Genetic variability and population structure of Grapevine leafroll-associated virus 3 isolates

Camilla Turturo,1 Pasquale Saldarelli,1 Dong Yafeng,2 Michele Digiaro,2 Angelantonio Minafra,1 Vito Savino1 and G. P. Martelli1

1Dipartimento di Protezione delle Piante e Microbiologia Applicata, Università degli Studi and Istituto di Virologia Vegetale del CNR, Sezione di Bari, via Amendola 165/A, 70126 Bari, Italy
2Istituto Agronomico Mediterraneo, Valenzano, Bari, Italy

The genetic variability and population structure of a collection of 45 Grapevine leafroll-associated virus 3 (GLRaV-3) isolates were studied by single-stranded conformation polymorphism (SSCP) and sequence analysis of the RNA-dependent RNA polymerase (RdRp), heat-shock protein 70 homologue (HSP-70) and coat protein (CP) genes. The distribution of SSCP profiles was not correlated with the geographical origin of the isolates, indicating that GLRaV-3 is a single, undifferentiated population. The majority of the isolates showed an SSCP profile and a population structure that were composed of a single predominant variant. However, 10% of the isolates for the RdRp and HSP-70 genes and 15% for the CP gene were composed of a combination of two or more variants. Estimation of genetic diversity and phylogenetic analysis disclosed the possible existence of vines with mixed infections by diverging sequence variants, showing, in some cases, possible recombination events. Furthermore, differences in the genetic diversity and constraints existing in the three regions analysed indicated a higher variability in the CP gene. The epidemiological and biological implications of this finding are discussed.

INTRODUCTION

Grapevine leafroll-associated virus 3 (GLRaV-3) is a widespread agent of leafroll (Rosciglione & Gugerli, 1986), an economically important disease of the grapevine that is second only to fanleaf degeneration (Martelli, 2003). This virus belongs to the genus Ampelovirus in the family Closteroviridae (Martelli et al., 2002). It is phloem-limited and transmitted in a semi-persistent manner by several species of mealybugs and soft-scale insects (Gugerli, 2003). Its RNA genome of 17 919 nt is organized into 13 ORFs that potentially encode the protease, replicase, movement, coat and replication-enhancing proteins (Ling et al., 1998, 2000). Functional evidence is only available for the CP cistron, whose product, when expressed in bacteria, is recognized by antibodies to intact virus particles (Ling et al., 1997).

Only one GLRaV-3 isolate (GenBank accession no. AF037268) has been sequenced completely (Ling et al., 1998, 2000). Very limited sequence information is available for other isolates with reference to the RdRp domain (GenBank accession no. AY424407; Fajardo et al., 2001), the heat-shock protein 70 homologue (HSP-70) (Saldarelli et al., 1998) and the p20 protein (Habili et al., 1995). By contrast, a great deal of sequence information exists on the variability and population structure of other members of the family Closteroviridae, mainly for Citrus tristeza virus (CTV), which occurs as a population of diverse isolates with different biological and epidemiological properties (Rubio et al., 1999, 2001; Kong et al., 2000).

An attractive reason for the study of closterovirus variability is the size and complexity of the genome, which uses combined strategies of polyprotein processing and subgenomic RNAs for its expression and evolution (Karasev, 2000). Furthermore, GLRaV-3 and CTV infect vegetatively propagated hosts that persist in the field for a long time and are potentially subjected to repeated infections through insects and as a result of cultural practices (e.g. top grafting). The potential impact of virus variability in such long-lived hosts is not understood. Finally, from such an investigation, useful information could be derived for the development of more efficient, nucleic acid-based diagnostic tools and to provide a better choice of possible transgenes for inducing resistance in grapevines.

In the present paper, the population structure and genetic variability of 45 GLRaV-3 isolates from 14 countries were investigated by single-stranded conformation polymorphism (SSCP) and sequence analysis of three different genomic regions: those encoding the RNA-dependent RNA polymerase (RdRp), HSP-70 and the coat protein (CP).
**METHODS**

**Virus isolates.** GLRaV-3 isolates were obtained from different varieties from 14 countries (see Supplementary Table, available in JGV Online). These were grown in a varietal collection of the Department of Plant Protection and Applied Microbiology of the University of Bari, which was kept under strict surveillance to avoid the establishment of mealybug infestations. By ‘isolate’, we mean a viral population from a single infected grapevine. All isolates were selected after identification by double antibody-sandwich ELISA using a polyclonal antiserum raised against the Italian GLRaV-3 isolate MT48.

**Total RNA extraction, RT-PCR and cDNA cloning.** Total RNA was extracted from cortical scrapings of dormant cuttings that were collected from five different branches of each vine in December 2001. The extraction method was according to Foissac *et al.* (2000), by absorption of total RNA on to silica particles after guanidine-buffer treatment of plant tissues. The primer pairs P3U (5′-CGCTCATGGTGAAGACGACG-3′) and P3D (5′-CTATGACAAAAATATGGGAGCAG-3′), LC1 (5′-CGCTACAGGCTGTGAAATTT-3′) and LC2 (5′-TTGTCCGGTACCAGATAT-3′), and CP3U (5′-ATGGCATTTGAACATTGGGC-3′) and CP3D (5′-CCGGCCGCTATAAACCCTCTTA-3′) were designed based on the nucleotide sequence of isolate GLRaV-3 NY1 (GenBank accession no. AF037268) to amplify fragments of 653, 546 and 484 bp from the RdRp domain of isolate GLRaV-3 NY1 (GenBank accession no. AF037268) to amplify fragments of 653, 546 and 484 bp from the RdRp domain of isolate GLRaV-3 NY1 (GenBank accession no. AF037268).

**SSCP and sequence analysis.** SSCP analysis of amplified PCR products was done following the protocol described by Palacio & Duran-Vila (1999) with minor modifications. A 5 μl aliquot of PCR product was combined with 15 μl denaturing solution [95% deionized formamide, 20 mM EDTA (pH 8·0), 0·05% xylene cyanol and 0·05% bromophenol blue], boiled for 10 min and chilled on ice. Denatured amplicons were separated by non-denaturing 10% (RdRp and HSP-70) or 12% (CP) PAGE at 200 V at 4 °C for 4 h. SSCP profiles were visualized by silver staining (Beidler *et al.*, 1982). Selected plasmids were purified by the ‘boiling method’ (Sambrook *et al.*, 1989) and sequenced automatically (MWG). Multiple nucleotide alignments were made by using CLUSTAL X (Thompson *et al.*, 1997) version 1.8, a Windows interface for the multiple sequence-alignment software CLUSTAL V. Numbers of synonymous and non-synonymous nucleotide substitutions were calculated by using the DIVERGE software in the Genetics Computer Group (GCG) package (Devereux *et al.*, 1984), imposing the Pamilo–Bianchi–Li method (Pamilo & Bianchi, 1993; Li, 1993). The mean similarity between two aligned sequences was calculated and plotted by using the GCG PLOTSIMILARITY program (Devereux *et al.*, 1984). MEGA version 2·1 (Kumar *et al.*, 2001) was used to calculate nucleotide distances, using the method of Jukes & Cantor (1969) for correction of superimposed substitutions. Phylogenetic analysis was done with the same software by using the neighbour-joining method (Saitou & Nei, 1987) with 1000 bootstrap replicates to assess the robustness of the nodes. Genetic diversity (mean nucleotide distance between two randomly selected sequence variants) within and between isolates was estimated by the method of Lynch & Crease (1990). Statistical analysis of genetic diversity was performed by a Wilcoxon non-parametric test (Sokal & Rohlf, 1993) using inter- and intra-isolate diversity values. Recombination events and identification of putative recombination junctions were detected by PHYLPRO version 0·8 (Weiller, 1998).

**RESULTS**

**Genetic variability in different genomic regions**

Analysis of the 45 isolates in the three genomic regions examined showed the existence of six, five and nine different
SSCP patterns in the RdRp, HSP-70 and CP genes, respectively (Fig. 1). Two categories of patterns were observed, one (hereafter indicated as 'simple') that was composed of two or three bands of similar intensity in the majority of the isolates (90% in the RdRp and HSP70 genes and 85% in the CP gene) and the other (hereafter indicated as 'complex') composed of four or more bands (Fig. 1). Given that the expected SSCP pattern should be composed of two or three bands (one of them being a 'conformomer', i.e. a conformational structure of the same sequence; Orita et al., 1989), it could be concluded that the majority of the isolates under study were composed of a single, predominant variant, whereas a more complex population structure existed with isolates showing a four-band pattern. The overall distribution of SSCP patterns showed a clear-cut prevalence of one pattern in the RdRp and CP genes, with frequencies of 0.57 and 0.53, respectively, whereas two patterns with frequencies of 0.44 and 0.33 were observed in the HSP-70 gene (Fig. 1). There was no apparent correlation between SSCP patterns and the geographical origin of the isolates. Amplification of the three regions were cloned from the following isolates: AA10, AUSG5, C6, FR1, IL1, MT12, NIG3, P1 and TU16 for the RdRp gene; AUSG5, C3, IL1, MT48, SY2, TU32 and USA6 for the HSP-70 gene; and AUSG5, C5, GR30, IL1, MN18, MT38, MT48, SS5, SS15, SY2, TU16 and USA6 for the CP gene. For each isolate and region, 10 clones, randomly selected, were analysed. Depending on the distribution of patterns of the 10 analysed clones, at least six clones were sequenced per isolate (Table 1). Criteria for choosing the isolates to be cloned were: (i) the need to analyse the population structure and the sequences of at least one representative isolate for each SSCP pattern; and (ii) to observe the population structures of isolates showing either 'simple' or 'complex' SSCP patterns. Clone distribution of isolates with patterns consisting of two or three bands had a single predominant (frequency of >0.9) haplotype or variant, identical to the original SSCP products (Fig. 2a), thus confirming the assumptions made after analysis of the original SSCP patterns (see above). Such a population structure would be expected according to the quasispecies theory for RNA viruses (Holland et al., 1992). By contrast, isolates with patterns consisting of four or more bands displayed clones with different SSCP haplotypes (from two to five), none of which was predominant (Fig. 2b).

**Table 1.** Clones sequenced and within-isolate genetic diversities in the three genomic regions

<table>
<thead>
<tr>
<th>Isolate</th>
<th>No. clones sequenced</th>
<th>RdRp</th>
<th>HSP-70</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RdRp</td>
<td>HSP-70</td>
<td>CP</td>
</tr>
<tr>
<td>AA10</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>AUSG5</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>C3</td>
<td>7</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>C5</td>
<td>6</td>
<td>0-029 ± 0-0016</td>
<td>0-032 ± 0-0017</td>
<td>0-033 ± 0-0055</td>
</tr>
<tr>
<td>C6</td>
<td>6</td>
<td>0-0271 ± 0-0045</td>
<td>0-0000 ± 0-0008</td>
<td>0-0000 ± 0-0008</td>
</tr>
<tr>
<td>FR1</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>GR30</td>
<td>7</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0554 ± 0-0083</td>
</tr>
<tr>
<td>IL1</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0203 ± 0-0051</td>
</tr>
<tr>
<td>MN18</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>MT12</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>MT38</td>
<td>7</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0102 ± 0-0011</td>
</tr>
<tr>
<td>MT48</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>SS5</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>SS15</td>
<td>6</td>
<td>0-0635 ± 0-0086</td>
<td>0-0492 ± 0-0073</td>
<td>0-0366 ± 0-0045</td>
</tr>
<tr>
<td>NIG3</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>P1</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>SY2</td>
<td>8</td>
<td>0-0000 ± 0-0000</td>
<td>0-0374 ± 0-0058</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>TU16</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>TU32</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
<tr>
<td>USA6</td>
<td>6</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
<td>0-0000 ± 0-0000</td>
</tr>
</tbody>
</table>
clones showing profiles A and B in Fig. 1c, respectively) gave a mean genetic diversity of 0·0014±0·0008 and 0·0026±0·0011, respectively. These values and the fact that clones with the same haplotype differed by no more than 5 nt from one another were taken as indications of the satisfactory accuracy of the SSCP conditions used, in agreement with previous reports (Rubio et al., 1999; Kong et al., 2000; Lin et al., 2003). Genetic diversity values in the three genomic regions were <0·1, indicating a limited genetic variability, similar to that reported for other plant viruses (Kong et al., 2000; García-Arenal et al., 2001). Among the three regions analysed, the mean within-isolate genetic diversity, calculated for all sequenced isolates, was significantly lower in the RdRp gene (0·0035±0·0005) than in the HSP-70 (0·0196±0·0023) and CP (0·0161±0·0018) genes. These differences were not clearly evident when the mean genetic diversity of the entire population was assessed, although the RdRp gene again showed the lowest value (RdRp, 0·0295±0·0044; HSP-70, 0·0340±0·0039; CP, 0·0492±0·0065). Therefore, statistical analysis using the Wilcoxon non-parametric test (Sokal & Rohlf, 1993) was carried out, which showed that differences among the mean genetic diversities for the entire population were significant (P<0·05) between RdRp and CP (P=0·00001) and between HSP-70 and CP (P=0·026), thus indicating that the CP gene was the most variable. Such differences in genetic variability prompted us to investigate the extent of selective pressure operating on each gene by estimating the ratio (dN/dS) of nucleotide substitutions at non-synonymous positions (dN) versus those at synonymous positions (dS). This analysis showed that all three regions were under high selective pressure, which was lower in the RdRp gene (RdRp, dN=2·0884; HSP-70, dN=1·11; CP, dN=1·13). However, the Wilcoxon analysis did not show any statistically significant differences among the three values.

Evidence for recombination events

Low values of within-isolate genetic diversity were observed for the majority of the isolates in the three genomic regions (Table 1). Exceptions were found in some isolates (C6 in RdRp; AUSG5, MT48 and SY2 in HSP-70; and GR30, IL1, SS5 and SS15 in CP) whose patterns were composed of more than four bands, all showing a within-isolate genetic diversity higher than between-isolate diversity (data not shown), especially in the HSP-70 and CP genes. Moreover, phylogenetic analysis carried out to dissect relationships among clones originating from the latter isolates showed incoherent topologies (Fig. 3a, b and c). For example, in the HSP-70 gene, some clones of isolate AUSG5 had smaller genetic distances from clones of other isolates (i.e. clones AUSG5-6 and SY2.4) than from clones originating from the parent isolate (Fig. 3b). Moreover, some isolates (i.e. AUSG5) had two highly diverging variants showing high genetic distance values in two (HSP-70 and CP) of the three genomic regions analysed (Fig. 3a, b and c). A similar population structure, previously described for CTV (Rubio et al., 2001), could originate from mixed infection of two diverging viral variants that subsequently underwent recombination. A careful analysis of the phylogenetic trees relative to the three genomic regions showed an incongruent topology of clones AUSG5-2 and IL1-1 in the HSP-70 and CP genes, respectively (Fig. 3b and c), as they did not cluster clearly with the other variants. Therefore, a systematic search for possible recombination events was performed by using PHYLPRO (Weiller, 1998), a program that estimates the coherence of phylogenetic relationships in a set of aligned sequences. Putative recombination events are displayed in the form of a peak pointing downwards in a graph, which allows allocation of the point of recombination to a precise region. Based on this, evidence for recombination was identified in the HSP-70 and CP regions (Fig. 4), but not in RdRp (data not shown). A new PHYLPRO analysis, done by excluding sequences with low phylogenetic correlation and introducing single sequences, identified ancestors of recombinant variants (data not shown), ultimately confirmed by analysis with the software PLOTSIMILARITY (Devereux et al., 1984). This analysis suggested that the HSP-70 gene of AUSG5-2 originated from a recombination event between
an AUSG5-3-like and an AUSG5-6-like variant around nt 370 (Fig. 5a). A similar analysis suggested that the CP gene of IL1-1 originated from a recombination event between an IL1-2-like and an NIG3-like variant around nt 195 (Fig. 5b). To determine whether these observed recombinant molecules originated from the in vitro ‘template switching’ activity of reverse transcriptase (Negroni & Buc, 2001) or from the Taq polymerase (Bradley & Hillis, 1997), a negative control consisting of a mixture of total RNA of two different isolates whose SSCP patterns and sequences were known was included in the RT-PCR assay. DNA amplified from the mixture was cloned and 10 obtained clones were analysed by SSCP. Results using the primers LC1/LC2 (Fig. 6a) and CP3U/CP3D (Fig. 6b) for the HSP-70 and CP regions, respectively, showed that most of the clones had the same SSCP pattern and sequence as the two ‘parental’ isolates. Sequencing of clones showing a ‘non-parental’ pattern (i.e. clone 7 in Fig. 6) did not reveal any in vitro recombination events.

**DISCUSSION**

The genetic variability and population structure of a collection of GLRaV-3 isolates from distant geographical areas were analysed in three genomic regions, spanning the genome in the 5‘→3’ direction and covering 9–2% of the
entire RNA genome. The SSCP technique, used for rapid differentiation of the isolates, was confirmed to be a useful tool for estimating genetic diversity, as reported previously (Kong et al., 2000; Rubio et al., 2001; Lin et al., 2003). However, a potential limit to our work could be the underestimation imposed by the primers used in the RT-PCR, which, although designed from very conserved regions, could operate selectively, thus excluding some variants. By contrast, considering the low rate of error of Taq DNA polymerase (Bracho et al., 1998), it could be assumed that its influence was negligible in the overall estimation of genetic diversity in the three genomic regions. This was also confirmed by the robustness of our experimental approach, which demonstrated that similar haplotypes had a maximum of five differences in a range of fragments of 484–653 nt. This approach allowed us to adopt SSCP analysis for the preliminary selection of isolates to be sequenced, assuming that isolates with similar patterns shared the same sequence. As also reported by Lin et al. (2003), this assumption let us extend the results of genetic diversity calculated on the sequenced isolates to all the GLRaV-3 isolates analysed here, even though not all were sequenced. SSCP profiles showed that the majority of the isolates had a ‘simple’ pattern in the three genomic regions, consistent with a population structure with a single, predominant variant. This situation occurred in 90 %
(RdRp and HSP70) and 85% (CP) of the isolates under study. No correlation between SSCP pattern and geographical origin was found. Analysis of population structures of those isolates having a pattern composed of several bands revealed the existence of mixed infections, which was confirmed by the high values of within-isolate genetic diversity and the inconsistent clustering in the phylogenetic tree of variants of these isolates. The fact that some of the isolates most probably originated from recombination events is consistent with the notion that recombination is an important evolutionary trait of members of the family Closteroviridae (Karasev, 2000), such as CTV (Rubio et al., 2001). However, compared with HSP-70 and CP, RdRp proved to be a less frequent recombination site. This observation, which was confirmed by the lowest values of within-isolate diversity ($D=0.0035$), entire population diversity ($D=0.0295$) and selective pressure ($dN/dS = 0.088$), suggested the existence of strong constraints in this genomic region, possibly imposed by interaction with host factors involved in virus replication. With CTV, Rubio et al. (2001) considered RdRp regions differing from those analysed by us and, contrary to our findings, reported that selective pressure was strongest in the CP cistron of the virus. They explained this finding by the existence of constraints imposed by virus–vector and/or virus–host interactions. In our case, the existence of the highest selective pressure on the RdRp gene, rather than on the HSP-70 and CP genes, is consistent with the view that the spread of GLRaV-3 depends primarily on infected propagative plant material rather than mealybug transmission, which is aspecific (Sforza et al., 2000; Gugerli, 2003) and less efficient than CTV transmission by aphids. Therefore, constraints imposed on the RdRp domain are likely to play a major role in the evolution of GLRaV-3. Finally, the lack of correlation of SSCP analysis and sequence data with the geographical origin of the isolates was consistent with the view that transmission by propagative material has a predominant role in the distribution of GLRaV-3.

**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge Dr M. Dell’Orco, Dr H.-X. Lin and L. Rubio for their suggestions and Professor F. García-Arenal for critically revising the manuscript. This study was supported by ‘Centro di Eccellenza: Genomica comparata – geni coinvolti in processi fisiopatologici in campo biomedico ed agrario (CEGBA)’.

**REFERENCES**


