Sequence diversity, population genetics and potential recombination events in grapevine rupestris stem pitting-associated virus in Pacific North-West vineyards

Olufemi J. Alabi,1 Robert R. Martin2 and Rayapati A. Naidu1

1Department of Plant Pathology, Washington State University, Irrigated Agriculture Research and Extension Center, Prosser, WA 99350, USA
2USDA-ARS Horticultural Crops Research Laboratory, Corvallis, OR 97330, USA

Grapevine rupestris stem pitting-associated virus (GRSPaV; genus Foveavirus, family Flexiviridae; Martelli & Jelkmann, 1998; Adams et al., 2004, 2005) is a graft-transmissible virus widely distributed in many grape-growing regions around the world (Minafra & Boscia, 2003). The positive-sense, single-stranded RNA of GRSPaV is 8725 nt in length, excluding the poly(A) tail, and encodes five open reading frames (ORFs; Meng et al., 1998; Zhang et al., 1998). The 3'-most ORF encodes a 28 kDa coat protein (CP), whereas the 5'-most ORF encodes the viral replicase polyprotein of 244 kDa. The ORFs 2, 3 and 4 encode polypeptides of 24, 13 and 8 kDa, respectively, which together constitute the ‘triple gene block’ (TGB) proteins TGBp1, TGBp2 and TGBp3, respectively (Meng et al., 1998). Recent studies have shown that TGBp1 has both a cytosolic and a nuclear distribution, whereas both TGBp2 and TGBp3 are associated with the endoplasmic reticulum (Rebelo, et al., 2008). The replicase polyprotein of GRSPaV contains two characteristic domains conserved in the alphavirus-like superfamily of the positive-strand RNA viruses (Koonin & Dolja, 1993) and two novel domains, AlkB and OTU, that are present in several other members of the family Flexiviridae (Makarova et al., 2000; Martelli et al., 2007). A putative ORF6 encoding a 14 kDa protein of unknown function, partially overlapping the CP at the C terminus, has been reported in some variants of GRSPaV (Meng et al., 1998; Zhang et al., 1998; Lima et al., 2006). Grapevine (Vitis spp.) is the only known natural host of GRSPaV. The virus can be spread via vegetative propaga-

INTRODUCTION

Grapevine rupestris stem pitting-associated virus (GRSPaV; genus Foveavirus, family Flexiviridae; Martelli & Jelkmann, 1998; Adams et al., 2004, 2005) is a graft-transmissible virus widely distributed in many grape-growing regions around the world (Minafra & Boscia, 2003). The positive-sense, single-stranded RNA of GRSPaV is 8725 nt in length, excluding the poly(A) tail, and encodes five open reading frames (ORFs; Meng et al., 1998; Zhang et al., 1998). The 3'-most ORF encodes a 28 kDa coat protein (CP), whereas the 5'-most ORF encodes the viral replicase polyprotein of 244 kDa. The ORFs 2, 3 and 4 encode polypeptides of 24, 13 and 8 kDa, respectively,
tion and grafting (Minafra & Boscia, 2003) and possibly through seeds of infected grapevine (Lima et al., 2007). Although GRSPaV has been detected in the pollen of infected grapevines (Rowhani et al., 2000b), its spread through pollen is not confirmed. The virus is not transmissible by mechanical inoculation and no biological vector has been reported. Studies have indicated the biological association of GRSPaV with rupestris stem pitting (RSP), one of the four disorders of the rugose wood complex (Martelli, 1993). RSP shows pitting symptoms on the woody cylinder below the graft union in ‘St George’ grapevines (Vitis rupestris Scheele). The aetiological relationship of GRSPaV with RSP, however, remains to be elucidated. Previous studies have shown that GRSPaV can occur as distinct variants (Rowhani et al., 2000b; Habili et al., 2006; Meng et al., 2006; Nolasco et al., 2006; Nakaune et al., 2008), some of which may not elicit RSP symptoms when graft-inoculated on the indicator ‘St George’ (Meng et al., 1999, 2006; Habili et al., 2006). In addition, some variants that are latent in Vitis vinifera cultivars and in most American Vitis species and hybrids can produce necrosis of the veinlets when graft-inoculated on V. rupestris × Vitis berlandieri 110 Richter (Bouyahia et al., 2005). Recently, GRSPaV, along with a novel marafivirus and several viroids, has been documented in Californian Syrah grapevines showing decline symptoms, although the role of GRSPaV in Syrah decline is yet to be resolved (Al Rwahnih et al., 2009).

The Pacific North-West (PNW) states of the United States (Washington, Oregon and Idaho) are collectively emerging as the second largest wine grape-growing region in the country. Wine grapes are largely grown as own-rooted vines in Washington and Idaho and on rootstocks in Oregon. Since exchange of planting material occurs among growers in the region, we have been carrying out studies to document grapevine viruses and their genetic variants for assessing the sanitary status of vineyards. Such information will aid the development of strategies for mitigating their negative impact on the sustainability of the wine grape industry in the region. Towards this objective, we assessed the genetic diversity of natural GRSPaV populations from PNW and compared them with isolates from other grape-growing regions. The results indicated that GRSPaV is highly variable in chronically infected, vegetatively propagated grapevines. The study provides evidence for the first time of the occurrence of potential recombination events in a foveavirus. The occurrence of mixed infection of genetically divergent variants within a single grapevine and recombination events identified in this study could, in part, contribute to the extensive sequence diversity and evolution that could impact grapevine health status and diagnosis of GRSPaV in grapevines. On a wider scale, the results contribute to an increased understanding of molecular population genetics of viruses infecting deciduous woody perennials.

RESULTS

Genetic diversity of GRSPaV populations from PNW

Primers RSP52F and RSP53R amplified a DNA fragment of approximately 905 bp encompassing the entire CP and the 62 and 63 bp upstream and downstream of the CP, respectively (Fig. 1). However, the flanking sequences were removed and only the CP gene (780 bp) sequences were used for analyses. The primers RSP4373F and RSP4711R amplified a fragment of approx. 339 bp specific to the helicase-encoding region (HR) and after removing the primer sequences, only a 299 nt sequence was used for further analyses. Both CP- and HR-specific fragments were amplified from a total of 34 grapevines, whereas a CP- or

![Fig. 1. (a) Genome organization of GRSPaV](image-url)
HR-specific fragment only was amplified from an additional 20 and 19 grapevines, respectively. Thus, a total of 84 CP sequences and 57 HR sequences derived from 54 and 53 individual grapevines, respectively, were included for the genetic diversity analyses. These sequences were deposited in GenBank under the accession numbers FJ943274–FJ943357 and FJ943358–FJ943414 for CP and HR, respectively (Supplementary Table S1, available in JGV Online).

In pairwise comparisons, the 84 CP sequences from PNW showed identities ranging between 80 and 100 % at the nt level and between 86 and 100 % at the aa level. The identities among the 57 HR sequences from the PNW ranged between 79 and 100 % and between 86 and 100 % at the nt and aa level, respectively. Similar ranges of values were obtained when comparisons were made with corresponding sequences in GenBank (Supplementary Table S2, available in JGV Online). These values are within the limits of species demarcation criteria in the family Flexiviridae, where isolates sharing greater than 72 % nt or 80 % aa sequence identities between their CP or polymerase genes are considered one species (Adams et al., 2005). Based on these results, it can be concluded that CP and HR sequences are specific to GRSPaV and divergent variants of the virus are present in different wine grape cultivars grown in PNW vineyards.

**GRSPaV isolates from the PNW comprise four major genetic lineages**

Initially, all 159 sequences of the CP (84 from this study and 75 in GenBank) and 94 sequences of the HR (57 from this study and 37 in GenBank) were included in assessing phylogenetic relationship of GRSPaV isolates. A comparison of CP- and HR-based phylogenetic trees revealed inconsistencies in the placement of 19 GRSPaV isolates (18 from PNW and one from GenBank) in different lineages with poor bootstrap support, suggesting that their genomic RNA might have originated from potential recombination events between different lineages. These events are described below to explain the inconsistencies. We omitted these interlineage recombinants and recalculated trees from the CP and HR sequences of other isolates together with sequences available in GenBank as described above, and the inferred phylogenetic trees are shown in Fig. 2.

The CP-based phylogenetic relationships of GRSPaV isolates from PNW were compared among themselves and with corresponding sequences from other grape-growing regions. This analysis, using the minimum evolution (ME) and neighbour-joining (NJ) methods with apple stem pitting virus (ASPV) CP as an outgroup, included a total of 140 CP sequences, 66 from the PNW and 74 from GenBank. The results showed segregation of GRSPaV CP sequences into four major lineages (Fig. 2a). We designated each of these lineages with a reference isolate to maintain a standardized nomenclature of GRSPaV sequence variant groups in analogy with a previous report by Meng et al. (2006). Thus, GRSPaV-1, GRSPaV-SG1, GRSPaV-BS and GRSPaV-VS lineages correspond to groups 2b, 2a, 3 and 1, respectively, as proposed by Nolasco et al. (2006) and Nakaune et al. (2008). The GRSPaV-1, GRSPaV-SG1, GRSPaV-BS and GRSPaV-VS lineages contained 11, 12, 21 and 22 sequences from PNW, respectively, and 20, 10, 15 and 29 from GenBank, respectively. These results indicate that GRSPaV sequences from the PNW were distributed in all four lineages with no geographical structuring, suggesting that they are polyphyletic. However, a greater number of sequences from PNW were aligned with GRSPaV-VS lineage, followed by GRSPaV-BS, GRSPaV-SG1 and GRSPaV-1 lineages. The mean genetic distance between the four groups ranged from 0.107 to 0.189, and those within each group ranged from 0.038 to 0.082.

A total of 74 sequences – 38 from PNW vineyards and 36 from GenBank – specific to HR were analysed phylogenetically. HR sequences also segregated into four lineages (Fig. 2b) with each lineage designated in a manner similar to that described for the CP-based phylogram. The GRSPaV-1, GRSPaV-SG1, GRSPaV-BS and GRSPaV-VS lineages contained 25, 12, 0 and 1 isolates from PNW, respectively, and 10, 16, 7 and 3 from GenBank, respectively. In contrast with CP sequences, many of the HR sequences from PNW clustered with the GRSPaV-1 lineage, followed by the GRSPaV-SG1 lineage and only one sequence was aligned with the GRSPaV-VS lineage and none with the GRSPaV-BS lineage. The mean genetic distance between the four lineages ranged from 0.118 to 0.215, and those within each lineage ranged from 0.031 to 0.080.

**Population genetic analysis shows variation in the two genomic regions of GRSPaV**

The population genetic parameters are listed in Table 1. Haplotype diversity (Hd) values for CP and HR sequences were close to 1.000 and π values (estimated by the average number of nucleotide differences between two random sequences in a population) for the two genomic regions ranged from 0.027 to 0.128. The highest genetic variation was observed in the CP (π=0.128), indicating that CP was the more variable of these two genomic regions. However, lower π values for each lineage suggest less variability within individual lineages, although differences in π values suggest genetic variation between them.

Nucleotide polymorphisms in the GRSPaV CP and HR populations were evaluated using Tajima’s D (Tajima, 1989) and Fu and Li’s D and F (Fu & Li, 1993) statistical tests to assess the influence of demographic forces on the population (Hey & Harris, 1999; Tajima, 1989; Tsompana et al., 2005). The significantly negative values of Tajima’s D and Fu and Li’s D and F statistical tests (Table 1) for CP sequences discount the neutral hypothesis but suggest the occurrence of demographic expansion of GRSPaV populations. This trend seemed less marked for the HR sequences than for the CP, since the Tajima’s D and Fu and Li’s D and F values for HR did not significantly deviate from zero, except for isolates belonging to GRSPaV-VS lineage.
Fig. 2. Phylogenetic analysis of GRSPaV isolates based on CP (a) and HR (b) sequences. The trees were constructed by the ME method using the NJ algorithm implemented by MEGA 4. The trees exclude interlineage recombinants identified in this study. The trees were rooted by using ASPV sequences as an outgroup. Bootstrap values (1000 replicates) are given at the branch nodes. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates are collapsed. The optimal trees with the sums of branch length (2.03681006 for HR and 2.58111728 for CP) are shown. Isolates from the PNW are in bold.
Table 1. Population genetic parameters and neutrality tests calculated for the HR and CP coding regions of GRSPaV

<table>
<thead>
<tr>
<th>Region (nt)</th>
<th>Lineage</th>
<th>n</th>
<th>S</th>
<th>Hd</th>
<th>π</th>
<th>θ</th>
<th>dS</th>
<th>dS/dS</th>
<th>Tajima’s D</th>
<th>Fu &amp; Li’s D</th>
<th>Fu &amp; Li’s F</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR (299)</td>
<td>All</td>
<td>94</td>
<td>145</td>
<td>0.998</td>
<td>0.092</td>
<td>0.098</td>
<td>0.695</td>
<td>0.019</td>
<td>0.027</td>
<td>-0.974</td>
<td>-1.645</td>
</tr>
<tr>
<td></td>
<td>GRSPaV-1</td>
<td>5</td>
<td>34</td>
<td>1.000</td>
<td>0.054</td>
<td>0.055</td>
<td>0.294</td>
<td>0.008</td>
<td>0.027</td>
<td>-0.147</td>
<td>-0.147</td>
</tr>
<tr>
<td></td>
<td>GRSPaV-SG1</td>
<td>32</td>
<td>89</td>
<td>0.992</td>
<td>0.070</td>
<td>0.077</td>
<td>0.356</td>
<td>0.021</td>
<td>0.059</td>
<td>-0.849</td>
<td>-1.159</td>
</tr>
<tr>
<td></td>
<td>GRSPaV-BS</td>
<td>10</td>
<td>36</td>
<td>1.000</td>
<td>0.031</td>
<td>0.043</td>
<td>0.199</td>
<td>0.010</td>
<td>0.049</td>
<td>-1.272</td>
<td>-1.236</td>
</tr>
<tr>
<td></td>
<td>GRSPaV-VS</td>
<td>47</td>
<td>94</td>
<td>0.995</td>
<td>0.032</td>
<td>0.071</td>
<td>0.483</td>
<td>0.020</td>
<td>0.041</td>
<td>-2.144*</td>
<td>-3.128*</td>
</tr>
<tr>
<td>CP (780)</td>
<td>All</td>
<td>159</td>
<td>427</td>
<td>0.999</td>
<td>0.128</td>
<td>0.097</td>
<td>0.874</td>
<td>0.032</td>
<td>0.037</td>
<td>-0.318</td>
<td>-2.640*</td>
</tr>
<tr>
<td></td>
<td>GRSPaV-1</td>
<td>63</td>
<td>258</td>
<td>0.995</td>
<td>0.045</td>
<td>0.070</td>
<td>0.205</td>
<td>0.010</td>
<td>0.046</td>
<td>-1.595†</td>
<td>-2.591*</td>
</tr>
<tr>
<td></td>
<td>GRSPaV-SG1</td>
<td>25</td>
<td>215</td>
<td>1.000</td>
<td>0.067</td>
<td>0.073</td>
<td>0.089</td>
<td>0.011</td>
<td>0.125</td>
<td>-0.768</td>
<td>-1.076</td>
</tr>
<tr>
<td></td>
<td>GRSPaV-BS</td>
<td>33</td>
<td>182</td>
<td>0.998</td>
<td>0.027</td>
<td>0.058</td>
<td>0.330</td>
<td>0.013</td>
<td>0.038</td>
<td>-2.188†</td>
<td>-2.672*</td>
</tr>
<tr>
<td></td>
<td>GRSPaV-VS</td>
<td>38</td>
<td>277</td>
<td>1.000</td>
<td>0.069</td>
<td>0.085</td>
<td>0.353</td>
<td>0.013</td>
<td>0.038</td>
<td>-1.209</td>
<td>-1.128</td>
</tr>
</tbody>
</table>

*Significant values that reject the null hypothesis of selective neutrality; P<0.05.
†Significant values that reject the null hypothesis of selective neutrality; P<0.02.
‡Significant values that reject the null hypothesis of selective neutrality; P<0.10.

The ratio of the rate of non-synonymous (dN) to synonymous (dS) substitutions for the CP and HR was less than 1.0 (0.037 and 0.027, respectively; Table 1), implying that these regions are under predominantly purifying selection. The distribution profiles of non-synonymous and synonymous substitutions were analysed separately for HR and CP (data not shown). The synonymous substitution curves for HR showed a generally consistent slope that implies that these substitutions were almost evenly distributed. The synonymous substitution curve for the CP shows a reduced number of substitutions in the last 100 aa compared with the first 150 aa, suggesting that substitutions in the CP are not selectively neutral. In the HR, few non-synonymous substitutions were observed between codons 10 and 45 with a fairly constant rate of substitutions in other portions. In the CP, a higher frequency of non-synonymous substitutions was observed in the first 100 codons of the N-terminal portion than in the other portion of the gene. These differences in the frequencies of non-synonymous substitutions indicated a bias in the distribution of these changes in the CP and HR.

GRSPaV CP sequences show intra-plant genetic variation with respect to viticultural practices

In a previous study (Meng et al., 2006), the population structure of GRSPaV was found to be different between grapevine scion and rootstock cultivars, with the former harbouring mixtures of distinct variants. Since the wine grape scion cultivars are grown on rootstocks in Oregon and as own-rooted plants in eastern Washington, we assessed the status of CP sequence diversity in cultivars grown under different viticultural practices (grafting versus own-rooting). Our results indicated that distinct CP sequences occurred as mixtures in 10 of 54 grapevines. Further analysis of these sequences from the 10 grapevines revealed that they belonged to two or more lineages (Table 2). Eight of these grapevines were from Oregon grown on rootstocks and only two were from eastern Washington grown as own-rooted plants. The variant sequences obtained from these 10 grapevines belonged to two or three different groups and they occurred in various proportions with one variant sequence predominating over the others in a given plant. Although the majority of samples in Oregon came from one cultivar (cv. Pinot Noir), samples from one other cultivar (cv. Chardonnay) also had genetically distinct variants in a single grapevine. We did not find the occurrence of all four variant groups in the same grapevine in this study. In contrast, the majority of CP sequences derived from grapevine samples collected from own-rooted scion cultivars in eastern Washington revealed no such intra-plant diversity. Only two samples, one each from cvs Syrah and Merlot, contained variant sequences belonging to two and three different lineages, respectively. These results indicate propensity of the occurrence of mixtures of distinct variants of GRSPaV in grafted plants. The profile of virus variants in individual grapevines revealed no specific correlation between variant groups and scion cultivars.

Discordance between the CP and HR phylogenies suggests recombination among GRSPaV sequence variants

Among the 34 isolates of GRSPaV from PNW vineyards, for which both CP and HR sequences are available, a group of 27 isolates contained a single variant. Phylogenetic relationships of these 27 isolates based on HR and CP concatenated sequences showed four major lineages (Fig. 3a) analogous to the tree presented in Fig. 2. The CP-based tree showed the same groupings as for the concatenated sequences (Fig. 3c), whereas the HR-based tree differed...
with respect to the position of 19 isolates (Fig. 3b). Since sequences of each of the 19 isolates comprised a single variant, their discordant phylogenies would suggest potential recombination events involving parents from different lineages. We investigated this possibility with the suite of programs included in the RDP3 package using concatenated lineages. We investigated this possibility with the suite of programs included in the RDP3 package using concatenated sequences of the HR and CP from all 27 isolates.

A total of 19 putative recombination events were detected by at least four programs of the RDP3 software. The recombinant isolates and their potential ‘parental sequences’ are listed in Table 3. The cross-over sites were identified at different locations for these recombinants, with 17 and 2 involving inter- and intra-lineage recombination events, respectively. The majority of these recombination events involved parental isolates from GRSPaV-1 and GRSPaV-VS lineages generating 10 recombinant sequences (EWSY1, EWSY2, EWSY3, EWCH1, EWCH2, EWCH3, EWCH4, EWCH6, ORPN15 and WWPN4). The other recombination events involved GRSPaV-SG1 and GRSPaV-1 lineages generating five recombinant sequences (EWCH7, ORPGB1, EWCS4, WWPN2 and WWPN3) and one each involving GRSPaV-1 and GRSPaV-BS, and GRSPaV-VS and GRSPaV-BS, resulting in ORPGR1 and GRSPaV-SY, respectively. In addition, intra-lineage recombination events occurred between ‘parental’ sequences from the GRSPaV-BS lineage resulting in two progeny sequences (EWCH9 and FBBGW1). A closer look at the various recombination events indicated proclivity for such events occurring between lineages GRSPaV-1 and GRSPaV-VS. In addition, 11 and 2 of 19 crossover sites were present in CP and HR sequences, respectively. Four crossover sites were observed between concatenated sequences of the HR and CP, indicating possible recombination events in portions of the virus genome between HR and CP.

**DISCUSSION**

In this study, the molecular diversity of field isolates of GRSPaV in wine grape cultivars grown in the PNW region was assessed relative to virus isolates from other grape-growing regions. A phylogenomic approach was used to analyse a total of 159 full-length CP sequences (84 from this study and 75 from GenBank) and 94 sequences specific to the HR of RNA-dependent RNA polymerase (RdRp) (57 from this study and 37 from GenBank). As only a few isolates from the USA and none from the PNW were included in previous studies, the work represents an analysis of GRSPaV isolates at a global level, extending previous investigations conducted in other grape-growing regions of the world (Meng et al., 2006; Nolasco et al., 2006; Habili et al., 2006; Lima et al., 2006; Nakaune et al., 2008). These comparative results show that the genetic diversity of GRSPaV in the PNW vineyards is considerably greater than reported in other regions, probably as a result of the introduction of planting materials from several sources outside the region.

The results reinforce that the global GRSPaV population is very diverse, with numerous, disparate strains segregating into four distinct lineages (Fig. 2a, b). Several factors could contribute to the perpetuation of such a complex and dynamic population structure. The perennial and clonal propagation of grapevines, chronic infection for many years without host mortality and the absence of a biological vector provide a conducive environment for the evolution of a population of genetically related variants, in the absence of selection pressure or a bottleneck for different variants generated by short replication times and the error-prone nature of the RdRp (García-Arenal et al., 2001). That grapevines often remain in the field for several years or decades coupled with cultural practices such as grafting onto rootstocks and topworking to new cultivars

**Table 2.** Proportions of GRSPaV variant mixtures present in individual grapevine isolates

The CP coding region was assessed. EWMR1 and EWSY10 were grown on their own roots; all other isolates were grown on rootstock.

<table>
<thead>
<tr>
<th>Isolate*</th>
<th>Sequence variant groups*</th>
<th>Nt identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1 (GRSPaV-VS)</td>
<td>Group 2a (GRSPaV-SG1)</td>
</tr>
<tr>
<td>EWMR1</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>EWSY10</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>ORCH1</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>ORPN1</td>
<td>–</td>
<td>7</td>
</tr>
<tr>
<td>ORPN7</td>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>ORPN9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>ORPN13</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>ORPN21</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>ORPN22</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>ORPN24</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

*Supplementary Table S1 provides a description of the virus isolates.
†The lineage of the corresponding sequence variant group is shown in parentheses and the numbers in each column show the number of clones obtained for each GRSPaV group in a given plant; –, absent.
could mean that new variants are introduced into a plant, providing additional opportunities for changes in the viral population dynamics within individual grapevines. The presence of genetically distinct variants in infected grapevines in the PNW and the presence of sequences from different genetic lineages within a scion cultivar grafted onto rootstock validates this argument.

Higher \( \pi \) values in the CP than in the HR (Table 1) suggest that the CP is the more variable of the two coding regions. Overall, the high level of \( \pi \) of GRSPaV isolates in both coding regions is largely due to the accumulation of synonymous substitutions (data not shown). Although the \( d_S/d_S \) ratio obtained for both CP and HR sequences was below 1.0, it was 1.4 times higher for CP (0.037) than for the HR (0.027), suggesting a stronger purifying selection operating on the HR. Similar high levels of nucleotide sequence variability yet low values of \( d_S/d_S \) have been reported in different genomic regions for other members of the family Flexiviridae (Chare & Holmes, 2004, 2006; Shi et al., 2004; Teycheney et al., 2005), further indicating that the attribute is common to members of this family.

Considering the critical role of the helicase domain in viral RNA replication and its possible involvement in RNA
Table 3. List of GRSPaV isolates showing putative recombination events in concatenated sequences of CP and HR

Recombinant and parental isolates obtained from PNW are in bold and the sources of isolates are indicated in Supplementary Tables S1 and S2. The suite of recombination detection programs used for the detection of recombination events and the corresponding average \( P \)-values for each event were: R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, CHIMAERA; S, SiScan; 3S, 3SEQ. Bootscan, RDP and SiScan are phylogeny-based methods, GeneConv, MaxChi, CHIMAERA and LARD are substitution-based methods. No recombination events were detected using LARD. \( P \)-values \( > 0.05 \) were considered significant; --, no recombination event detected.

<table>
<thead>
<tr>
<th>Recombinant isolate</th>
<th>Parental isolate*</th>
<th>Major</th>
<th>Minor</th>
<th>R ( \times 10^{-05} )</th>
<th>G ( \times 10^{-05} )</th>
<th>B ( \times 10^{-04} )</th>
<th>M ( \times 10^{-07} )</th>
<th>C ( \times 10^{-07} )</th>
<th>S ( \times 10^{-21} )</th>
<th>3S ( \times 10^{-26} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWSY1</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>EWSY2</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>EWSY3</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>EWCH1</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>EWCH2</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>EWCH4</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>EWCH5</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>ORPN15</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>WWPN4</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>GRSPaV-SY (GRSPaV-VS)</td>
<td>1 x 10^{-05}</td>
<td>3 x 10^{-05}</td>
<td>1 x 10^{-04}</td>
<td>6 x 10^{-10}</td>
<td>3 x 10^{-07}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>EWCH7</td>
<td>EWCS3 (GRSPaV-SG1)</td>
<td>GRSPaV-1 (GRSPaV-VI)</td>
<td>3 x 10^{-06}</td>
<td>5 x 10^{-06}</td>
<td>2 x 10^{-06}</td>
<td>8 x 10^{-03}</td>
<td>7 x 10^{-03}</td>
<td>2 x 10^{-21}</td>
<td>3 x 10^{-26}</td>
<td></td>
</tr>
<tr>
<td>EWCH9</td>
<td>WWPN5 (GRSPaV-VS)</td>
<td>GRSPaV-1 (GRSPaV-VS)</td>
<td>3 x 10^{-06}</td>
<td>5 x 10^{-04}</td>
<td>2 x 10^{-06}</td>
<td>8 x 10^{-03}</td>
<td>7 x 10^{-03}</td>
<td>3 x 10^{-03}</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>FBGW1</td>
<td>WWPN5 (GRSPaV-VS)</td>
<td>Unknown (GRSPaV-VS)</td>
<td>3 x 10^{-06}</td>
<td>5 x 10^{-04}</td>
<td>2 x 10^{-06}</td>
<td>8 x 10^{-03}</td>
<td>7 x 10^{-03}</td>
<td>3 x 10^{-03}</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>GRSPaV-SY</td>
<td>FBGW1 (GRSPaV-VS)</td>
<td>Unknown (GRSPaV-VS)</td>
<td>3 x 10^{-06}</td>
<td>5 x 10^{-04}</td>
<td>2 x 10^{-06}</td>
<td>8 x 10^{-03}</td>
<td>7 x 10^{-03}</td>
<td>3 x 10^{-03}</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>ORPG1</td>
<td>EWSY9 (GRSPaV-VI)</td>
<td>Unknown (GRSPaV-VS)</td>
<td>3 x 10^{-06}</td>
<td>5 x 10^{-04}</td>
<td>2 x 10^{-06}</td>
<td>8 x 10^{-03}</td>
<td>7 x 10^{-03}</td>
<td>3 x 10^{-03}</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>ORPG2</td>
<td>EWCS3 (GRSPaV-SG1)</td>
<td>GRSPaV-1 (GRSPaV-VI)</td>
<td>3 x 10^{-06}</td>
<td>5 x 10^{-04}</td>
<td>2 x 10^{-06}</td>
<td>8 x 10^{-03}</td>
<td>7 x 10^{-03}</td>
<td>3 x 10^{-03}</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>ORPG3</td>
<td>EWCS4 (GRSPaV-SG1)</td>
<td>GRSPaV-1 (GRSPaV-VI)</td>
<td>3 x 10^{-06}</td>
<td>5 x 10^{-04}</td>
<td>2 x 10^{-06}</td>
<td>8 x 10^{-03}</td>
<td>7 x 10^{-03}</td>
<td>3 x 10^{-03}</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>WWPN2</td>
<td>GRSPaV-1 (GRSPaV-VI)</td>
<td>EWCS3 (GRSPaV-SG1)</td>
<td>3 x 10^{-06}</td>
<td>5 x 10^{-04}</td>
<td>2 x 10^{-06}</td>
<td>8 x 10^{-03}</td>
<td>7 x 10^{-03}</td>
<td>3 x 10^{-03}</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

*Minor’ and ‘Major’ parents refer to the parental isolates contributing the smaller and larger fractions of the recombinant’s sequence, respectively. Lineages of parental isolates (in italics) are derived from Fig. 3(a). An ‘unknown’ parent is so called probably because isolates distantly resembling one of the recombinant’s parents have been sampled.
unwinding and cap formation (Martelli et al., 2007), it is likely that HR is less tolerant of mutations that could affect its function. In contrast, a lack of biological vector for transmission makes the CP of GRSPaV more tolerant to sequence variability, since the capsid proteins of vector-borne plant viruses are subject to greater purifying selection on amino acid changes than viruses transmitted by other routes (Chare & Holmes, 2004; Rubio et al., 2001). It is also plausible that the C-terminal third of the CP is less tolerant of mutations due to assembly or particle structural limitations. Given the high degree of the global genetic variation among GRSPaV sequences, it is unlikely that the estimated error rate of $1 \times 10^{-4}$ to $2 \times 10^{-5}$ errors per base pair for Taq polymerase (Lundberg et al., 1991; Bracho et al., 1998) would have contributed to the apparent nucleotide divergence obtained in this study and previous reports (Meng et al., 2006; Nolasco et al., 2006; Habili et al., 2006; Lima et al., 2006; Nakaune et al., 2008).

Phylogenetic analysis of GRSPaV isolates characterized in this study further supports the existence of the same four phylogroups (Meng et al., 2006; Nolasco et al., 2006; Nakaune et al., 2008), irrespective of the protein employed for the analysis (Fig. 2). These isolates were distributed in all four groups, confirming a lack of clustering by geographical origin. Such a situation could be due to the spread of GRSPaV exclusively through the distribution of vegetative cuttings, in contrast with geographical delineation of virus variants in some vector-borne viruses (Karan et al., 1994; Varsani et al., 2008). Although the phylograms derived in this study using the entire CP were similar to HR-based phylograms, the placement of some GRSPaV isolates showed differences between CP- and HR-based phylograms due to possible recombination events in the genome (discussed below). In view of such discordant gene phylogenies among GRSPaV isolates, it seemed appropriate to use one of these two genes for comparative sequence analysis and we favour the whole CP gene sequences as the best representation of the phylogeny of global isolates of GRSPaV. Such CP-based phylogeny has been used for the classification of several different virus groups (Karan et al., 1994; Pfosser & Baumann, 2002; Fajardo et al., 2007).

The discordant gene phylogenies obtained for CP and HR sequences of GRSPaV isolates (Fig. 3) provide strong evolutionary evidence for recombination. Factors that can cause different genes to give different topologies have been well reviewed (Jeffroy et al., 2006; Rokas & Carroll, 2006; Castresana, 2007). These include the stochastic nature of mutation, lineage sorting, phylogenetic reconstruction artefacts and methodological problems related to the assessment of homology. Hence, similar, rather than identical, trees should be expected from different genes (Penny et al., 1982; Huerta-Cepas et al., 2007). In order to investigate recombination events among GRSPaV isolates, the two genomic regions were concatenated and analysed by a suite of programs representing phylogenetic-, substitution- and distance-based methods for robust assessment of definitive evidence of recombination. Several potential breakpoints for recombination have been identified in CP and HR sequences. It is unlikely that these recombinants were RT-PCR artefacts, since a minimum of three independent clones originating from different PCR products were sequenced to avoid artificial results arising from RT, PCR and cloning. Besides, the occurrence of nearly identical recombinant breakpoints in different isolates and the highly significant statistical support provided for each event strongly suggest that the recombinants are genuine. In addition, Nolasco et al. (2006) reported a recombination event involving GRSPaV isolates belonging to two of the four CP variant groups from Portugal and Slovenia. However, support for this putative recombinant event came from only four of the eight algorithms implemented by the RDP program.

As the CP and HR sequences represent only about 12% of the GRSPaV genome, the likelihood of recombination events in the portions of virus genome not covered by present analyses cannot be excluded. Thus, the actual proportion of recombination events may be much greater and the extent of recombination in the global GRSPaV population can be realized when the full genomic sequences of a representative set of isolates from different grape-growing regions are available. A possible explanation for recombination events observed in this study is that the scion cultivar could be carrying one variant and certain rootstocks could be carrying another and, after grafting, these variants might have mixed together in the same grapevine. Mixed infections provide the opportunity for recombination to occur between different strains and are therefore important in increasing the diversity of GRSPaV within populations. Such recombination events appear to be common in the genomes of viruses infecting vegetatively propagated perennial crops such as citrus (Rubio et al., 2001; Weng et al., 2007). In a broader context, recombination in viruses infecting perennial crops offers potentially significant advantages for an increased genetic diversity and adaptability, because deleterious mutations accumulated due to the lack of proofreading activity of an RdRp can be offset by recombination of the error-free parts of co-infecting genomes (Garci-Arenal et al., 2001; Vives et al., 2005). Thus, RNA recombination is likely to play a significant role in GRSPaV variation and evolution. Consequently, every field isolate of GRSPaV could be unique in each gene sequence due to the highly variable nature of the virus. Previous studies have documented genetic diversity at the intra-host level in perennial woody hosts (Magome et al., 1999; d’Urso et al., 2000). A recent study has shown that plum pox virus has evolved into several distinct populations to the point that in different parts of the same Prunus tree, the composition of these populations was different over a period of 13 years after inoculation (Jridi et al., 2006). It will be interesting to pursue similar studies with GRSPaV to better understand the dynamics of genetic
diversity and the potential association between different *Vitis* spp. and cultivars and virus population structures.

The role of GRSPaV in the aetiology of RSP disorder is still unresolved (Nakaune et al., 2008), and therefore, the possible biological implications of genetic variants of GRSPaV are difficult to assess presently. A recent study in Australia (Habili et al., 2006) has shown no association between the type of symptom expressed on the indicator host (cv. St George) and the sequence diversity in the virus genome. This is further compounded by the presence of GRSPaV isolates causing latent infections in the *V. rupestris* sources widely used for biological indexing (Minafra et al., 2000; Meng et al., 1999; Petrovic et al., 2000). It is likely that such latent infections could be the result of antagonism between genetically divergent variants present as mixed infections in a grapevine resulting in erroneous conclusions of biological indexing assays. Conversely, GRSPaV variants present at low frequencies might play an important biological role, for example in virus pathogenesis. Recent studies have shown the association of GRSPaV with grapevine rupestris vein-feathering virus and grapevine Syrah virus-1 in grapevines showing Syrah decline symptoms (Al Rwahnih et al., 2009). Thus, from a practical point of view, the results presented in this study provide a foundation for further exploring the biological role of GRSPaV sequence variants in disorders like Syrah decline and veinal necrosis. The likely recombination events in the CP and HR sequences, as reported in this study, and the frequent occurrence of GRSPaV as mixed infections with other known and currently uncharacterized graft-transmissible agents (Credi, 1997; Komar et al., 2006; Naidu et al., 2006; Proser et al., 2007) add additional layers of complexity to the assessment of the biological behaviour of different genetic variants of GRSPaV. The availability of an infectious clone of GRSPaV would provide additional possibilities for establishing the aetiological relationship between GRSPaV and RSP and understanding the biological implications of molecular diversity in the virus genome.

**METHODS**

**Virus isolates.** Seventy-three GRSPaV isolates from different grapevine cultivars were included in this study (Supplementary Table S1). Virus derived from a single grapevine was considered as one isolate. The GRSPaV isolates were collected between 2005 and 2007 from different vineyards (names withheld due to grower confidentiality) in Washington and Oregon. Leaf petioles were randomly collected from different parts of individual grapevines and pooled for extractions to minimize possible variations due to uneven distribution of the virus in infected tissues. Samples from eastern Washington originated from own-rooted wine grape cultivars and those from western Washington and Oregon were from grapevines grafted onto a rootstock. Other samples were obtained from a collection maintained at the North-West Grape Foundation Service, Proser, WA.

**Sample extraction and RT-PCR.** Petiole sample extracts prepared with the aid of a HOMEX 6 homogenizer (BIOREBA AG) were used in a one step–single tube RT-PCR assay (Rowhani et al., 2000a) for amplification of two distinct regions of the virus genome (Fig. 1). Two sets of primers, designed based on the consensus sequence of multiple viral variants reported previously (Meng et al., 2006; Nolasco et al., 2006, Nakaune et al., 2008), were used to amplify the two regions of the virus genome. The primer pair RSP4373F (5'-GATTAGGGTCC-AGTTGTTTCC-3') and RSP4711R (5'-ATCCAAAGGACCCTTTTGACC-3') corresponds to nt 4373–4711 (GenBank accession no. AF057136) within the HR of ORF1 and the primer pair RSP52F (5'-TGAAGGCTTTAGGGT Tag-3') and RSP53R (5'-CTTAACCCAGCCTTGAAAT-3') corresponds to nt 7709–8613 (AF057136), encompassing the entire CP gene and flanking sequences upstream in the TGBp3 and downstream in the 3' untranslated region.

**Cloning, sequencing and sequence analysis.** Amplicons were cloned into pCR2.1 vector (Invitrogen) and transformed into *Escherichia coli*. Plasmid DNA was purified from positive recombinant clones using the QIAprep spin mini-prep kit (Qiagen). Three independent clones per isolate were sequenced in both orientations. Three clones of an isolate with sequence identity above 99% were considered to be a single sequence. Sequences derived from additional clones of an isolate were included in the analyses when differences were greater than 2%. Alignments of nucleotide sequences were done using CLUSTAL W (Thompson et al., 1994) with default settings, and sequence identities were obtained using Vector NTI Advance 11 program (Invitrogen).

**Phylogenetic analysis.** Evolutionary relationships were inferred from multiple sequence alignments calculated by CLUSTAL W using the ME method (Rzhetsky & Nei, 1992) with 1000 bootstrap replications. The evolutionary distances were computed using the Kimura two-parameter model (Kimura, 1980). The ME trees were searched using the close neighbour-interchange algorithm (Nei & Kumar, 2000) at a search level of 1. The NJ algorithm (Saitou & Nei, 1987) was used to generate the initial tree, for which all positions containing gaps and missing data were eliminated from the dataset. These phylogenetic analyses tools were implemented by MEGA version 4.0 (Tamura et al., 2007). Corresponding sequences of GRSPaV isolates available in GenBank (Supplementary Table S2) were included in these analyses. Estimates of mean evolutionary divergence over sequence pairs within and between groups were calculated using the MEGA4 program. Corresponding sequences of ASPV (GenBank accession no. D21829) were used as an outgroup.

**RNA polymorphism and evolution.** DnAsP version 4.90.1 (Rozas et al., 2003) was used to estimate Tajima’s D (Tajima, 1989) and Fu and Li’s D and F (Fu & Li, 1993) statistical tests to examine the hypothesis of neutral selection operating on the CP and HR sequences. We also estimated several population genetic parameters including nucleotide polymorphism (π; estimated by the average number of nucleotide differences between two random sequences in a population), haplotype diversity (Hd; the frequency and number of haplotypes in a population), the statistic θ from the number of segregating sites (S) (Waterson, 1975) and the average rate of non-synonymous and synonymous substitutions. The distribution of dS and dN along the coding regions was analysed using the SNAP program (http://www.hiv.lanl.gov; Korber, 2000).

**Recombination analysis.** The occurrence of recombination events in CP and HR sequences was investigated by using the suite of programs included in RD93 Beta 27 (Martin et al., 2005).

**ACKNOWLEDGEMENTS**

The authors are grateful to Dr Ulrich Melcher (Oklahoma State University, Stillwater, OK) and Dr Roy French (USDA-ARS, Lincoln,
NE) for their constructive comments and suggestions. This work was supported in part by Washington Wine Commission’s Wine Advisory Committee, the Washington State University’s Agricultural Research Center and Extension Issue-focused Team internal competitive grant, Washington State Department of Agriculture Nursery Assessment Funds, Washington State Commission on Pesticide Registration, USDA-ARS North-West Center for Small Fruits Research and USDA-CSREES Viticulture Consortium-West. PPNS # 0527, Department of Plant Pathology, College of Agricultural, Human, and Natural Resource Sciences Agricultural Research Center Project No. WNPO 0616, Washington State University, Pullman, WA 99164-6240, USA.

REFERENCES


Varsani, A., Shepherd, D. N., Monjane, A. L., Oowor, B. E., Erdmann, J. B., Rybicki, E. P., Petersschmitt, M., Bridson, R. W., Markham, P. G. & other authors (2008). Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. *J Gen Virol** **89**, 2063–2074.


