ANNOTATED SEQUENCE RECORD

Complete nucleotide sequence of a South African isolate of *grapevine leafroll-associated virus* 3 reveals a 5′UTR of 737 nucleotides

H. J. Maree · M.-J. Freeborough · J. T. Burger

Grapevine leafroll-associated virus 3 (GLRaV-3) is the type species of the genus *Ampelovirus* (family *Closteroviridae*) [1]. It is an economically important virus that is known to only infect *Vitis* spp. and that has a negative impact on the wine and table grape industries worldwide. To date, there has been only one report that claims the complete nucleotide sequence of GLRaV-3, by Ling et al. in 2004 (isolate NY-1, AF037268) [2]. In that report the 5′ untranslated region (UTR) was found to be 158 nt in length. Here we report that the sequence of GLRaV-3, isolate GP18[1], has a 5′UTR of 737 nt. This extended UTR was found in all other South African isolates of GLRaV-3 that were analysed.

Grapevine material (Cabernet Sauvignon) was harvested in the Somerset West wine-producing region in South Africa from a monitored vineyard. RNA of isolate GP18 was extracted from a vine that displayed symptoms for the first time (Pietersen, pers comm). GP18 double-stranded RNA (dsRNA) was extracted from phloem tissue of wooded canes using an adapted cellulose (CF11) column method first described by Hu et al. [3]. RT-PCR was performed with primer sets designed to cover a large portion of the genome (nucleotides 1,835–17,905 of AF037268) in 10 overlapping clones (Fig. 1). Amplicons were cloned and sequenced and a consensus sequence was generated using BioEdit [4]. Primers were not included in the consensus sequence assembly of these overlapping sequences.

Poly(A) tailing was performed on dsRNA in an attempt to reach the 5′- and 3′- terminal ends (Fig. 1) [5]. The 3′-terminus of GP18 was found to be similar to that of the NY-1 isolate. However, we were unable to reach the reported 5′-end of the NY-1 isolate and consistently found amplicons that started at +50 nt. After further experimentation with PCR conditions, a range of amplicons were generated that extended beyond the NY-1 sequence's 5′-end. A possible explanation for this inconsistency, using poly(A) tailing, might be the high uracil content in the 5′-region upstream of the +50 site in the NY-1 strain that leads to stretches of adenines on the negative strand, which could act as priming sites for the oligo(dT) primer and lead to fragments shorter than the true genomic size.

To determine the 5′ end of the GP18 genome, total RNA was extracted from grapevine phloem tissue and subjected to RNA ligase-mediated rapid amplification of cDNA Ends (RLM-RACE) using the FirstChoice® RLM-RACE kit (Ambion, USA) as per the manufacturer's instructions. The amplicon generated for the 5′-end of the GP18 genome was larger than expected (Fig. 1). It was cloned and four clones sequenced. The 5′UTR was found to extend beyond the sequence reported for the NY-1 isolate [2]. The adapter ligation reaction was repeated on the same total RNA treated with calf intestine alkaline phosphatase (CIP) and tobacco acid pyrophosphatase (TAP), and a further five clones were sequenced. All nine clones showed the first 365 nucleotides of the NY-1 sequence and an additional 579 nucleotides upstream of the 5′-end.

The efficacy of RLM-RACE to determine the 5′-termini of multiple ssRNA viruses from total RNA in a single reaction was also investigated. RLM-RACE was performed on total RNA extracted from three grapevine plants (K1, K5, K6) that contained a mixture of viruses (GLRaV-2, -3 and grapevine rupestris stem pitting-associated virus

H. J. Maree · M.-J. Freeborough · J. T. Burger (✉)
Department of Genetics, Stellenbosch University,
Private Bag X1, Matieland 7602, South Africa
e-mail: jtb@sun.ac.za

---

[1] The GenBank accession number for the sequence reported in this paper is EU259806.
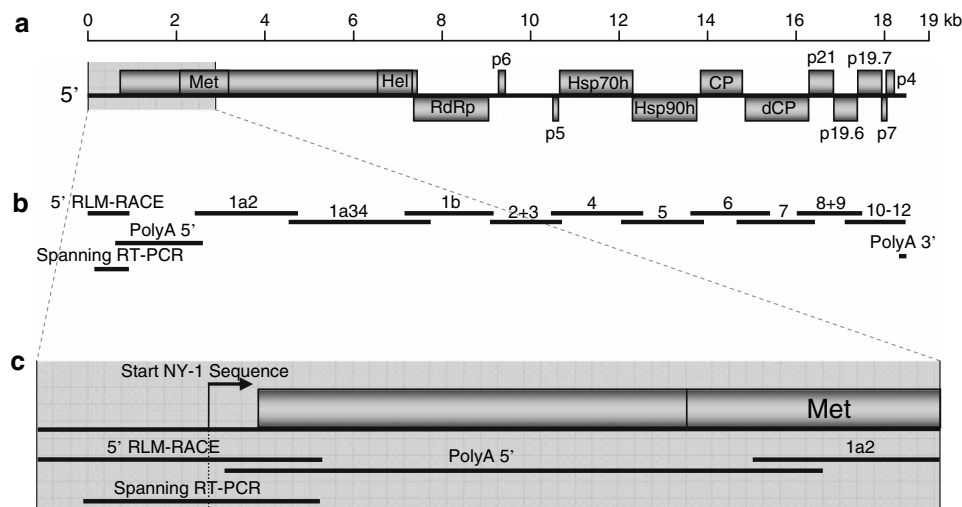
Fig. 1 **a** Schematic representation of the genome organisation of GLRaV-3 GP18 (drawn to scale). **b** Lines indicate the regions cloned and sequenced by different techniques. Lines 1a2, 1a34, 1b, 2 + 3, 4, 5, 6, 7, 8 + 9, 10–12 are representative of the clones generated by RT-PCR to sequence the majority of the genome. "PolyA 5'" and "PolyA 3'" represent the clones generated using poly A tailing and "5′ RLM-RACE" show the area amplified using RLM-RACE. "Spanning RT-PCR" represents the area generated by RT-PCR to indicate that other isolates also has the extended 5′UTR. **c** Enlargement of the 5′ area also shows the start of the NY-1 sequence compared to the GP18 sequence

(GRSPaV)) to determine their 5′-termini in the same reaction. Sequencing results indicated that GLRaV-2 (AY881628) and GRSPaV (AY881626) have the same 5′-ends as previously reported. However, GLRaV-3 showed the same extended 5′-end that was found in the GP18 isolate. The absence of similar extended 5′-regions for GLRaV-2 and GRSPaV confirms that the GP18 result is not an artefact of the RLM-RACE reaction.

Moreover, RT-PCR was used to amplify a fragment spanning the 5′-end of the NY-1 sequence for additional samples (K1, K2, GP16 and KK1; Fig. 1). A PCR product of 786 nt was amplified, and sequence results indicated that these four samples contained the same extended 5′UTR.

The complete genome of GLRaV-3 isolate GP18 is 18,498 nucleotides long and has a sequence identity of 93% with the NY-1 sequence over nucleotides

Table 1 Position of untranslated regions (UTRs) and open reading frames (ORFs) on the isolate GP18 (EU259806) sequence

UTR or ORF	Position on the genome	Size (nt)	% Sequence identity to NY-1 isolate	
			nt	aa
5′UTR	1–737	737	80	–
1a (Met/Hel)	738–7,451	6,714	94	95
1b (RdRp)	7,369–9,066	1,698	95	97
2	9,287–9,442	156	90	80
3	10,509–10,646	138	93	98
4 (Hsp70 homologue)	10,665–12,314	1,650	94	97
5 (Hsp90 homologue)	12,307–13,758	1,452	92	93
6 (CP)	13,848–14,789	942	92	95
7 (dCP)	14,852–16,285	1,434	92	90
8	16,296–16,853	558	93	97
9	16,850–17,383	534	91	89
10	17,390–17,929	540	90	86
11	17,932–18,042	111	90	89
12	18,039–18,221	183	97	97
3′UTR	18,222–18,498	277	97	–

Percentage nucleotide and amino acid sequence identity to isolate NY-1 (AF037268) are given

580–18,498. See Table 1 for nucleotide and amino acid sequence identities between GP18 and NY-1. The GP18 sequence is 579 nt longer than NY-1 at the 5′-end and has a 5′UTR of 737 nt. The extended 5′UTR has an adenine/uracil content of 68.4%, with a high uracil content of 48.5%. The only other virus of the genus *Ampelovirus* that has been completely sequenced is little cherry virus 2 (LChV-2, AF531505), whose genome was found to contain a region of 539 nt, 5′ of the ORF1a. The 5′-region of LChV-2 and 5′UTR of GLRaV-3 GP18 is much larger than the 5′UTRs of other members of the family *Closteroviridae*: GLRaV-2, closterovirus (AY881628) 105 nt; BYV, closterovirus (AF190581) 107 nt; CTV, closterovirus (DQ272579) 107 nt and LIYV, crinivirus (NC_003617) 97 nt. The 5′UTR of GLRaV-3 GP18 contains two small ambisense ORFs with no similarity to the LChV-2 ORF0, and it is possible that these ORFs are not expressed [6]. This leaves one to speculate about the function of such a large 5′UTR, which will be further investigated. A further significant difference between the sequence of GP18 and NY-1 is the 82-nt overlap between ORF1a and ORF1b. In the GP18 sequence ORF1b can still be expressed as a +1 frameshift.

The putative function for some of the ORFs was predicted using the Pfam 22.0 domain search software [7]. The domains predicted by this software were similar to those previously described for the NY-1 isolate, with the only differences being the lack of a p-protease domain and the presence of a 2OG-Fe(II) oxygenase domain (aa 1,938–2,199) in ORF 1a. Further investigation is needed to elucidate this discrepancy.

The function of this large 5′UTR is not known, however its role in replication cannot be discounted and needs to be further investigated. Furthermore, analysis of the expression of ORF1b needs to be evaluated to determine the mechanism by which this protein is translated.

References

1. Martelli GP, Agranovsky AA, Bar-Joseph M, Boscia D, Candresse T, Coutts RHA, Dolja VV, Falk BW, Gonsalves D, Jelkmann W, Karasev AV, Minafra A, Namba S, Vetten HJ, Wisler GC, Yoshikawa N (2002) The family C*losteroviridae* revised. Arch Virol 147(10):2039–2044
2. Ling K-S, Zhu H-Y, 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
3. Hu JS, Gonsalves D, Teliz D (1990) Characterization of closterovirus-like particles associated with leafroll disease. Phytopathology 128:1–14
4. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98
5. Meng B, Li C, Goszczynski DE, Gonsalves D (2005) Genome sequences and structures of two biologically distinct strains of *Grapevine leafroll-associated virus* 2 and sequence analysis. Virus Genes 31(1):31–41
6. Rott ME, Jelkmann W (2005) Little cherry virus-2: sequence and genomic organization of an unusual member of the *Closteroviridae*. Arch Virol 150:107–123
7. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A (2006) Pfam: clans, web tools and services. Nucleic Acids Res Database Issue 34:D247–D251