Complete nucleotide sequence of peanut clump virus RNA 1 and relationships with other fungus-transmitted rod-shaped viruses

E. Herzog, H. Guilley, S. K. Manohar, M. Dollet, K. Richards, C. Fritsch and G. Jonard

1 Institut de Biologie Moléculaire des Plantes du CNRS et de l’Université Louis Pasteur, 12 rue du Général Zimmer, Strasbourg 67084 Cedex and 2 IRHO, Laboratoire de Phytovirologie des Régions Chaudes, CIRAD-ORSTOM, BP 5035, 34032 Montpellier Cedex 01, France

The complete nucleotide sequence of RNA 1 of the tentative furovirus peanut clump virus (PCV) has been determined by characterization of cloned cDNA and by direct RNA sequencing. The sequence is 5897 nucleotides in length and contains three long open reading frames (ORFs). The 5’-terminal proximal ORF has the potential to encode a polypeptide of $M_r$ 130942 (P131) containing methyltransferase and RNA helicase homologous domains and displaying homology with large nonstructural proteins of alpha-like viruses, which are known or thought to be involved in virus replication. The P131 ORF is followed in-frame by a second ORF which is probably expressed by partial readthrough of the UGA termination codon of the P131 ORF to produce a polypeptide of $M_r$ 191044 (P191). The readthrough region of P191 contains the characteristic ‘core’ RNA polymerase motif, indicating that the PCV replicase proteins are expressed as a pair of overlapping proteins as in the tobamoviruses, tobraviruses and the furovirus soil-borne wheat mosaic virus (SBWMV). Sequence comparisons indicate that P131 and P191 are most closely related to the replicase proteins of SBWMV and the hordeivirus barley stripe mosaic virus (BSMV) but are only distantly related to the replicase of the furovirus beet necrotic yellow vein virus (BNYVV). The 3’-terminal proximal ORF can encode a putative polypeptide of $M_r$ 14556 (P15) which displays homology to small cysteine-rich proteins of hordeiviruses and SBWMV. We have corrected four errors in the sequence of PCV RNA 2 published previously by Manohar et al. (Virology 195, 33-41, 1993). One of these changes causes two small ORFs near the 3′ terminus of RNA 2 to be fused together to create an ORF for a putative polypeptide of $M_r$ 16833 (P17) which displays extensive homology with the third protein of the triple gene block of BSMV RNA β.

Introduction

Peanut clump virus (PCV) is a plus-sense RNA virus which has been classified as a tentative member of the furovirus family, a group of viruses with rigid rod-shaped virions, multipartite genomes and soil-borne fungi as vectors (Brunt, 1991). Other accepted or possible members of the furovirus family include the type member, soil-borne wheat mosaic virus (SBWMV), beet necrotic yellow vein virus (BNYVV), Hypochoeris mosaic virus, oat golden stripe virus, Nicotiana velutina mosaic virus, potato mop-top virus, rice stripe necrosis virus and sorghum chlorotic spot virus (Brunt, 1991). PCV has two genome components (Thouvenel et al., 1976; Manohar et al., 1993) and is believed to be transmitted by the plasmodiophoromycete fungus Polymyxa graminis (Thouvenel & Fauquet, 1981). There is a second probable furovirus of groundnut, Indian peanut clump virus (IPCV), which has rod-shaped virions of similar size to those of PCV and an apparently bimodal genome (Reddy et al., 1983). Although PCV and IPCV are not serologically related, recent sequence analysis (Wesley et al., 1994) has revealed that the coat protein of IPCV is 61% identical to that of PCV.

In a previous publication, we presented the complete sequence of PCV RNA 2 and showed that it contained the viral coat protein cistron nearest the 5′ terminus, followed by an open reading frame (ORF) for a 39-K protein and a triple gene block (Manohar et al., 1993). At the same time, the sequence of the 3′ terminal 316 residues of PCV RNA 1 was determined and shown to be 95% identical to the 3′-noncoding region of RNA 2 (Manohar et al., 1993). In this paper, we describe the complete sequence of PCV RNA 1 (5897 residues) and present a revised sequence for RNA 2, in which several errors in the published sequence (Manohar et al., 1993)
Methods

Virus and viral RNA. Purification of PCV isolate PCV2 from systemically infected Nicotiana benthamiana and extraction of viral RNA were as described by Manohar et al. (1993).

Synthesis and cloning of double-stranded cDNA. Unless otherwise noted, manipulation of recombinant DNA and related procedures followed conventional techniques (Sambrook et al., 1989). Total PCV RNA was 3'-polyadenylated (Manohar et al., 1993) and cDNA synthesis was primed using PstI-linearized oligo(dT)-tailed pUC9 (Pharmacia) (Heidecker & Messing, 1983). Recombinant plasmids containing double-stranded (ds) cDNA were prepared as described (Heidecker & Messing, 1983) with minor modifications (Bouzoubaa et al., 1985). Other recombinant cDNA plasmids were produced by using synthetic oligomers complementary to known portions of the RNA 1 sequence in order to prime first-strand synthesis. Second-strand synthesis was by the method of Gubler & Hoffman (1983) with minor modifications (Bouzoubaa et al., 1987). Alternatively, the cDNA was 3'-polyadenylated with terminal transferase and second-strand synthesis was primed with the synthetic deoxyoligonucleotide 5' GCCAAGCTT 18.

DNA sequencing. DNA sequence was determined by the dideoxynucleotide triphosphate chain termination method (Sanger et al., 1977) adapted to use with dsDNA templates (Zhang et al., 1988). The terminal sequences of the inserts were obtained using the direct and reverse universal primers. Long cDNA inserts of interest were transferred to pBluescript II pKS(-) [pBKS(-)] and sets of clones containing nested deletions extending in from each end of the insert were generated by mild treatment of linearized plasmid DNA with exonuclease III (Henikoff, 1987), using the Pharmacia nested-deletion kit and following the supplier’s instructions. Sequences near the extremities of the 5'- and 3'-terminally truncated inserts of members of the nested deletion sets were obtained using the universal primers. In certain cases, sequence was also obtained using synthetic oligomers designed to hybridize to appropriate internal sequences on the viral RNA. The sequence of the first 16 residues of RNA 1 was determined by the dideoxyribonucleotide triphosphate chain termination method using reverse transcriptase to prime cDNA synthesis from oligo 1 (complementary to nucleotides 95 to 115) on viral RNA template (Fichot & Girard, 1990; Deborde et al., 1986). Sequence data were tabulated and analysed using UWGCG programs (Devereux et al., 1984) on a VAX microcomputer.

Results and Discussion

Sequence analysis of PCV RNA 1

In the course of analysis of the sequence of PCV RNA 2, a recombinant cDNA clone (pPC1-47) containing the 3'-terminal 704 residues of PCV RNA 1 was isolated and the sequence of the 3'-terminal 316 residues (highly homologous with the corresponding region of RNA 2) was reported (Manohar et al., 1993). Additional RNA 1-specific cDNA clones were obtained by several different methods. Total PCV RNA was 3'-polyadenylated (Manohar et al., 1993) and cDNA synthesis was primed using PstI-linearized oligo(dT)-tailed pUC9. Recombinant plasmids containing ds cDNA were prepared (Heidecker & Messing, 1983) and those plasmids found by restriction enzyme analysis to contain long cDNA inserts were nick-translation and used as probes in Northern hybridization experiments with viral RNA to identify those harboring cDNA inserts derived from RNA 1. The cDNA insert of one such clone was excised with BamHI and HindIII (both polylinker sites) and inserted between the BamHI and HindIII polylinker sites of pBKS(-) (Stratagene) to produce pPCI-18 (Fig. 1).

Recombinant cDNA plasmids pPC1-12 and pPC1-15 were obtained by priming first-strand cDNA synthesis with a 23-mer (oligo 3; Fig. 1) complementary to nucleotides 4412 to 4434 of RNA 1. The cDNA was then 3'-polyadenylated with terminal transferase and second-strand synthesis was primed with the synthetic deoxyoligonucleotide 5' GCCAAGCTT 18. Protruding extremities were filled-in by treatment with the Klenow fragment of Escherichia coli DNA polymerase and the reaction product was ligated into SmaI-linearized pBKS(-). Another clone (pPC1-16.1) was produced by priming first-strand cDNA synthesis with a 20-mer (oligo 2, Fig. 1) complementary to residues 1814 to 1833 of RNA 1, treatment of the RNA-cDNA hybrids with RNAase H, followed by second-strand DNA synthesis (Gubler & Hoffman, 1983; Bouzoubaa et al., 1987). After filling in the protruding extremities with Klenow fragment, the product was cloned into SmaI-linearized pBKS(-).

![Fig. 1. Genetic map of PCV RNA 1 and map of cDNA inserts characterized in determining the sequence. ORFs are represented by numbered hollow rectangles. The origin of the putative gene products P131, P191 and P15 are indicated in bold lettering with P191 postulated to arise from translation suppression of the P131 termination codon (see text). The cloned cDNA inserts characterized in the course of sequence determination are represented by thick lines below the genetic map. Thin lines correspond to 3'-poly(A) extensions added for cloning purposes. Synthetic oligodeoxyribonucleotides used in cloning and sequence determination are represented by thin lines below the genetic map.](image-url)
The sequences near the extremities of the cDNA inserts of clones pPCI-12, -15, -16.1, -18 and -47 were determined by the dideoxynucleotide chain termination method using the universal direct and reverse primers. For those clones containing long cDNA inserts, different amounts of insert sequence proximal to the 5' or 3' terminus were eliminated by unidirectional exonuclease III digestion (Henikoff, 1987) of plasmid DNA which had been linearized by digestion at appropriate polylinker restriction sites. The partially deleted DNA was then circularized by ligation to generate a set of clones containing a graded series of 5'- or 3'-terminally nested insert deletions. This procedure permitted sequence analysis of regions of the cDNA inserts initially too distant from the insert extremities to be sequenced with the universal primers (Henikoff, 1987). Sequence data were obtained for both strands of the inserts. Confirmation of the sequence at the junctions between the initiation and termination codons revealed the presence of independent cDNA clones. The sequence of residues 1 to 13459 (P131) is 130942. The second ORF is immediately adjacent to ORF 1 and in the same reading frame. ORF 2 ends at UAA (5148 to 5150) and has a polypeptide of Mr 60120. As will be shown below, the suppressible UGAs in TRV RNA 1 and pea early browning tobavirus (UUAUGACGGUUUCGG and AAAUGACGGUGGUCG, respectively; Hamilton et al., 1987; MacFarlane et al., 1989) also display considerable homology with the PCV-SBWMV context. The

Genetic organization of PCV RNA 1

Fig. 2 shows the complete nucleotide sequence (5897 nt) of RNA 1. Analysis of the distribution of translation initiation and termination codons revealed the presence of three long ORFs in the plus strand (Fig. 1). No ORF of more than 300 nucleotides was present on the minus strand. The first plus-strand ORF begins with AUG (132 to 134) and terminates with UGA (3567 to 3569). The calculated Mr of the corresponding putative protein (P131) is 130942. The second ORF is immediately adjacent to ORF 1 and in the same reading frame. ORF 2 ends at UAA (5148 to 5150) and has a polypeptide coding capacity of Mr 60120. As will be shown below, the sequence context of UGA (3567 to 3569) strongly suggests that it can undergo suppression during translation. The calculated Mr of the resulting ORF 1–ORF 2 translational fusion protein (P191) is 191044. Products of a size corresponding approximately to P131 and P191 have been detected following in vitro translation of viral RNA of PCV (E. Herzog and C. Fritsch, unpublished observations) and IPCV (Mayo & Reddy, 1985) in rabbit reticulocyte lysate.

The 3 '-proximal ORF 3 of PCV RNA 1 is separated from ORF 2 by an intergenic region of 75 residues. ORF 3 starts with AUG (5225 to 5227) and ends with UAA (5597 to 5599) and, if expressed, should encode a polypeptide of Mr 14556 (P15).

The 3'-noncoding region of RNA 1 is 298 residues in length. As noted previously (Manohar et al., 1993), the 3'-terminal 276 residues display greater than 95% identity with the corresponding region of RNA 2 with the last 96 residues identical. The 5'-noncoding region of RNA 1 is 131 residues in length and the first six residues at the 5' terminus (5' GUAUUC) are identical in PCV RNA 1 and 2.

P131 (and the common portion of the overlapping P191 species) contain an array of domains displaying sequence homology to domains of the putative RNA-dependent RNA polymerases of a host of plus-strand RNA viruses (see Koonin & Dolja, 1993, for a review). In particular, the N-terminal portion of P131 contains the consensus sequence for a putative methyltransferase of the type characteristic of the tobamo-like viruses while the C-terminal portion of the P131 contains the conserved purine nucleotide triphosphate-binding motif and associated consensus sequences which are present in many plus-strand RNA viruses and are thought to be a marker for RNA helicase activity (Koonin & Dolja, 1993). The portion of P191 following the putatively suppressible UGA codon (i.e. ORF 2) contains the well known ‘GDD’ sequence and associated motifs characteristic of the RNA-dependent RNA polymerases of all plus-strand RNA viruses (Koonin & Dolja, 1993). The relative positions of the aforesaid domains on P131 and P191 are identical to their arrangement in the corresponding replicases (also expressed by a readthrough mechanism) of alpha-like viruses such as tobacco mosaic virus (TMV; Goelet et al., 1982), tobacco rattle virus (TRV; Hamilton et al., 1987) and SBWMV (Shirako & Wilson, 1993). It is noteworthy that the nucleotide sequence flanking the suppressible UGA codon of PCV RNA 1 is identical to that flanking the analogous suppressible UGA in SBWMV RNA 1 with the longest stretch of conserved sequence on the 3'-terminal side of the termination codon (AAAAUGACGGUUGGUGUC; the termination codon is underlined). The sequences flanking the suppressible UGAs in TRV RNA 1 and pea early browning tobavirus (UUAAUGACGGUUUCGG and AAAUGACGGUGGUCG, respectively; Hamilton et al., 1987; MacFarlane et al., 1989) also display considerable homology with the PCV-SBWMV context. The
Fig. 2. For legend see facing page.
context of the suppressible UAG in TMV, on the other hand, is quite different (CAAUAGCAUUA). Possibly, the conserved sequence flanking the suppressible UGA codons plays a role in translation readthrough, downstream of the suppressible codon in TMV (Skuzeski et al., 1989), although the nature of the key conserved residues intervening in the process is evidently different from those involved in TMV. The close similarity between the above two codons is evident, as has been shown to be the case for the two codons of TMV (Skuzeski et al., 1991), although the nature of the key conserved residues intervening in the process is evidently different from those involved in TMV.

The putative P15 species encoded by the 3'-proximal ORF has a rather high content of charged residues, particularly in the C-terminal portion, but does not have a particularly high content of cysteine residues, as do small proteins encoded by similarly placed ORFs in the genomes of a number of other rod-shaped plant viruses (Rupasov et al., 1989). However, detailed sequence alignments reveal regions of homology of the PCV RNA 1 P15 with certain of these small proteins (see below).

**Sequence homologies**

The sequence of PCV P131 was used as query to search the sequence database for similar proteins. The greatest homologies (Table 1) were with the Mr ~ 150000 nonstructural protein of barley stripe mosaic virus (BSMV) RNA 1 (Rupasov et al., 1989). However, detailed sequence alignments reveal regions of homology of the PCV RNA 1 P15 with certain of these small proteins (see below).
Table 1. Amino acid sequence comparisons of coding regions of PCV RNA 1 and corresponding coding regions of several other rod-shaped RNA viruses

<table>
<thead>
<tr>
<th>Percentage of identical amino acids*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SBWMV</strong></td>
</tr>
<tr>
<td>38</td>
</tr>
</tbody>
</table>

* Values in the table are the percentage of identical amino acids between the PCV sequence and the indicated sequences after alignment with the program Gap (Devereux et al., 1984).

† The sequence of PCV P131 was compared to the sequences of the SBWMV RNA 1 M, 150000 protein (Shirako & Wilson, 1993), the BSMV RNA α M, 130000 protein (Gustafson et al., 1989), the TRV RNA 1 M, 134000 protein (Hamilton et al., 1987), the TMV M, 126000 protein (Goelet et al., 1982) and amino acids 1 to 1507 of the BNYVV RNA 1 237000 M, protein (Bouzoubaa et al., 1987).

‡ The PCV RNA 1 readthrough domain (ORF 2) amino acid sequence was compared to the readthrough domain sequences of SBWMV RNA 1 (M, 59000), TRV RNA 1 (M, 60000) and TMV RNA 1 (M, 57000), the BSMV RNA γ M, 87000 protein (Gustafson et al., 1989) and amino acids 1508 to 2109 of the BNYVV RNA 1 M, 237000 protein.

§ The sequence of PCV RNA 1 P15 was compared to the SBWMV RNA 2 M, 19000 protein, the BSMV RNA γ M, 17000 protein, the TRV RNA 1 M, 16000 protein and the BNYVV RNA 2 M, 14000 protein (Bouzoubaa et al., 1986).


BSMV MNAYTSOCVCC ......... GKK .......... RKRRHRYSETNKLK ......... ELVYYYKLF QPQCAINGVSGCPC ......... CS1 ......... AEHACQDPLTV ...RFOQQRH 85

PSLV MSTDSICCGPR ...... CACVDGPK .............. SI KCVSK ......... YRISE/YKTI/ ......... L. DVKCPJ ......... PADC ......... AAF ......... VLFHPKLTM ......... DGYCGEKH 75

SBMV MSTVQGIVSCG ...... RYEDGK .......... YARRIDRIAENL ......... TFS GT ......... ATVVRK .......... LEK .......... SEA/ ......... NAIRKSVASS ...... EDSV ......... GCDDSSSVSKL 154


BSMV MNAYTSOCVCC ......... GKK .......... RKRRHRYSETNKLK ......... ELVYYYKLF QPQCAINGVSGCPC ......... CS1 ......... AEHACQDPLTV ...RFOQQRH 85

PSLV MSTDSICCGPR ...... CACVDGPK .............. SI KCVSK ......... YRISE/YKTI/ ......... L. DVKCPJ ......... PADC ......... AAF ......... VLFHPKLTM ......... DGYCGEKH 75

Fig. 3. Sequence homology among the 17000 to 19000 M, cysteine-rich proteins of BSMV, poa semilatent virus (PSLV; Agranovsky et al., 1992) and SBWMV and the PCV RNA M, 150000 protein. Cysteine residues conserved in the BSMV M, 17000 protein, the PCV RNA M, 19000 protein and the BNYVV RNA 2 M, 19000 protein are indicated by asterisks above the sequence. Residues conserved in all four sequences are shown in the consensus line (CON). The sequence alignment was generated with the program Pileup (Devereux et al., 1984).
homology between PCV RNAs 1 and 2 and the 3'-terminal regions of SBWMV RNAs 1 and 2. In particular, the 3'-proximal 27 residues of both the SBWMV and PCV RNAs can be folded into an identical pseudoknot secondary structure (Fig. 4). The 3'-terminal 80 residues of SBWMV RNA display sequence and structural homology with the 3'-terminal valyl-tRNA-like structures of sunnhemp mosaic tobamovirus RNA and several tymoviruses (Shirako & Wilson, 1993). However, the corresponding upstream sequences of PCV RNA 1 and 2 cannot be folded into a structure analogous to the tRNA-like structures mentioned above.

Revised sequence of PCV RNA 2

In the course of further work on PCV RNA 2, several errors were discovered in the published sequence (Manohar et al., 1993). The modifications are as follows: CT in place of TC at positions 3328 to 3329 (numbering refers to the original published sequence); an extra G between nucleotides 3702 to 3703; the G at position 3270 eliminated and a T added between nucleotides 3996 and 3997. The above changes in the PCV 2 sequence, which resulted from transcription errors, have been incorporated into the revised PCV 2 database entry (L07269).

The extra T between nucleotides 3996 and 3997 modifies the genetic map of RNA 2. In the original sequence (Manohar et al., 1993) the third ORF of the triple gene block encoded a hypothetical polypeptide of \( M_r \sim 9000 \) (P9) which was followed by an ORF for a hypothetical \( M_r \sim 7000 \) species (P7) (see Fig. 4 in Manohar et al., 1993). In the revised sequence these two small ORFs are fused together and can encode a hypothetical polypeptide of \( M_r \) 16833 (P17) which would be the third protein in the PCV RNA 2 triple gene block (Fig. 5). P17 displays 35% sequence identity with the third triple gene block protein encoded by BSMV RNA \( \beta \) (data not shown).

Relationships with other furoviruses

Sequence analysis has made it clear that genes and gene cassettes of RNA viruses have been subject to considerable reassortment in the course of evolution (e.g. Morozov et al., 1989; Koonin & Dolja, 1993). Thus, it is more appropriate to consider separately the evolutionary relationships of functionally distinct portions of the genome rather than the genome as a whole. Of the two other furoviruses for which sequence data is available, the putative replicase proteins P131 and P191 of PCV are

---

**Fig. 4.** Sequence homology and conserved secondary structure at the 3' termini of PCV RNA 1 and SBWMV RNA 1. Regions of the sequence which can base pair to generate a pseudoknot are indicated by thick and thin lines. Similar structures can be drawn for PCV RNA 2 and SBWMV RNA 2.

**Fig. 5.** Genetic organization of characterized members of the furovirus family and BSMV. BNYVV RNAs 3 and 4, which are not required for leaf infection (Koenig et al., 1986), are not shown. Regions of the putative replicase proteins thought to be functionally homologous are indicated by light hatching (containing the putative methyltransferase and helicase domains) and heavy hatching (containing the core polymerase domain; the position of the conserved ‘GDD’ motif is indicated by the circle). The small homologous ‘cysteine-rich’ proteins are indicated by stippling. Termination codons which can undergo translational suppression are identified by asterisks. TGB, triple gene block; CP, coat protein.
much more closely related to the corresponding species of SBWMV than to the \( M_r \) 237000 protein of BNYVV. Both the SBWMV and PCV replicase proteins are in turn very closely related to the \( M_r \) 130000 and \( M_r \) 87000 proteins of BSMV (see Fig. 5 for virus genetic maps). Thus the PCV replicase genes, like those of SBWMV and BSMV, derive from one lineage, associated with the tobamoviruses and related viruses, whereas the BNYVV replicase is derived from a lineage associated with Rubella virus (Koonin & Dolja, 1993). However, with respect to proteins known or thought to be involved in cell-to-cell movement, the situation is quite different. The \( M_r \) ~ 37000 putative movement protein encoded by SBWMV RNA 1 is of the dianthovirus type (Shirako & Wilson, 1993) whereas the known or supposed movement proteins of PCV, BNYVV and BSMV are encoded by triple gene blocks (Pettet et al., 1990; Gilmer et al., 1992). Thus, with regard to both the sequence homology displayed by their replicases and the nature of their cell-to-cell movement proteins, SBWMV (the type member of the furovirus group) and BNYVV are distant. With respect to the nature of their movement proteins, PCV clusters with BNYVV and BSMV (with particularly close homology with the latter), whereas SBWMV is distinct from all three of the above. To further complicate the situation, all four viruses (PCV, SBWMV, BNYVV and BSMV) carry genes for an \( M_r \) 14000 to 19000 protein with a number of conserved cysteine residues near the 3' extremity of one of the genomic RNAs (see Fig. 5). There is some sequence homology among these species for PCV, BSMV and SBWMV, but less with the \( M_r \) 14000 species of BNYVV. It should also be noted that there is considerable homology between the coat proteins of PCV and BSMV (Wesley et al., 1994).

Finally, the three furoviruses possess one feature in common which is not shared with BSMV: the presence on RNA 2 of an additional ORF immediately following the 5'-proximal coat protein cistron. This ORF is expressed from full-length RNA 2 by translational readthrough of the coat protein termination codon to produce the \( M_r \) ~ 75000 and \( M_r \) ~ 84000 readthrough proteins of BNYVV and SBWMV, respectively (Ziegler et al., 1985; Hsu & Brakke, 1985), or by a leaky scanning mechanism in the case of P39 of PCV (E. Herzog and C. Fritsch, unpublished). There is no sequence homology between these species but the readthrough domain of BNYVV RNA 2 has been shown to be involved in vector transmission (Tamada & Kusume, 1991). Transmission of SBWMV and PCV without recourse to the fungal vector is often accompanied by deletions in the corresponding ORFs of SBWMV (Shirako & Brakke, 1984; Hsu & Brakke, 1985; Shirako & Ehara, 1986) and PCV (Manohar et al., 1993), providing circumstantial evidence for a similar role in these two viruses. Thus all three species may represent independent adaptations to permit otherwise rather dissimilar viruses to be fungus transmissible.

The authors thank D. L. Beck and R. L. S. Forster for pointing out a possible error in the PCV RNA 2 sequence.

References


(Received 6 May 1994; Accepted 5 July 1994)