Arch Virol
DOI 10.1007/s00705-010-0793-y

ORIGINAL ARTICLE

# Three genetic grapevine leafroll-associated virus 3 variants identified from South African vineyards show high variability in their 5′UTR

**A. E. C. Jooste · H. J. Maree · D. U. Bellstedt · D. E. Goszczynski · G. Pietersen · J. T. Burger**

Received: 29 April 2010 / Accepted: 27 August 2010


**Abstract** Three genetic variants of grapevine leafroll-associated virus 3 (GLRaV-3) were identified in vineyards of the Western Cape, South Africa. The GLRaV-3 variants were identified by single-strand conformation polymorphism (SSCP) profiles generated from a region amplified in ORF5. ORF5 sequence data confirmed the three genetic variant groups, and a specific SSCP profile was assigned to each variant group. The results of SSCP analysis of this region in ORF5 showed that this method gives a fast and reliable indication of the GLRaV-3 variant status of a plant, which in many instances showed mixed infections. The full genome sequence of one representative of each variant group i.e. isolates 621 (group I), 623 (group II) and PL-20 (group III), was determined by sequencing overlapping cloned fragments of these isolates. The sequences of genomic 5′ ends of these isolates were determined by RLM-RACE. Sequence alignment of the 5′UTRs indicated significant sequence and length variation in this region between the three South African variant groups. Alignment of the Hsp70h and CP gene regions of these isolates with those of isolates from elsewhere in the world, followed by phylogenetic analysis, further supported the presence of three variants of GLRaV-3 in South Africa and the presence of two or three additional variant groups elsewhere in the world.

The GenBank/EMBL/DDBJ accession numbers of the sequences reported in this paper are GQ352631, GQ352632 and GQ352633.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00705-010-0793-y) contains supplementary material, which is available to authorized users.

A. E. C. Jooste (✉) · D. E. Goszczynski · G. Pietersen
ARC-Plant Protection Research Institute, Queenswood,
Private Bag X134, Pretoria 0121, South Africa
e-mail: JoosteE@arc.agric.za

H. J. Maree · J. T. Burger
Department of Genetics, Stellenbosch University,
Private Bag X1, Matieland, Stellenbosch 7602, South Africa

D. U. Bellstedt
Department of Biochemistry, Stellenbosch University,
Private Bag X1, Matieland, Stellenbosch 7602, South Africa

## Introduction

Grapevine leafroll-associated virus 3 (GLRaV-3) is the main causative agent of grapevine leafroll disease (leafroll) in South Africa and spreads rapidly in local vineyards [18]. As GLRaV-3 is consistently associated with leafroll, it is viewed to be an important etiological agent of this economically important disease [2]. *Grapevine leafroll-associated virus 3* is the type species of the genus *Ampelovirus* in the family *Closteroviridae* [17]. Virions are flexuous, filamentous particles, about 1,800 nm long, with a positive-sense single-stranded RNA genome organised into 13 open reading frames [12].

Research worldwide has shown the existence of several molecular variants of GLRaV-3. The population structure and genetic variability of 45 GLRaV-3 isolates, from different grapevine varieties and 14 different countries, by single-stranded conformation polymorphism (SSCP) and sequence analysis of three genomic regions, RdRp, Hsp70h and coat protein (CP), was investigated [22]. The results for the RdRp and Hsp70h regions showed that 10% of the isolates analysed had mixed-variant infections, whilst 15% of the isolates had mixed infections when the CP region was analysed [22]. Multiple alignment of sequences deposited in GenBank revealed that the sequences used in the Italian study had nucleotide identities of above 90%

Published online: 10 September 2010

between isolates in the regions studied. High diversity was noted in other studies, such as the divergent strain of GLRaV-3 (GLRaV-3-Tempr; GenBank accession no. DQ314610) found in a grapevine accession in the cultivar Tempranillo from a Spanish vineyard [1]. The GLRaV-3-Tempr isolate was almost 20% divergent from the NY-1 isolate at the nucleotide level in the sequenced 3′ end of ORF1 [1]. GLRaV-3-infected juice grapes (*Vitis labruscana* 'Concord' and *Vitis labruscana* 'Niagara') from Washington State revealed nucleotide identities of 94–98% and amino acid identities of 97–98% in the Hsp70h gene of the NY-1 isolate [20]. A survey of leafroll-associated viruses showed 74.1–100% identity at the nucleotide level and 85.9–100% identity at the amino acid level between five GLRaV-3 isolates from New York and 25 isolates from other geographic regions [5]. Phylogenetic analysis of the HSP70h gene showed at least five possible variant groups in one study [5]. A study on the viral variants in the 'Waltham Cross' table grape variety revealed at least two GLRaV-3 variants; one clone (WC-HSP-2) shared 93.2% nucleotide identity with NY-1 [12], and two other clones (WC-HSP-10 and WC-HSP-28) were only 72.3% identical to NY-1 [19]. A nucleotide identity of 97.6% was reported between the Chilean isolate Cl-766 and NY-1 [4]. Another study reported significant variability between New Zealand isolates where, to date, four genetic variants have been identified [3]. A study on Portuguese grapevine varieties infected with GLRaV-3 identified five GLRaV-3 variant groups based on CP gene sequences [8].

In a previous study of South African isolates of GLRaV-3, single-strand conformation polymorphism (SSCP), restriction enzyme (RE) SSCP, cloning and sequencing techniques were used to identify two clearly divergent molecular groups of the virus [10]. The first molecular variant, represented by isolate 621, was very similar to the NY-1 isolate of Ling et al. [12], while sequence data of the second molecular variant, represented by isolate 623, were very similar to the complete genome sequence of the South African isolate GP18 [16]. The molecular divergence between these two variant groups was especially high in the 5′-terminal part (partial sequences of the 5′UTR and ORF1a of the virus, where nucleotide sequences differed by 35%). Sequence data of the remaining coding regions showed nucleotide similarities above 90% between the variant groups. The two variant groups could be distinguished by unique SSCP profiles generated from an amplified region in ORF 5 [10].

The complete genome length of a South African GLRaV-3 isolate, GP18 (EU259806), was reported to be 18,498 nt [16]. The extended length of the 5′UTR, consisting of 737 nt, differed from that reported previously by Ling et al. [12] and Engel et al. [4], where a 5′UTR of 158 nt for both isolates NY-1 (AF037268) and Cl-766 (EU344893) was described. There was 93% sequence identity between the genome sequences of isolate GP18 and NY-1 [16].

In this study, the variability of GLRaV-3 in infected plants in South African vineyards was further explored. The aim of the study was to use the SSCP technique to identify GLRaV-3 variants from infected vineyards and to obtain full genome sequences of three representatives of these GLRaV-3 variants, namely isolates 621, 623 and PL-20, including their 5′UTRs. The phylogenetic relationships of these three isolates was determined using Hsp70h and CP gene sequences, which were aligned with those of isolates obtained from elsewhere in the world.

## Materials and methods

### Virus sources

A field survey was conducted in five motherblocks in different grape-producing areas of the Western Cape, South Africa, namely, Stellenbosch, Paarl, Wellington, Rawsonville and Worcester. GLRaV-3-infected plants were selected based on different spatial distribution patterns of leafroll recorded in a survey done in motherblocks from 2001–2007 [18]. In numerous disease foci (also referred to as disease clusters), the infection point or starting point of disease spread began from a single plant. From this single plant, mealybugs transmitted the disease to adjacent plants in rows and across rows forming clusters of infection. The plants collected in this survey all represent the initial infected plant of any given focus of leafroll infection.

Based on SSCP profiles, twelve of these infected plants (3, 4, 5, 7, 12, 15, 16, 17, 20, 32, 48, and 50) were selected for further investigation (Table 1). The existence of an extended 5′UTR was studied in these twelve plants and in 57 additional field-collected GLRaV-3 isolates.

### dsRNA isolation, RT–PCR, cloning and SSCP

Isolation of dsRNA, SSCP analysis and cloning were carried out as described earlier [7]. SSCP profiles were generated in 15% polyacrylamide gels from amplified products from ORF5 of GLRaV-3, genome position 12,592–12,801 (primer pair H420/C629 in Jooste & Goszczynski [10]).

### Cloning of fragments of isolates 621, 623 and PL-20 and assembly into whole-genome sequences

Based on the SSCP results from the field survey, the complete nucleotide sequences of three GLRaV-3 isolates were determined: GLRaV-3 isolate PL-20, from a *Vitis vinifera* cv. Cabernet Sauvignon plant (Table 1), and

**Table 1** A description of the twelve plants selected for further investigation, in which cultivar, year planted and collection region are listed

| Plant number | Position description (block/row/plant no.) | Cultivar | Region | Year planted |
|---|---|---|---|---|
| 3 | 52/28/95 | Cabernet Sauvignon | Paarl | 1997 |
| 4 | 52/30/64 | Cabernet Sauvignon | Paarl | 1997 |
| 5 | 52/53/26 | Cabernet Sauvignon | Paarl | 1998 |
| 7 | 52/63/89 | Cabernet Sauvignon | Paarl | 1998 |
| 12 | 10/9/87 | Cabernet Sauvignon | Worcester | 1996 |
| 15 | 10/23/10 | Cabernet Sauvignon | Worcester | 1997 |
| 16 | 10/29/2 | Cabernet Sauvignon | Worcester | 1997 |
| 17 | 10/31/35 | Cabernet Sauvignon | Worcester | 1997 |
| 20 | 10/39/99 | Cabernet Sauvignon | Worcester | 1992 |
| 32 | 19b/6/15 | Merlot | Rawsonville | 1994 |
| 48 | 1/23/63 | Cabernet Sauvignon | Stellenbosch | 1991 |
| 50 | 1/29/83 | Cabernet Sauvignon | Stellenbosch | 1991 |

GLRaV-3 isolates 621 and 623, from *Vitis vinifera* cv. C. Sauvignon and *Vitis vinifera* cv. Ruby Cabernet, respectively [10].

dsRNA and total RNA were isolated from these samples and used for RT-PCR to clone fragments of isolates 621, 623 and PL-20 by using standard molecular techniques.

Primer design for cloning and sequencing of the fragments of the genomes of isolates 621 and 623 was done with the Genefisher program [6], using the published sequences of isolate NY-1 [12] as a template. The sequencing strategy for fragments of isolates 621 and 623 was based on the amplification, cloning and sequencing of 18 overlapping clones in the genomes of isolates 621 and 623, with product sizes ranging from 283 to 986 bp (primer sequences not shown). Cloning of PCR-amplified products was carried out using a pGEM®-T easy cloning system (Promega). At least three clones of each amplicon were sequenced in both directions. Consensus sequences of isolates 621 and 623 similar to the nucleotide length of NY-1 [12] were assembled with DNAMAN version 6 sequence analysis software (Lynnon Biosoft, 1996) from the overlapping sequences generated from the cloned fragments.

The sequence of isolate PL-20 was compiled by amplifying and sequencing ten overlapping cloned fragments (Fig. 3). Primer design was done with Genefisher [6] and OligoExplorer 1.1.0 software (http://www.genelink.com/tools/gl-oe.asp) with the GP18 isolate sequence as a template [16]. Primers designed to amplify the fragments of isolate PL-20 in ten overlapping clones are listed in Table 3. The PCR products (online resource Table 3) were ligated into the pDrive vector using a Qiagen® PCR cloning kit according to the manufacturer's instructions, and the ligation mixtures were used to transform competent DH5α cells. At least three clones of each amplicon were sequenced in both directions with the SP6 and T7 primers, and with additional primers where required. The additional sequencing primers that were designed to sequence the larger cloned fragments are listed in online resource Table 4.

To determine the sequences of the 5′ and 3′ ends of isolates 621 and 623, poly(A) tailing was attempted [11, 12]. The 5′ends of isolates 621, 623 and PL-20 were amplified using RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) with the First Choice® RLM-RACE kit (Ambion, USA) according to the manufacturer's instructions, using total RNA extracted from 2 g of phloem tissue using the CTAB method [23]. The genome-specific reverse primers used with the 5′ RLM-RACE kit to perform the 5′ RACE outer and inner PCRs were LR3 365Rev, 5′ CGTCCGCTTCACCCCTTTGG 3′ (used in the outer PCR reaction) and LR3 868, 5′ GGGTGTGAAGTCAGAT AACTT 3′ (in the inner PCR reaction). The PCR products of the 5′UTR region were ligated into the pDrive vector and sequenced as described above. Sequences were analysed and assembled with Vector NTI *v10* (Invitrogen) and BioEdit software [9], and consensus sequences of the 5′ and 3′ ends of isolates 621, 623 and PL-20 were compiled.

The 3′ ends of isolates 621 and 623 were successfully determined by polyadenylation and reverse transcription of purified dsRNA [11]. The 3′ end of isolate PL-20 was cloned with the extension of polyA cDNA.

The whole genome sequences of GLRaV-3 isolates 621, 623 and PL-20 were deposited in the GenBank database and assigned accession numbers GQ352631, GQ352632 and GQ352633. The predicted functions of the ORFs were confirmed with the conserved domain search on the NCBI website [14, 15]. The nucleotide (nt) and amino acid (aa) sequences of the ORFs of these isolates and those of the Chilean isolate C1-766 and the South African isolate GP18 were compared with those of the USA isolate NY-1, expressed as percentage sequence identity to isolate NY-1 and listed in tabular form (Table 2).

Confirmation of 5′UTR in field-collected samples

Primers were designed to amplify a 362–422-nt fragment (depending on the variant group) in the 5′UTR of selected GLRaV-3 isolates by RT-PCR. These primers were designed in conserved regions, identified by multiple sequence alignment of the 5′UTRs of isolates 621, 623, PL-20 and GP18 and designated GL3.5F (5′ TGCTC TAGTAGGATTCGAAC 3′) and GL3.342R (5′ CCCA ACACGATAAAGAGAAC 3′). PCR products were sequenced, and the sequence data were analysed to identify variant groups.

Phylogenetic analysis

In order to assess the relationship of the three South African isolates, their Hsp70h and CP gene sequences were used in a phylogenetic analysis in which the Hsp70h and CP sequences of a large number of isolates from elsewhere in the world were included. The origins of the Hsp70h and CP sequences and GenBank accession numbers are shown in additional online data in Tables 5 and 6, respectively. These Hsp70h and CP sequences of isolates 621, 623 and PL-20 as representatives of the South African variant groups I, II and III, respectively, were aligned with the Hsp70h and CP sequences downloaded from GenBank (online resource Tables 6 and 7), utilizing the software package BioEdit [9]. These sequences were aligned using the Clustal W (v 1.4) alignment function embedded within the BioEdit package, and the alignment was refined manually. Phylogenetic analyses of the aligned Hsp70h and CP sequence matrices were performed using PAUP 4.0b10 [21]. In both analyses, the GLRaV-1 Hsp70h and CP sequence (GenBank accession no. AF 195822) was used as the outgroup. A heuristic search (1,000 replicates) using TBR branch swapping with all characters weighted equally was performed to search for the shortest possible trees from both data matrices. A bootstrap analysis (1,000 replicates) using TBR branch swapping was performed to establish clade support. Branches with bootstrap values ≥75% were considered well supported, whilst values between 75% and 50% were considered moderately supported. Values below 50% were considered weakly supported and, in line with other phylogenetic analyses, were not indicated on phylograms.

## Results

Field survey, SSCP analysis

SSCP profiles obtained for the isolated dsRNA of 46 GLRaV-3-infected plants showed two distinct profiles (Fig. 1a, lanes 4 and 5), as described previously [10], as well as additional (hitherto unreported) profiles (Fig. 1a–d). The SSCP profiles consisted of ‘simple’ and ‘complex’ patterns, as described previously [22]. ‘Complex’ and atypical profiles were selected for further analysis in this study. The SSCP profiles of the twelve plants selected for further analysis are indicated by arrows in Fig. 1a–d (3, 4, 5, 7, 12, 15, 16, 17, 20, 32, 48 and 50). The SSCP profiles of plants 4 and 5 represent the two variants, isolates 621 (group I) and 623 (group II), identified previously [10].

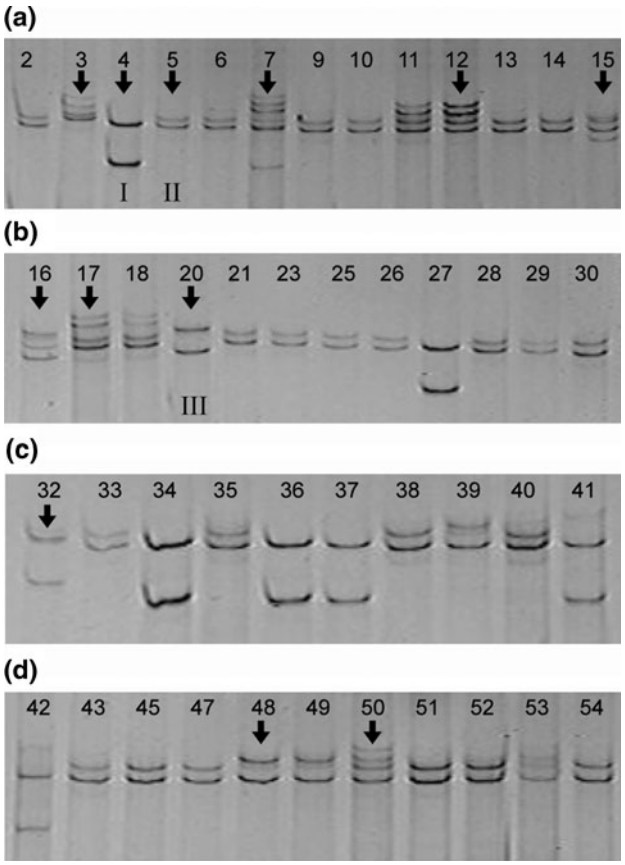


**Fig. 1** SSCP results of plants collected in vineyards in the **a** Paarl and Worcester districts, *lanes* 2–10 and 11–15, respectively, **b** Worcester and Wellington districts, *lanes* 16–20 and 21–30, respectively, **c** Rawsonville district, *lanes* 32–41 and **d** Stellenbosch district, *lanes* 42–54

Ten clones from each of the twelve plants were generated and analysed individually by SSCP. A minimum of four clones per plant were selected for sequencing of this region (indicated by circled numbers in Fig. 2). Variability in SSCP profiles between the clones was detected in most of the plants (4, 7, 12, 15, 16, 32, and 48), illustrating that combinations of variants occurred in different plants. In some of the plants (5, 17, 20 and 50), the SSCP profiles of all 10 selected clones were identical. The nucleotide sequences of 55 clones and five reference sequences (Group I: NY-1, Cl-766 and 621, Group II: GP18 and 623) were compared. Sequence alignment showed that the

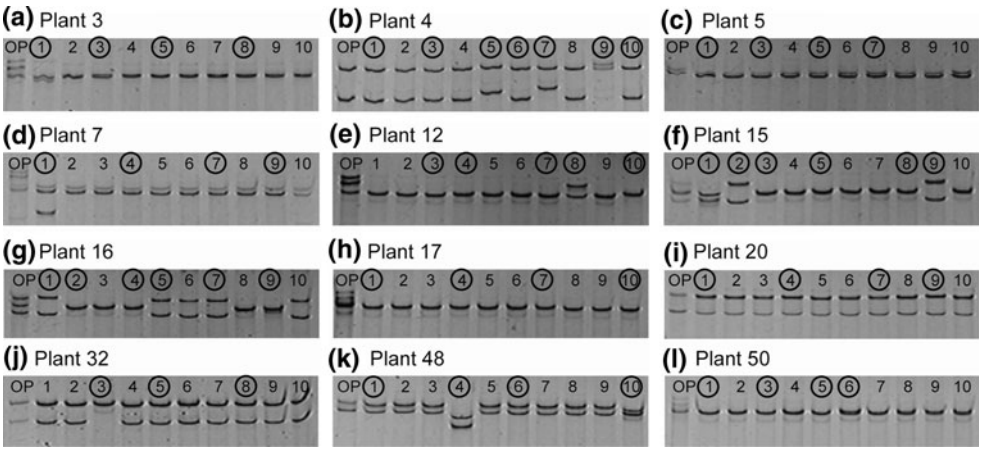


**Fig. 2** SSCP profiles of ten clones of each of the twelve plants. Clones indicated with *circles* were sequenced. The first *lane* of each gel represents the original SSCP profile (OP) of each plant

majority of the clones grouped with the variant group II isolates 623 and GP 18 (not shown). A third variant group was identified based on sequence results and distinct SSCP profiles detected in plants 15, 16, 20 and 32. The SSCP profiles shown in Fig. 2 of clones 15.2, 15.9, 16.1, 16.5, 16.7, 20.1, 20.4, 20.7, 20.9, 32.5 and 32.8 represent this third group of variants. Based on these results, the viral isolate from plant 20 (referred to as isolate PL-20) was selected as a representative of the third variant group. Sequences of other clones from plants 15, 16 and 32 clustered with sequences of clones in group I. The nucleotide sequences from each clone and the SSCP profiles (Fig. 2) of individual clones of the three variant groups correlated. SSCP profiles of clones from plant 12 were all similar, except for 12.8, which had a single nucleotide change, resulting in a different SSCP profile.

Clones derived from plant 4, i.e. 4.1, 4.3, 4.6 and 4.10, showed SSCP profiles typical for variant group I and clustered into the group that was similar to isolates 621, NY-1 and Cl-766. Clones 4.5, 4.7 and 4.9 showed atypical SSCP profiles for group I, and sequence results confirmed that these clones were similar to clones from variant group II.

The SSCP profiles determined previously for the three variant groups are indicated in the profiles of plants 4, 5 and 20 (Fig. 1a, b).

## Whole-genome sequencing results and variability in the 5′UTR

Whole genome sequences of isolates 621, 623 and PL-20 were successfully generated, each representing one of the variant groups identified in the SSCP studies. The whole genome organisation of the three GLRaV-3 variants was similar to that described previously for isolates NY-1, Cl-766 and GP18 [4, 12, 16]. Similar to isolate GP18, all three isolates contained an extended 5′UTR [16]. ORF1a of isolates 621 (group I) and 623 (group II) started at nucleotide position 738 on the respective genomes and at position 673 on the PL-20 genome (group III). In this region, the methyltransferase (MET), AlkB (2OG-FeII) and helicase (HEL) domains were found in all three isolates. The other domains were similar to those of the NY-1, GP18 and Cl-766 isolates [4, 12, 16] and included the two intergenic regions from positions 9,058–9,286 and 9,443–10,508.

The variation of the three variant groups compared to isolate NY-1 is shown in Table 2. The percent sequence identity in the nt and aa sequences in ORF1a showed no clear differentiation between group I and II variants. In ORF2, PL-20 (variant group III) showed an 87.8% difference at the nt level from NY-1, but the predicted aa sequence was 100% similar to NY-1. In ORF2, the aa prediction for variant group II isolates GP18 and 623 differed by almost 20% from that for the group I and III variants. The genomic regions near the 3′UTR, namely ORF10, ORF11 and ORF12, showed the most variation between variants in the nt and aa sequences, and the group III variant showed 80%, 67% and 78% aa identity, respectively, to the NY-1 isolate in these regions.

The 5′UTR sequences of isolates 621, 623 and PL-20 were successfully sequenced by RLM-RACE. The full genome length of isolates 621 and 623 was 18,498 nt with a 5′UTR length of 737 nt, identical to that reported for the South African GP18 isolate [16]. Multiple sequence alignment of the 5′UTR (Fig. 3) indicated significant variation amongst the three variant groups, with two regions differing substantially amongst them. In the first region, two insertions were observed in the sequence of the group I and III variants (isolates 621 and PL-20) compared to the group II variants, in which a 61-nt insertion between positions 205 and 267 and an insertion of 4 nt between positions 273 and 278 was found. In the second region, deletions between positions 567 and 661 in variant group I (isolate 621) of 65 nt and in variant group III (isolate PL-20) of 132 nt were found. The sequence data of the 5′UTR

**Table 2** Percent nucleotide (nt) and amino acid (aa) sequence identity of the GLRaV-3 isolates, Cl-766 (EU344893), 621 (GQ352631), GP18 (EU259806), 623 (GQ352632) and PL-20 (GQ352633), representing the three variant groups, as compared to isolate NY-1 (AF037268)

| GLRaV-3 isolate | ORF1a Met/Hel nt/aa | ORF1b RdRp nt/aa | ORF2 p6 nt/aa | ORF3 p5 nt/aa | ORF4 Hsp70h nt/aa | ORF5 Hsp90 h nt/aa | ORF6 CP nt/aa | Variant group |
|---|---|---|---|---|---|---|---|---|
| Cl-766 | 95.7/95.5 | 99.0/99.1 | 98.7/98.0 | 97.1/100 | 98.8/98.4 | 99.2/99.0 | 99.0/99.0 | I |
| 621 | 96.2/96.5 | 99.4/100 | 98.7/98.0 | 97.1/100 | 98.8/98.7 | 99.2/99.2 | 99.3/99.0 | |
| GP18 | 94.1/95.3 | 95.4/97.6 | 90.4/80.4 | 93.5/97.8 | 94.6/96.9 | 92.8/93.2 | 92.5/94.6 | II |
| 623 | 93.4/95.6 | 90.3/95.7 | 91.0/82.4 | 93.5/97.8 | 94.8/97.8 | 92.9/93.4 | 92.8/95.2 | |
| PL-20 | 87.5/90.9 | 92.3/100 | 87.8/100 | 92.0/91.1 | 90.0/94.7 | 90.8/91.3 | 91.5/96.5 | III |
| | ORF7 dCP nt/aa | ORF8 p21 nt/aa | ORF9 p19.6 nt/aa | ORF10 p19.7 nt/aa | ORF11 p7 nt/aa | ORF12 p4 nt/aa | 3′UTR nt | |
| Cl-766 | 99.2/98.1 | 98.9/98.4 | 99.4/99.4 | 98.7/97.2 | 98.2/97.2 | 93.4/91.7 | 97.5 | I |
| 621 | 99.1/98.3 | 99.5/100 | 99.6/99.4 | 98.7/96.6 | 97.3/94.4 | 93.4/91.7 | 96.8 | |
| GP18 | 92.1/89.9 | 93.5/97.3 | 91.6/88.7 | 90.6/86.0 | 91.0/88.9 | 97.3/96.7 | 97.5 | II |
| 623 | 92.3/90.6 | 93.7/97.3 | 91.6/88.7 | 89.8/84.9 | 91.9/88.9 | 97.8/95.0 | 97.1 | |
| PL-20 | 88.3/89.3 | 90.1/94.6 | 90.8/89.3 | 83.0/79.9 | 77.5/66.7 | 84.2/78.3 | 94.9 | III |

was confirmed with clones generated from the RLM-RACE reaction as well as clones from primer set GL3.5 and LR 365Rev, which amplified a product of 941 bp from positions 5-946 on the GP18 genome.

The variation in the 5′UTR sequences presented here correlated with the three variant groups shown by SSCP profiles generated from amplicons of ORF5, and sequence data of the 5′UTRs of the three variants (621, 623 and PL-20) reflected the three variant groups. The 5′UTR is highly variable between the GLRaV-3 molecular variants compared to the rest of the genome. Intergroup variation between the group I (621) and group II (623) variants in the 5′UTR was as much as 30%, while variant group II (623) and group III (PL-20) differed by 22%. Group I (621) and III (PL-20) varied by 33%.

## Confirmation of 5′UTR sequence results

Amplified RT-PCR products obtained from the twelve selected plants, using primer pair GL3.5F and GL3.342R, were of the expected sizes for the different variant groups, i.e. 422 bp for isolates of variant groups I and III and a smaller product, 362 bp, for isolates of group II variants. These PCR products were sequenced, and nucleotide sequences of 382 nt and 317 nt (excluding primer sequences) were compiled. The multiple sequence alignment of the twelve plants showed an insertion of 65 nt for group I and III isolates (plants 4 and 20).

In addition to these plants, 57 GLRaV-3-infected plants, collected randomly from 10 motherblocks in different grapevine-growing regions of the Western Cape, were analysed in the same area of the 5′UTR (not shown). Fifty-two of these plants grouped in the group II variant clade with GP18 and 623, four of these plants grouped with the group I variant 621, and one plant grouped in the group III clade with PL-20.

## Phylogenetic analysis of the Hsp70h and CP gene sequences

The alignment of the Hsp70h gene sequences of the isolates 621, 623 and PL-20 with the sequences from GenBank resulted in a matrix that was 933 bp in length. Phylogenetic analysis revealed that 213 of the 933 characters were constant, 545 (58%) characters were found to be parsimony-uninformative, while 175 (19%) characters were parsimony-informative. The heuristic search retrieved 8,960 trees with a tree length of 988, a consistency index (CI) of 0.906 and a retention index (RI) of 0.786. As many of the GenBank depositions of CP sequences were not complete, the alignment that was used in the phylogenetic analysis was restricted to the first 499 bp of the CP gene. The analysis revealed that 174 of the 499 characters were constant, 228 (45.7%) characters were found to be parsimony-uninformative, and 97 (19.4%) characters were parsimony-informative. The heuristic search retrieved 16 trees with a tree length of 458, with a CI of 0.856 and a RI of 0.876. The trees that were retrieved were used to generate a strict consensus tree. One of the shortest trees retrieved for both regions is shown in Fig. 4. This representation of the phylogenetic analysis was chosen because it shows branch lengths, which indicate the actual numbers of differences between the sequences that were included in the analysis. Those nodes that collapsed in the strict consensus tree are


**Fig. 3** A schematic representation of the genome organisation of GLRaV-3 and the position of the ten overlapping clones of isolate PL-20, illustrating the sequencing strategy (*top*). The multiple alignments (*bottom*) of the complete 5′UTR sequences of isolates 621, 623, PL-20 and GP18 is shown below this. This reveals a first deletion in isolates 623 and GP18 (group II) of 61 and 4 nt, a second deletion of 65 nt in isolate 621 and PL-20 (group I and III) and a further deletion of 51 nt in the isolate PL-20 sequence (group III)

indicated with arrows. The phylogeny confirms that, based on strong bootstrap support of key nodes, three distinct variants of GLRaV-3 occur in South Africa.

The phylogenetic position of the three South African variants in the Hsp70h region (Fig. 4a), in relation to GLRaV-3 from other geographic regions, showed that the group I variant, isolate 621, grouped with the NY-1 [12] and Cl-766 [4] isolates with very little variation between these isolates. The variant I clade included sequences from isolates from the USA and China (NY-1, N1-1, C6-1, C4-1,

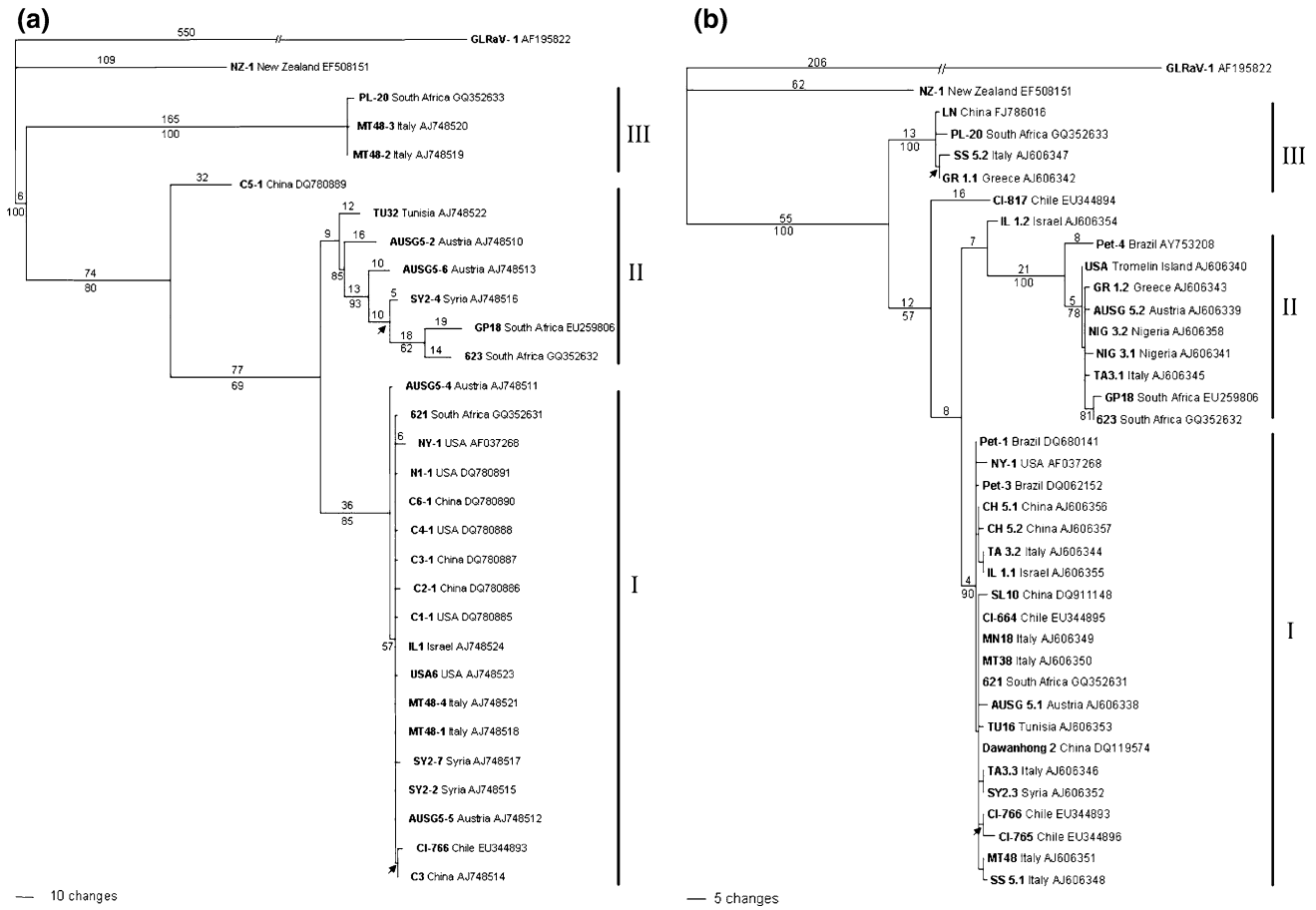


**Fig. 4** One of the shortest trees of a heuristic search performed on the aligned Hsp70h (**a**) and coat protein (**b**) sequence matrixes. Branch lengths that are longer than 5 are shown *above* branches and bootstrap values are indicated *beneath*. Branches that collapsed in the strict consensus tree are shown with *arrows*. The branch length of the outgroup is not drawn to scale. The three variant groups studied here are indicated as groups I, II and III

C3-1, C2-1, C1-1, USA6, C3), Chile (Cl-766), Israel (IL1), Italy (MT48-4, MT48-1), Syria (SY2-2, SY2-7), and Austria (AUSG5-5). The group II variant, isolate 623, was sister to isolate GP18 [16]. Other accessions in the group II clade were from Austria (AUSG5-2, AUSG5-6), Syria (SY2-4) and Tunisia (TU32). Branch lengths were longer in this clade, indicating greater sequence heterogeneity, and even the two South African isolates showed some sequence divergence. The South African variant group III, as represented by isolate PL-20, grouped with two accessions from Italy (MT48-2, MT48-3). The branch lengths in this clade indicated that there was very little variation between these isolates. Two isolates, C5-1 and NZ-1, each resolved in isolated positions on their own in the phylogeny. They appear to represent two further variants of GLRaV-3 in addition to the three variants present in South Africa and elsewhere in the world.

Phylogenetic analysis of the CP gene region (Fig. 4b) showed that the three South African isolates also grouped into three well-supported monophyletic clades, as in the Hsp70h analysis. In the CP analysis, the majority of isolates grouped in the variant group I clade, together with isolate 621 and NY-1. This group included isolates from Brazil (Pet-1, Pet-3), China (SL10, CH 5.1, CH 5.2, Dawanhong 2), Italy (TA 3.2, TA 3.3, MN18, MT38, MT48, SS 5.1), Israel (IL1.1), Chile (Cl-664, Cl-766, Cl-765), Austria (AUSG 5.1), Tunisia (TU16), and Syria (SY 2.3). In the group II clade, isolates 623 and GP18 grouped together with isolates from Italy (TA 3.1), Nigeria (NIG 3.1, NIG 3.2), Austria (AUSG 5.2), Greece (GR 1.2), Tromelin Island (USA) and Brazil (Pet-4). Isolates that grouped with the third variant, isolate PL-20, were from China (LN), Italy (SS 5.2) and Greece (GR 1.1). The New Zealand isolate NZ-1, the Chilean isolate Cl-817 and the Israeli IL 1.2 isolate resolved in isolated positions, but with poor bootstrap support.

## Discussion

As in another study [22], the SSCP technique proved a useful tool for studying the genetic diversity of South

African GLRaV-3 isolates. In our study, the results showed that SSCP analysis of a region in ORF5 gave a fast and reliable indication of the GLRaV-3 variant status of a plant. This region in ORF5 was successfully amplified in all GLRaV-3-infected plants (>200 plants) collected during two local surveys in red grapevine cultivars (data not shown). Based on this gene region, three main GLRaV-3 variant groups were identified in South African vineyards. The three variant groups were confirmed by sequencing, and distinct SSCP profiles could be assigned to each variant group. SSCP and sequencing results showed complete correlation in all of the genomic regions studied. The SSCP profiles detected for group II variants were heterogeneous, indicating greater sequence complexity within this group. In some of the analysed plants, "complex" SSCP profiles were detected that, in most cases, contained a single dominant GLRaV-3 variant. Plants infected with a mixture of variants were also detected. This information is in agreement with a previous study [22] in which combinations of variants in the same isolate were also detected.

The variation detected by SSCP was confirmed by sequencing the whole genomes of three isolates representative of the three variant groups. The three GLRaV-3 variant groups correlated with previous reports that more than one group occurred in South Africa. Prosser et al. [19] reported that the WC-HSP-2 isolate from South Africa was 93.2% identical to the NY-1 isolate. Similarly, Maree et al. [16] reported that isolate GP18 was 93% similar to NY-1. Jooste & Goszczynski [10] found that the percentage identity between isolates 621 and 623 was 91.8–96.2% in ORFs 4-7 and that the most and least divergent fragments were from ORF5 and ORF7 (91.8–92.3% and 96.0–96.2%).

Phylogenetic analysis of the aligned sequences of the Hsp70h gene region of these South African isolates and those of isolates from elsewhere in the world further confirmed that three variant groups of GLRaV-3 occurred in South Africa and that these variants also occur elsewhere in the world. The two further isolates, one from New Zealand (NZ-1) and one from China (C5-1), appear to represent two further variants. Five GLRaV-3 variant groups were also identified by Fuchs et al. [5] in their phylogenetic analysis of partial Hsp70h sequences available in GenBank. The five variant groups in the Fuchs study were represented by isolates NY-1, GP18 and MT48-2, found in our groups I, II and III, respectively, and also C5-1 and NZ-1. Two isolates, C5-1 and NZ-1, each also resolved in isolated positions on their own in the phylogenetic analysis in the Fuchs study. The phylogenetic analysis of the CP largely corroborated the identification of the three variant groups present in South Africa in spite of the fact that most of the CP sequences from elsewhere in the world, with the exception of NY-1, Cl-766 and GP18, were not from the same source as the Hsp70h gene sequences. Again the NZ-1 isolate grouped in an isolated position, confirming the unique identity of a further variant in New Zealand, but the isolated position of two further isolates, Cl-817 from Chile and IL 1.2 from Israel, may prove to present two further variant groups of GLRaV-3. Thus, the phylogenetic analysis of the CP region indicates six possible variants instead of the five indicated by the phylogenetic analysis of the Hsp70h gene. Further whole-genome sequencing of viral isolates from other regions of the world will have to be undertaken to establish whether, in addition to the three isolates that are represented in South Africa, two or three further GLRaV-3 variants occur.

The high divergence detected in the 5′UTR between isolates 621 and 623 supports a previous report in which a 417-nt fragment, which included 88 nt of the 5′UTR and an adjacent 329 nt of ORF1a, only showed a 81.8% identity between isolates 621 and 623 [10]. In the present study, the sequence of the 5′UTR of GLRaV-3 showed three clear variant groups of GLRaV-3 that correlated with the identification of the three variant groups by SSCP profiles, whole genome sequencing and phylogenetic analysis of South African isolates. The third variant, represented by PL-20, is especially interesting, as it contains a shorter 5′UTR, resulting in a genome of 18,433 nt, 65 nt shorter than the sequences of isolates 621 (variant group I), 623 (variant group II) and GP18 (variant group II). A similar study on variability of citrus tristeza virus (CTV) isolates showed that the 5′UTR of CTV could be used for the classification of sequences into three groups [13].

RT-PCR of all South African field samples tested thus far confirmed the presence of an extended 5′UTR for GLRaV-3, as described previously [16]. Sequence results of the 5′UTR from 69 of these field collected isolates clearly indicated that variability between isolates of the same variant group is low, but higher between isolates from different variant groups.

The possibility of potential folding and the significance of the 5′UTR were investigated in this study. Bioinformatic analysis of the 5′UTRs, i.e. conserved functional motifs, ORF analysis and predicted secondary structure of these three molecular variants, were unable to predict a possible function for these 5′UTRs.

Three genetic variants of GLRaV-3, represented by isolates 621, 623 and PL-20, were identified from South African vineyards. We conclude that the 5′UTR of the GLRaV-3 genome is a key region that can be used to study the variation amongst variants of the virus. The full-length sequences of these three isolates and partial 5′UTR sequence data from field isolates confirm the existence of an extended 5′UTR [16]. Currently, these four sequences represent the only complete full-length sequences of GLRaV-3. Sequencing of the 5′UTRs of other GLRaV-3

variants occurring elsewhere in the world may therefore be of particular value in future studies of the variation in GLRaV-3 and warrants further investigation.

**Acknowledgments** The research in this study was funded by the South African industry partner, Winetech, in collaboration with the National Research Foundation's THRIP programme and ARC-PPRI. We acknowledge the assistance of staff at the ARC-PPRI and Stellenbosch University who provided technical assistance. A special word of thanks to Elsa van Niekerk, who assisted with the graphics of the paper.

## References

1. Angelini E, Bertazzon N, Bazzo I, Borgo M (2006) Molecular characterization of a divergent strain of grapevine leafroll-associated virus 3. In: Extended abstracts 15th Meeting of ICVG. Stellenbosch, South Africa, pp 148–150
2. Boscia D, Greif C, Gugerli P, Martelli GP, Walter B, Gonsalves D (1995) Nomenclature of grapevine leafroll-associated putative closteroviruses. Vitis 34:171–175
3. Chooi KM, Pearson MN, Cohen D, Pong JCH (2009) Sequence variation in *Grapevine leafroll-associated virus 3* (GLRaV-3) New Zealand isolates. In: Extended abstracts of the 16th Meeting of ICVG. Dijon, France, pp 290–291
4. Engel EA, Girardi C, Escobar PF, Arredondo V, Domininguez C, Perez-Acle T, Valenzuela PDT (2008) Genome analysis and detection of a Chilean isolate of *Grapevine leafroll associated virus-3*. Virus Genes 37:110–118
5. Fuchs M, Martinson TE, Loeb GM, Hoch HC (2009) Survey for the three major leafroll disease-associated viruses in Finger Lakes vineyards in New York. Plant Dis 93:395–401
6. Giegerich R, Meyer F, Sleiermacher C (1996) GeneFisher—software support for the detection of postulated genes. Proc Int Conf Intell Syst Mol Biol 4:68–77
7. Goszczynski DE, Jooste AEC (2002) The application of single-strand conformation polymorphism (SSCP) technique for the analysis of molecular heterogeneity of grapevine virus A. Vitis 41:77–82
8. Gouveia P, Esteves F, Teixeira Santos M, Fonseca F, Eiras-Dias JE, Nolasco G (2009) Assessment of GLRaV-3 variants occuring in Portuguese grapevine varieties according to the coat protein gene. In: Extended abstracts of the 16th meeting of ICVG. Dijon, France, pp 316–317
9. Hall TA (1999) BioEdit: a user-friendly biological sequence and genomic alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98
10. Jooste AEC, Goszczynski DE (2005) Single-strand conformation polymorphism (SSCP), cloning and sequencing reveal two major groups of divergent molecular variants of grapevine leafroll-associated virus 3 (GLRaV-3). Vitis 44(1):39–43
11. Ling KS, Zhu HY, Drong RF, Slightom JL, McFerson JR, Gonsalves D (1998) Nucleotide sequence of the 3′-terminal two-thirds of the grapevine leafroll-associated virus 3 genome reveals a typical monopartite closterovirus. J Gen Virol 79:1299–1307
12. Ling KS, Zhu HY, Gonsalves D (2004) Complete nucleotide sequence and genome organization of grapevine leafroll-associated virus 3, type member of the genus Ampelovirus. J Gen Virol 85:2099–2102
13. López C, Ayllón MA, Navas-Castillo J, Guerri J, Moreno P, Flores R (1998) Molecular variability of the 5′- and 3′-terminal regions of *Citrus tristeza virus* RNA. Phytopathology 88:685–691
14. Marchler-Bauer A (2009) CDD: specific functional annotation with the conserved domain database. Nucleic Acids Res 37:205–210
15. Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32:327–331
16. Maree HJ, Freeborough M-J, Burger JT (2008) Complete nucleotide sequence of a South African isolate of *grapevine leafroll-associated virus 3* reveals a 5′UTR of 737 nucleotides. Arch Virol 153:755–757
17. Martelli GP, Agranovsky AA, Bar-Joseph M, Boscia D, Candresse T, Coutts RHA, Dolja VV, Falk BW, Gonsalves D et al (2002) The family *Closteroviridae* revised. Arch Virol 147(10):2039–2044
18. Pietersen G (2004) Spread of grapevine leafroll disease in South Africa—a difficult, but not insurmountable problem. Wynboer—a technical guide for wine producers. June 2004. URL: http://www.wynboer.co.za/recentarticles/articles.php3
19. Prossser SW, Goszczynski DE, Meng B (2007) Molecular analysis of double-stranded RNAs reveals complex infection of grapevines with multiple viruses. Virus Res 124:151–159
20. Soule MJ, Eastwell KC, Naidu RA (2006) First report of *Grapevine leafroll associated virus-3* in American *Vitis* spp. Grapevines in Washington State. Plant Dis 90:1461
21. Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland
22. Turturo C, Salderelli P, Yafeng D, Digiaro M, Minafra A, Savino V, Martelli GP (2005) Genetic variability and population structure of *grapevine leafroll-associated virus 3* isolates. J Gen Virol 86:217–224
23. White EJ, Venter M, Hiten NF, Burger JT (2008) Modified Cetyltrimethylammonium bromide method improves robustness and versatility: the benchmark for plant RNA extraction. Biotechnol J 3:1424–1428