The nucleotide sequence and genomic organization of grapevine virus B

P. Saldarelli, A. Minafra and G. P. Martelli

Dipartimento di Protezione delle Piante, Università degli Studi and Centro di Studio del CNR sui Virus e le Virosi delle Colture Mediterranee, Via Amendola 165/A, 70126 Bari, Italy

Grapevine virus B (GVB) is a tentative member of the genus Trichovirus. The 5'-terminal region of the RNA genome of GVB comprises 5437 nucleotides and has been sequenced by the dideoxynucleotide chain termination method. Evidence was obtained that the RNA is capped. Two putative open reading frames (ORFs) were identified. ORF 1 coded for a 194.7 kDa polypeptide with conserved motifs of replication-related proteins of positive-strand RNA viruses (i.e. methyltransferase, helicase and RNA-dependent RNA polymerase, in that order from the N to the C terminus). ORF 2 encoded a 20 kDa polypeptide that did not show any significant sequence homology with protein sequences from the databases. The biological function of this polypeptide was not determined. Although the 20 kDa product was expressed as a fusion protein with glutathione S-transferase in Escherichia coli and an antiserum produced, it could not be identified in GVB-infected plant tissue extracts. The GVB genome had the same size as that of apple chlorotic leaf spot virus (ACLSV), the type species of the genus Trichovirus, but differed substantially in the number (five compared to three), size and order of genes. Differences existed also in the extent of sequence homology between polymerases, which did not cluster together in tentative phylogenetic trees. The results of this study show that definitive and tentative trichovirus species differ molecularly to an extent that may warrant a taxonomic revision of the genus.

Introduction

Grapevine virus B (GVB), a possible member of the genus Trichovirus (Martelli et al., 1994), has filamentous particles 800 nm long and a single type of coat protein (CP) subunit with a molecular mass of 21.6–23 kDa, according to whether estimates are from the amino acid sequence of the CP cistron (Minafra et al., 1994) or from electrophoretic mobility of purified virus preparations (Boscia et al., 1993). The genome is a single-stranded RNA of about 7600 nucleotides in size (Boscia et al., 1993). GVB may be the causal agent of corky bark, a major disease of the grapevine (Boscia et al., 1993, Garau et al., 1995; Greif et al., 1995; Bonavia et al., 1996), and is transmitted in nature by several species of the pseudococcid mealybug genera Pseudococcus and Planococcus (Boscia et al., 1993; Garau et al., 1995).

As reported in the present paper, sequencing of GVB genomic RNA was completed, and the potential expression products of the 5' region compared with those of other RNA plant viruses.

Methods

Viral RNA and total nucleic acid extraction. The CVB isolate was the same as that used in previous studies (Boscia et al., 1993; Minafra et al., 1994). It was propagated in Nicotiana occidentalis and purified as described (Boscia et al., 1993). Viral RNA was extracted from purified virus preparations and total nucleic acids from infected plants according to Diener & Schneider (1966) and White & Kaper (1989), respectively.
**cDNA cloning and sequencing.** A library of 25 clones previously produced by random primed cDNA synthesis (Saldarelli et al., 1993) was used for sequencing. Additional clones, corresponding to non-represented regions of the genomic RNA, were obtained by cDNA synthesis primed by the following specific oligonucleotides: GB1, 5' TCTATCAGCTTTGCGG 3', nt 4147-4162; GB2, 5' TGTCAATGTACCCAGTTGGC 3', nt 3032-3051; GB5, 5' TCGATTAGCTGGTGGCAG 3', nt 1687-1706. Double-stranded cDNA was cloned in the Smal site of pUC18 and recombinant plasmids were used to transform competent *Escherichia coli* DH5α cells (Sambrook et al., 1989).

Dideoxynucleotide chain termination sequencing was performed on denatured plasmid templates with T7 DNA polymerase (Sequenase; US Biochemical) and [35S]-dATP, according to the manufacturer's instructions. Whenever necessary, subclones were generated from larger clones by digestion with appropriate enzymes and cloned in the polylinker of pUC18. Each nucleotide was read in both directions on the two strands of a single plasmid. The 5' end of the viral RNA was sequenced by dideoxy-terminated reverse transcription on 5 µg of total RNAs extracted from infected *N. occidentalis* tissue, or 11 µg of purified viral RNA. Primer GB10 (5' CAGGGCTGCACTTGCAACAC 3'), complementary to nt 73-92 of the viral RNA, was labelled with polynucleotide kinase and [32P]dATP and extended by reverse transcription (Meshi et al., 1983). The 5'-terminal nucleotide was determined by terminal deoxynucleotidyl transferase (TdT) treatment according to DeBorde et al. (1986).

Sequence data were assembled and analysed with the DNA Strider program (Marck, 1988). Comparison of the first two ORFs of GVB RNA with PIR (release 47.0) was done with the FASTA (Pearson & Lipman, 1988), BLAST (Altschul et al., 1990) and BLITZ (Smith & Waterman, 1981) programs. Multiple alignments and clustering analysis were performed both with the CLUSTAL V program (Higgins & Sharp, 1988) and with the PILEUP program of the GCG programs package (Devereux et al., 1984). Tentative phylogenetic trees and bootstrap analyses were carried out with the SEQBOOT, PROTDIST, NEIGHBOR and CONSENSE programs of the PHYLIP package (Felsenstein, 1989).

**Expression of the 20 kDa polypeptide as a fusion protein in *E. coli* and antiserum production.** An XbaI–HindIII cDNA fragment of clone 3rp (see Fig. 1), coding for 155 of the 179 amino acids of the putative 20 kDa product of ORF 2, was made blunt-ended, cloned in the Smal site of plasmid pGEX 3X (Smith & Johnson, 1988) and used to express the viral product in *E. coli* strain BL21 (Studier et al., 1990). Partial purification of the fusion protein was obtained by fractionating total bacterial proteins in SDS–PAGE (Laemmli, 1970), staining in ice-cold 0.25 M-potassium acetate (Hager & Burgess, 1980) and excising the overexpressed product as a single band (Niesbach-Klosgen et al., 1990).

Approximately 300 µg of fusion protein was emulsified with an equal volume of incomplete Freund's adjuvant and injected subcutaneously at multiple sites in the back of two rabbits. Three weeks later, a booster injection with 400 µg of fusion protein was given with the same quantity of incomplete Freund's adjuvant. The rabbits were bled a week later. Detection of 20 kDa-related proteins in infected *N. occidentalis* plants was carried out with total (200 mg of plant tissue homogenized in 5 vols of 1 × Laemmli buffer) or fractionated protein extracts (Deom et al., 1990).
Fig. 2. For legend see opposite.
Results

Nucleotide sequence and open reading frames

The cloning strategy for GVB RNA sequencing is shown in Fig. 1. The 5'-terminal sequence of the viral RNA consisted of 7598 nt, excluding the poly(A) tail. This total is in excellent agreement with that (about 7600 nt) estimated by electrophoresis of viral RNA extracts (Boscia et al., 1993). The 5'-terminal nucleotides were sequenced directly on total RNA extracted from infected N. occidentalis, or on purified viral RNA.

Both types of RNA preparations gave two terminal 'strong stop' reverse transcription run-off products (Fig. 3a) suggesting the presence of a cap (Ahlquist & Janda, 1984). TdT treatment (DeBorde et al., 1986) confirmed that the RNA is capped and showed the first nucleotide to be adenine (Fig. 3b).

The overall A + U and G + C content of genomic RNA was 53.5% and 46.5%, respectively, with a slight predominance of adenine (29%). Seven nucleotide variations were found among different clones. Five of these variations, translationally silent, corresponded to transitions A → G at positions 1868, 4001 and 4937; C → U at position 2030; and a transversion A → C at position 701. The other two, consisting of transitions C → U at position 1855 and A → G at position 4129, resulted in the change of two amino acids with the same biochemical properties (Ala → Val and Arg → Lys respectively).
Starting from the 5' end, a non-translated region of 47 nt rich in A + T (68%) preceded the initiation codon (nt 48-50) of ORF 1, which is in a favourable translation context with A in the -4 and -1 positions and C in the +2 position (Lutcke et al., 1987). ORF 1 encoded a polypeptide 1707 amino acids long with a deduced molecular mass of 194 768 Da ('194K'). An opal stop codon (nt 5169-5171) terminated the ORF.

ORF 2, which overlapped ORF 1 by 23 nt (Fig. 1), had an AUG codon at position 5146-5148 and an ochre stop codon at position 5683-5685. This potential ORF coded for a 179 amino acid product with a deduced molecular mass of 20 287 Da ('20K'). The AUG initiation codon of ORF 2, which immediately follows an in-frame amber stop codon, was not in a favourable translation context (Lutcke et al., 1987; Kozak, 1987).

Expression of the 20K protein in E. coli

Following induction with IPTG, bacteria containing the recombinant plasmid pX-20k over-expressed a protein with a molecular mass of about 45 kDa, as could be expected from the fusion of a viral product of about 18 kDa with the carrier glutathione S-transferase (GST; 27 kDa) (not shown). This protein was purified and used to raise two different antisera. Both antisera clearly reacted with the bacterial antigen (not shown) but, despite repeated attempts, failed to reveal the presence of the putative ORF 2 product in total or fractionated protein extracts from GVB-infected plants.

Homologies with other filamentous plant viruses

Comparison of the amino acid sequence of ORF 1 with known viral proteins revealed extensive homologies with
RNA virus replication-related proteins. Starting from the N-terminal region (aa 66–242), the putative methyltransferase domain of the 'Sindbis-like' supergroup of positive-strand RNA viruses was identified with the four domains described by Rozanov et al. (1992). Typical motifs of nucleic acid helicases (Gorbalenya & Koonin, 1989) were located in the central part (aa 920–1140) of this polypeptide, whereas the GDD motif of RNA-dependent RNA polymerase (Koonin, 1991) was in the C-terminal region between aa 1390–1621. Homologies with putative papain-related thiol proteinases of positive-strand RNA viruses (Gorbalenya et al., 1991) were also found in a block of amino acids (751–853) immediately preceding the helicase domain (Rozanov et al., 1995). Search for homologies of the 20K product of ORF 2 with known proteins yielded consistently negative results.

Phylogenetic analysis using amino acid signatures of CPs (Fig. 4a) and polymerase domains (Fig. 4b) of different potex-, carla-, capillo- and trichoviruses revealed differences within the genus Trichovirus. It was confirmed that GVB CP is closely related (estimated bootstrap value 59) to CP of GVA and ACLSV (Minafra et al., 1994), but does not cluster with PVT CP (Fig. 4a). The reverse was true for GVB polymerase which clustered with that of PVT, but was clearly separated (estimated bootstrap value of 63) from that of ACLSV.

**Discussion**

GVB is the second trichovirus species, after ACLSV, for which the RNA has been completely sequenced (German et al., 1990; Sato et al., 1993). Although the size of the genome of both these viruses is virtually the same (7552–7555 nt for ACLSV and 7598 nt for GVB), distinctive differences exist in genome organization and in the size, number and order of genes.

As compared with ACLSV (Fig. 1), GVB has two additional ORFs (ORF 2 and ORF 5) and smaller genes coding for replication-related proteins (1947 kDa compared to 2165 kDa) and movement proteins (365 kDa compared to 505 kDa). Differences exist also in the level of sequence homology between polymerases, which, contrary to capsid proteins (Fig. 4a), do not cluster together in phylogenetic trees (Fig. 4b). In this respect, GVB seems closer to PVT than ACLSV.

The expression product of ORF 2 (20K protein) is most unusual for, apparently, it does not have any significant homology with known protein sequences in current databases. The biological function of this protein is therefore undetermined. The fact that the 20K polypeptide was not identified in GVB-infected plant tissue extracts can be explained if this protein is either expressed transiently, or in quantities below the detection level afforded by the extraction technique used. ORF 2 is not unique to GVB as it also occurs in grapevine virus A (GVA), another tentative trichovirus (unpublished information).

The occurrence in the GVB genome of a 3′-terminal ORF missing in ACLSV and PVT (the other two trichoviruses sequenced so far) suggests that GVB be assigned the status of tentative species in the genus Trichovirus (Martelli et al., 1994). The results of the present study now show that the molecular gap between definitive (ACLSV and PVT) and putative (GVA and GVB) trichovirus species is wider. Therefore, the molecular biological and epidemiological differences among the trichoviruses may warrant the taxonomic revision of the genus.

Research supported by the National Research Council of Italy. Special Project RAISA, Subproject no. 2, Paper no. 2779. Grateful thanks are expressed to Dr S. Namba for the unpublished information on PVT sequence.

**References**


Received 22 March 1996; Accepted 31 May 1996