatively widespread in the Southern United States. This high infection rate and the perennial nature of the host might be important factors contributing to the recombination and thus evolution of the virus. It is the first ampelovirus identified in Rubus spp. Considering that extant ampeloviruses are known to be mealybug-transmitted and implicated in serious diseases in other woody crops (i.e. grapevine, pineapple, stone fruits), our current focus is on identification of its vector and possible role (along with other viruses) in the complex etiology of blackberry yellow vein disease.

Acknowledgements

The authors acknowledge the financial support provided by NIFA-SCRI (2009-51181-06022), USDA-NCNP (11-8100-1572) and Special Research Initiative Program of the MAFES, Mississippi State University. We thank Ms. Yingxue (Shirley) Xie for her assistance in the lab work.
Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.viruses.2013.09.039.

References


