Complete nucleotide sequence and organization of the RNA genome of groundnut rosette umbravirus

M. E. Taliansky, D. J. Robinson and A. F. Murant
Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

Complementary DNA clones representing the entire genome of groundnut rosette umbravirus (GRV) were obtained and sequenced. GRV RNA comprises 4019 nucleotides and contains four large open reading frames (ORFs). The second ORF from the 5' end includes sequences that encode motifs characteristic of viral RNA-dependent RNA polymerases and is probably expressed by a -1 frameshift mechanism as a fusion protein with the product of the 5'-most ORF. The other two ORFs are almost completely overlapping in different reading frames, and are probably expressed from subgenomic RNA. One of the putative products has significant sequence similarity with viral movement proteins. None of the putative proteins encoded by GRV RNA seems to be a structural protein. In genome organization and in the amino acid sequences of its potential products, the RNA of GRV is similar to that of carrot mottle mimic umbravirus, and to the umbravirus-like RNA-2 of pea enation mosaic virus.

Introduction

Groundnut rosette virus (GRV) is a member of the genus Umbravirus (Murant et al., 1995). This is a group of imperfectly characterized plant viruses, each of which depends on an unrelated helper luteovirus (or luteo-like virus) for transmission by aphids in a circulative, non-propagative manner. Among the other members are bean yellow vein-banding virus, carrot mottle virus (CMoV), lettuce speckles mottle virus and tobacco mottle virus (Murant et al., 1995). Another umbravirus with properties very like those of CMoV, but distinct from it, was recently isolated in Australia. The genome RNA sequence of this virus, provisionally named carrot mottle mimic virus (CMoMV), has been determined (Gibbs, 1995). The umbraviruses, unlike their helper luteoviruses, are mechanically transmissible, and do not rely on their helper viruses for multiplication in plants. The infectivity of umbraviruses is associated with single-stranded RNA (ssRNA), but no conventional virus-like particles have been seen in preparations from plants infected only with an umbravirus (Murant et al., 1969, 1973; Falk et al., 1979), but it has been unclear whether these are virus particles of an unusual kind or cytopathological structures.

GRV occurs in groundnuts in Africa south of the Sahara, where it is part of the complex of agents that cause groundnut rosette disease, the most destructive virus disease of groundnut in Africa (Reddy et al., 1985 a). In nature, GRV is transmitted by Aphis craccivora, but only from plants that also contain the helper virus, groundnut rosette assistor luteovirus (GRAV). Infective ssRNA of GRV has not been obtained free from host plant RNA. However, infected plants yield abundant double-stranded RNA (dsRNA), and electrophoretic analysis revealed three major dsRNA species not present in healthy plants (Reddy et al., 1985 b; Murant et al., 1988). The largest of them, dsRNA-1 (approx. 4-6 kbp), was presumed to be the double-stranded form of the single-stranded genomic RNA of GRV; dsRNA-2 (approx. 1-3 kbp), which has at least some sequences in common with dsRNA-1 (Murant et al., 1988), may represent the dsRNA form of a subgenomic RNA species. DsRNA-3 (approx. 0-9 kbp) has been shown to represent a satellite RNA, which is largely responsible for the symptoms of rosette disease (Murant et al., 1988). Isolates of GRV from which the satellite has been eliminated replicate in infected plants (Murant et al., 1988). However, the satellite plays an essential role, as yet unexplained, in the GRAV-dependent aphid transmission of GRV (Murant, 1990). The complete nucleotide sequences of ten variants of the GRV satellite RNA were presented by Blok et al. (1994), and that of the capsid protein gene of GRAV by Scott et al. (1996). In this paper, we report the complete
nucleotide sequence of the GRV genome, describe its organization and compare it with the genomes of other plant viruses.

Methods

Virus culture. The satellite-free culture MC1 used in this work was derived by Murant & Kumar (1990) from a GRV isolate obtained from a Malawian groundnut plant showing symptoms of chlorotic rosette. MC1 was propagated in *Nicotiana benthamiana* by manual inoculation, under a licence issued by the Scottish Office Agriculture and Fisheries Department.

Preparation of dsRNA and total ssRNA from GRV-infected plants. DsRNA was prepared from 100 g portions of GRV-infected *N. benthamiana* leaf tissue by a slight modification of the method of Murant et al. (1988). The aqueous phase after phenol extraction was adjusted to

<table>
<thead>
<tr>
<th>Table 1. Primers used for cloning and sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer designation</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Sense-strand primers</td>
</tr>
<tr>
<td>S1</td>
</tr>
<tr>
<td>S2</td>
</tr>
<tr>
<td>S3</td>
</tr>
<tr>
<td>S4</td>
</tr>
<tr>
<td>S5</td>
</tr>
<tr>
<td>Complementary-sense primers</td>
</tr>
<tr>
<td>C1</td>
</tr>
<tr>
<td>C2</td>
</tr>
<tr>
<td>C3</td>
</tr>
<tr>
<td>C4</td>
</tr>
<tr>
<td>C5</td>
</tr>
<tr>
<td>C6</td>
</tr>
<tr>
<td>C7</td>
</tr>
</tbody>
</table>

Fig. 1. cDNA cloning and sequencing strategy. The open bar represents the GRV genome and the filled bars represent the cDNA clones obtained (a) by random primed first-strand cDNA synthesis on denatured dsRNA as template, followed by second-strand cDNA synthesis, (b) by first-strand synthesis with sequence-specific primers on total ssRNA as template, followed by PCR amplification, (c) by using d(T)5G as primer for first-strand synthesis on polyadenylated dsRNA as template, followed by PCR amplification. Arrows s1–s5 and c1–c7 represent the primers used for cloning and sequencing, as listed in Table 1.
Fig. 2. For legend see page 2339.
Fig. 2. For legend see opposite.
Fig. 2. Nucleotide sequence of GRV RNA and, below, the amino acid sequences encoded by the four large ORFs. Variant nucleotides found at four positions are indicated above the main sequence, and any resulting altered amino acids are indicated lower-case letters. Asterisks indicate termination codons. A sequence motif characteristic of zinc-dependent carboxypeptidases is in italics, and the eight conserved motifs of RNA-dependent RNA polymerases defined by Koonin & Dolja (1993) are underlined.

18.5% (v/v) ethanol and chromatographed twice on Whatman CF-11 cellulose. Total ssRNA was isolated from GRV-infected N. benthamiana as described by Blok et al. (1994).

### Synthesis and cloning of cDNA

To produce the first series of cDNA clones, dsRNA was denatured with methylmercuric hydroxide and used as template for synthesis of first-strand cDNA (M. Dickenson, personal communication). DsRNA (1 μg) was treated in a total volume of 10 μl with 10 mM-methylmercuric hydroxide in the presence of 0.5 mg/ml random deoxyribonucleotide hexamers (Boehringer) at room temperature for 10 min and then frozen in liquid nitrogen. Upon removal from the liquid nitrogen, 90 μl of pre-warmed reverse transcription reaction mixture was added, to give final concentrations of 50 mM-Tris-HCl, pH 8.3, 75 mM-KCl, 3 mM-MgCl₂, 10 mM-dithiothreitol, 1 mM of each deoxynucleotide triphosphate, 0.3 units/μl human placental ribonuclease inhibitor (Boehringer) and 8 units/μl Moloney murine leukemia virus reverse transcriptase (BRL). The mixture was incubated at 37 °C for 1 h and extracted with phenol. Second-strand cDNA was synthesized by the method of Gubler & Hoffman (1983), rendered blunt-ended with T4 DNA polymerase, ligated into EcoRV-digested pKR (Waye et al., 1985), and transformed into competent E. coli JM101 cells (Stratagene) by standard methods (Sambrook et al., 1989).

### Sequence determination

Cloned cDNAs were sequenced on an ABI model 373A Stretch DNA sequencer (Perkin Elmer) using the PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing kit. Universal M13 (--40) and Reverse M13 primers were used to sequence the terminal parts of the cDNA inserts, and additional primers were
hydroxide, and ssRNA templates were heated at 70 °C for 5 min and completely sequenced on both strands.

■ Primer extension. Terminal sequences were mapped by primer extension, essentially as described by Agranovsky et al. (1994). dsRNA templates were denatured by treatment with 10 mM-methylmercuric hydroxide, and ssRNA templates were heated at 70 °C for 5 min and chilled on ice immediately before use.

■ Sequence analysis. Sequence data were assembled and analysed with the aid of the UWCGG programs (Devereux et al., 1984). The programs FASTA, TFASTA, MPRC and BLASTP were used to search databases for sequence similarity. For sequence comparisons, the GAP and BESTFIT programs were used. CLUSTALV was used to generate multiple alignments of amino acid sequences. RNA secondary structure predictions were made with FOLDRNA and SQUIGGLES.

■ Northern blot hybridization. Total RNA preparations were denatured with formaldehyde and formamide, and electrophoresed in 1-2% agarose gels in TBE buffer. RNA was transferred to Hybond-N nylon membrane (Amersham) by capillary blotting with 20 x SSC and hybridized as described by Sambrook et al. (1989). GRV-specific probes were prepared by incorporation of [3P]dATP using a Random Primers DNA Labelling kit (Life Technologies).

Results

CDNA cloning

Because GRV does not form conventional nucleoprotein particles, the first series of cDNA clones was generated from the virus-specific dsRNA isolated from infected N. benthamiana. Of the clones obtained, five had inserts longer than 400 bp that were shown to be GRV-specific by dot-blot hybridization (Sambrook et al., 1989) to RNA from GRV-infected plants but not to that from healthy plants. Database searches with the nucleotide sequences determined from these inserts revealed similarities with sequences in RNA-2 of pea enation mosaic virus (PEMV). One insert in particular, gr21, showed strong similarity with the 5' end of PEMV RNA-2, including the three 5'-terminal G residues. Thus, gr21 was presumed to represent the 5' terminus of GRV RNA, and the other four inserts were arranged along the GRV genome by analogy with PEMV RNA-2 (Fig. 1a). The location of the 5' end of GRV RNA was confirmed by primer extension from oligonucleotide C5 (Table 1). To cover the gaps between the sequences represented in the first series of clones, a second series of clones was generated by PCR with synthetic primers (Fig. 1, Table 1). Finally, to obtain a clone representing the extreme 3'-terminal region of GRV RNA, dsRNA was polyadenylated and cDNA synthesis was primed with d(T)35G, on the assumption that the 3'-terminal residue of GRV RNA is C as in PEMV RNA-2. The 3'-terminal sequence was confirmed by primer extension of oligonucleotide S5 (Table 1) on denatured dsRNA as template.

Sequence and genome organization

Fig. 2 shows the complete nucleotide sequence of GRV RNA, comprising 4019 nucleotides (nt). The double-stranded form of this RNA would have $M_r$ about $2.6 \times 10^5$, which is comparable to the $3.0 \times 10^6$ estimated by PAGE (Reddy et al., 1985). Fig. 3a depicts the size and arrangement of open reading frames (ORFs); the sequences of the predicted protein products are shown in Fig. 2. At the 5' end, a short non-coding region of 12 nt is followed by an ORF (ORF1) that encodes a potential product of $M_s$ 31 000 (31K). A second ORF overlaps the final ten codons of ORF1 in a different reading frame, and has the capacity to encode a polypeptide of $M_s$ 64 000 (64K). This putative 64K polypeptide is composed of 119 amino acid residues preceding the first methionine codon, followed by another 445 amino acids and terminates at an amber codon (nt 2529). Following the second ORF, a 112 nt non-translated region precedes ORF3, which encodes a putative product with an $M_s$ of 27 000 (27K). Superimposed in a different reading frame on this ORF is another (ORF4), encoding a putative product with an $M_s$ of 28 000 (28K). ORF4 begins 61 nt after ORF3, extends 85 nt beyond its termination codon, and is followed by a long 538 nt 3' non-coding region. A similar genome organization has been demonstrated for PEMV RNA-2 (Demler et al., 1993, 1994) and for CMoMV (Gibbs, 1995) (Fig. 3).

Products of ORFs 1 and 2

Database searches with the amino acid sequence of the putative product of ORF1 (31K) revealed 52% similarity with the ORF1 product of PEMV RNA-2 ($M_s$ 33 000) and 56% similarity with the analogous product of CMoMV RNA ($M_s$ 37 000). None of these three sequences showed significant similarity with any other viral or non-viral amino acid sequences in the database. However, the putative product of GRV ORF1 contains a sequence motif (italics in Fig. 2) that exactly corresponds to one of the two signature patterns of zinc-dependent carboxypeptidases, namely H(S,T,A,G)-X3-(L,I,V,M)-X2-(L,I,V,M,F,Y,W)P(F,Y,W). Such a motif is not present in the ORF1 products of PEMV RNA-2 or CMoMV RNA, and it is unclear whether that in GRV RNA has any functional significance.

The 64K polypeptide encoded by GRV ORF2 was identified as a probable RNA-dependent RNA polymerase. Amino acid sequence comparisons with other plant virus proteins showed that it has high degrees of similarity with virus-specific RNA polymerases encoded by PEMV RNA-2 (72% similarity), dianthoviruses, carmoviruses, necroviruses and subgroup I luteoviruses. These viruses belong to class II of positive-strand RNA viruses according to the classification of Koonin & Dolja (1993), which is based on a tentative phylogeny of their RNA-dependent RNA polymerases. All eight conserved motifs of RNA-dependent RNA polymerases of positive-strand RNA viruses, as defined by Koonin & Dolja (1993), were identified in the GRV ORF2 product (underlined in Fig. 2).

Duplex unwinding by an RNA helicase is thought to be involved in replication of viral RNA genomes over 6 kb in size.
but no virus-coded helicase has been identified in viruses with small genomes, such as dianthoviruses, carmoviruses, tombus-viruses and luteoviruses (Gorbalenya & Koonin, 1989; Koonin & Dolja, 1993). No motif typical of a helicase domain could be identified in the putative products of GRV ORFs 1 or 2 (or of ORFs 3 or 4).

The overlapping arrangement of ORFs 1 and 2, together with the lack of an AUG initiation codon near the 5' end of ORF2, suggest that ORF2 may be expressed by a -1 frameshift from ORF1, as has been proposed for the analogous ORFs of PEMV (Demler et al., 1993) and of CMoMV (Gibbs, 1995). This idea is supported by the presence, immediately upstream of the UAG stop codon of GRV ORF1, of the ‘shifty’ heptanucleotide AAAUUUU. This motif is similar to the consensus frameshift signal of a number of plant and animal viruses (Miller et al., 1995). Such a frameshift would yield a fusion protein with an Mr of 94000 from ORFs 1 and 2.

Products of ORFs 3 and 4

Amino acid sequence comparisons showed that the putative product of GRV ORF3 (27K) has 50% and 42% similarity, respectively, with the corresponding proteins encoded by PEMV RNA-2 and CMoMV RNA. For the ORF4 product (28K), the equivalent values were 74% and 56%. Database searches with the sequence of the GRV ORF4 protein also revealed long stretches of similarity with the ‘P3a’ proteins of cucumoviruses (Fig. 4), which have been shown to be involved in cell-to-cell movement (Kaplan et al., 1995), as well as with the putative movement proteins of several other plant viruses. A motif typical of plant virus movement proteins of the ‘30K superfamily’ has previously been identified in the ORF4 protein of PEMV RNA-2 (Mushegian & Koonin, 1993). Thus, the 28K GRV ORF4 protein is probably involved in movement of the virus from cell to cell.

Database searches with the sequence of the 27K GRV ORF3 protein revealed no significant similarity with any other viral or non-viral proteins, except the corresponding proteins encoded by PEMV RNA-2 and CMoMV RNA. The function of these proteins is therefore uncertain. Neither GRV nor CMoMV form conventional nucleoprotein particles, and PEMV RNA-2 is encapsidated by an RNA-1-encoded protein (Demler et al., 1993). Thus, the ORF3 proteins are probably not particle proteins. There are examples in which more than one virus-coded protein is involved in movement of the virus within infected plants (e.g. Petty et al., 1990; Traynor et al., 1991; Nelson et al., 1993; Hilf & Dawson, 1993; Taliansky & Garcia-Arenal, 1995) and the ORF3 proteins may play a role, for example, in long distance virus movement.

The location of the overlapping ORFs 3 and 4 towards the 3' end of GRV RNA suggests they may be expressed from subgenomic mRNA, which might correspond to the 13 kbp dsRNA previously identified in GRV-infected plants (Murant et al., 1988). To confirm the presence of such subgenomic RNA in infected plants, total RNA extracts were analysed by Northern blot hybridization. Probes derived from clones gr21 and gr11, which are near the 5' end of GRV RNA, revealed...
predicted for subgenomic RNA (Fig. 5). It is not clear whether both species function as subgenomic mRNAs or whether one is a degradation product.

**Non-coding regions**

The ends of the GRV genomic RNA have partial identity with those of the GRV satellite RNA: 12 of the first 16 nt of the genomic RNA at the 5' terminus and 7 of the final 9 at the 3' terminus are identical with those in the satellite RNA.

**Genomic RNA**

5' ggGGGUUUCAAC ATGG . . . . AUGaCaCCC 3'

Satellite RNA

5' GGGUUUCAAC ATGG . . . . AUGGcGCCC 3'

In the genomic and satellite RNAs, the three 5'-most residues in both plus and minus strands are Gs, and in this respect they resemble many viral RNAs, including PEMV RNA-2 and carnation mottle carmovirus RNA.

The internal non-coding region in GRV RNA, between ORFs 2 and 3, might be expected to contain a promoter sequence for synthesis of the putative subgenomic mRNA for the ORF 3 and 4 proteins. However, no similarity was found with comparable regions in the genomes of dianthoviruses, only one RNA species, corresponding in size to the genomic RNA (Fig. 5). However, probes derived from clone grpmp1, which contains ORFs 3 and 4, or from the 3'-terminal clone gr51 revealed two additional RNA species, both about the size predicted for subgenomic RNA (Fig. 5). It is not clear whether both species function as subgenomic mRNAs or whether one is a degradation product.

**Non-coding regions**

The ends of the GRV genomic RNA have partial identity with those of the GRV satellite RNA: 12 of the first 16 nt of the genomic RNA at the 5' terminus and 7 of the final 9 at the 3' terminus are identical with those in the satellite RNA.

**Genomic RNA**

5' ggGGGUUUCAAC ATGG . . . . AUGaCaCCC 3'

Satellite RNA

5' GGGUUUCAAC ATGG . . . . AUGGcGCCC 3'

In the genomic and satellite RNAs, the three 5'-most residues in both plus and minus strands are Gs, and in this respect they resemble many viral RNAs, including PEMV RNA-2 and carnation mottle carmovirus RNA.

The internal non-coding region in GRV RNA, between ORFs 2 and 3, might be expected to contain a promoter sequence for synthesis of the putative subgenomic mRNA for the ORF 3 and 4 proteins. However, no similarity was found with comparable regions in the genomes of dianthoviruses,
carmoviruses, tombusviruses, necroviruses or luteoviruses, although there was some similarity with the intergenic regions of PEMV RNA-1 and RNA-2. In particular, two blocks of sequence, which were previously shown to be duplicated (in the arrangement ABAB) in the intergenic region of PEMV RNA-1 and RNA-2. In particular, two blocks of RNA-2 and to occur once (arranged as BA) in the intergenic region of the luteovirus-like PEMV RNA-1 (Demler et al., 1993), also occur unduplicated (arranged as BA) in the intergenic region of GRV (Fig. 6). These sequences may play some role as elements of a subgenomic RNA promoter.

Discussion

The close relationship between GRV RNA and PEMV RNA-2 is evident from the strong amino acid sequence similarities of all their potential translation products, as well as some similarities in the nucleotide sequences of their non-translated regions. A third RNA closely related to these two is the genomic RNA of CMoMV. All three have a similar genome organization, which is different from those of all other plant viruses that have been examined. Particularly notable is their lack of a capsid protein gene. For the umbravirus GRV and the presumed umbravirus CMoMV, this accords with their failure to produce conventional nucleoprotein particles, and is probably a defining feature of the genus Umbravirus. One of the most important functions of virus particle proteins is to protect the viral RNA. Virus variants that are defective in particle protein production, such as NM-type isolates of tobacco rattle virus, are often unstable and accumulate in plants only to low levels (Harrison & Robinson, 1981). In contrast, umbraviruses multiply efficiently in plants and are moderately stable in extracts. For example, extracts of GRV-infected N. clevelandii leaves were infective when diluted up to 1000-fold and remained infective for a day at room temperature and for 15 days at 4 °C (Reddy et al., 1985b). These properties suggest that umbravirus RNA may be associated in infected plants with a structure that affords it some measure of protection, perhaps an 50 to 60 nm membranous particles described by Murant et al. (1969) and Falk et al. (1979).

Beet western yellows virus ST9-associated RNA (Chin et al., 1993) shares some biological properties with umbraviruses in that it is capable of independent replication in infected plants, but is dependent on its helper luteovirus for aphid transmission (Passmore et al., 1993). However, it does not have what now appears to be the typical umbravirus genome organization, nor did we identify obvious sequence similarities between it and GRV RNA, apart from conserved RNA polymerase motifs.

Although PEMV RNA-2 is regarded as one component of a bipartite genome, the other component of which includes a capsid protein gene, it clearly has strong affinities with umbraviruses. Indeed, PEMV RNA-2 is capable of infecting plants systemically in the absence of RNA-1 (Demler et al., 1994, 1996), and such infections resemble those with umbraviruses. Moreover, PEMV RNA-1 resembles the genomes of luteoviruses (Demler & de Zoeten, 1991) and the parallel with the luteoviruses that mediate the aphid transmission of umbraviruses is obvious. Thus, as suggested by Demler et al. (1993), it seems likely that PEMV has evolved from a luteovirus–umbravirus complex, similar to that involving GRV and GRAV, with increased mutual dependence of the two parts. PEMV RNA-2 depends on RNA-1 for encapsidation, whereas PEMV RNA-1 depends on RNA-2 for cell-to-cell movement.

Amino acid sequence comparisons showed that the putative RNA-dependent RNA polymerase of GRV, the ORF2 product, belongs to the so-called supergroup 2 of RNA polymerases (Koonin & Dolja, 1993), which also includes those of carmoviruses, tombusviruses, subgroup I luteoviruses, dianthoviruses and necroviruses. Thus, based primarily on similarities in RNA-dependent RNA polymerases, which are the only universally conserved proteins of positive-strand RNA viruses, the genus Umbravirus might be placed in the family Tombusviridae (Mayo & Martelli, 1993; Russo et al., 1994), or at least considered as a genus related to the Tombusviridae.

Note added in proof. The genome RNA sequence of the virus referred to here as carrot mottle mimic virus is reported in Gibbs, M. J., Cooper, J. I. & Waterhouse, P. M. (1996). The genome organization and affinities of an Australian isolate of carrot mottle umbravirus. Virology (in press).

We thank Ken Taylor for assistance with sequencing. SCRI is grant-aided by the Scottish Office Agriculture, Environment and Fisheries Department. Work on this project was supported by the Overseas Development Administration through the Centre for Arid Zone Studies, Bangor.

References


Received 4 March 1996; Accepted 26 April 1996