

1095

JA 1515

**The coat protein of Indian peanut clump virus:
relationships with other furoviruses and with
barley stripe mosaic virus**

S. V. Wesley^{1,*}, M. A. Mayo¹, C. A. Jolly¹, R. A. Naidu², D. V. R. Reddy²,
M. K. Jana³, and V. K. Parnaik⁴

¹ Scottish Crop Research Institute, Invergowrie, Dundee, U.K.

² ICRISAT, Patancheru, Andhra Pradesh.

³ Indian Institute of Technology, Kharagpur, West Bengal, and

⁴ Center for Cellular and Molecular Biology, Hyderabad, Andhra Pradesh, India

Accepted September 28, 1993

Summary. The 5'-most open reading frame of the c.4kb RNA-2 of Indian peanut clump furovirus (IPCVC) encodes a protein of 208 amino acids. This protein is thought to be the coat protein of IPCVC because its amino acid composition and M_r closely resemble those reported for IPCVC coat protein and because its amino acid sequence is 61% identical to that of the coat protein of peanut clump virus (PCV) from West Africa. The extent of the sequence identity between IPCVC and PCV coat proteins confirms previous conclusions that the viruses are distinct rather than strains of one virus. The sequences of the coat proteins of IPCVC and PCV were between 18% and 26% identical to those of other furoviruses and those of unrelated tobamoviruses and tobamoviruses. In contrast, the coat protein sequences were 37% (IPCVC) and 36% (PCV) identical to that of the coat protein of barley stripe mosaic hordeivirus (BSMV). This similarity between the coat proteins of viruses from different groups (= genera) is unusual but is consistent with previous reports of sequence relatedness in various genes between certain furoviruses and BSMV.

Introduction

Peanut clump is a soil-borne disease of groundnut crops [26]. In West Africa, the disease is caused by infection with peanut clump virus (PCV) which is transmitted by the plasmodiophoromycete fungus *Polymyxa graminis* [27]. PCV particles are rod-shaped and of two lengths and PCV has been classified with other fungus-transmitted bi-partite genome viruses with rod-shaped

* Present address: ICRISAT, Patancheru, Andhra Pradesh, India.

particles in the furovirus group [3]. In India, a similar but serologically distinct virus, named Indian peanut clump virus (IPCV), induced clump disease [20]. No serological relation between PCV and IPCV could be detected by immunosorbent electron microscopy [18, 20, 21], by immunoblotting [15] or by ELISA [18]. The results of nucleic acid hybridization tests suggested between 23% and 41% sequence homology between any of three strains of IPCV and PCV [21]. Some authors have considered IPCV and PCV to be strains of one virus [4] but others [21, 22] have referred to them as separate viruses.

IPCV is thought to have a bi-partite genome consisting of a c. 5.4kb and a c. 4kb species [21]. Smaller RNA species were detected in preparations from purified virus particles [16] but their origins were not established. Translation *in vitro* of IPCV RNA-2 yielded coat protein and little else [16]. This suggested that the coat protein gene was near the 5'-end of the RNA, which is the position in the genome of the coat protein genes of other furoviruses [4]. In this paper we report the nucleotide sequence of the part of RNA-2 which encodes the coat protein of IPCV and comparisons among the amino acid sequences of coat proteins of other viruses with rod-shaped particles including the recently determined sequence of PCV coat protein [14].

Materials and methods

Propagation and purification of IPCV

The Hyderabad isolate of IPCV [18] was propagated in *Nicotiana clevelandii* and purified as described by Reddy et al. [21].

Preparation and cloning of IPCV RNA

RNA was extracted from purified virus particles as described by Mayo and Reddy [16]. It was annealed with random composition hexa-deoxynucleotides and reverse transcribed [11], and the resulting cDNA was cloned in *Sma*I-cut pUC19 [17]. Cloned cDNA was used as a probe in Northern blotting [23] of RNA extracted from IPCV particles in order to assign the cDNA clones to RNA-1 or RNA-2.

Nucleotide sequence determination

Sequence was determined by using dideoxy chain termination as described by Mayo et al. [17]. Sequence at the 5'-end of the RNA was obtained by primer extension as described by Fichot and Gerard [7]. The oligonucleotide used was complementary to the sequence between nucleotides 38 and 53 (Fig. 1). The sequence was assembled as described by Mayo et al. [17].

Amino acid sequence analysis

Sequences were compared by using programs in the GCG package [5] and CLUSTALV [13]. The significance of the clustering detected by CLUSTALV was assessed by bootstrapping 1000 samples. Amino acid sequences for comparisons were of the coat proteins of beet necrotic yellow vein furovirus (BNYVV) [1], barley stripe mosaic hordeivirus (BSMV) [12], *Nicotiana velutina* mosaic furovirus (NVMV) [19], peanut clump furovirus (PCV) [14], soil-borne wheat mosaic furovirus (SBWMV) [24] and tobacco mosaic virus (TMV) [28].

Results

Figure 1 shows the nucleotide sequence of the 5'-terminal 1128 residues of RNA-2. The sequence has been submitted to the EMBL database under Accession Number 76658. The non-coding leader is 501 nucleotides long and is followed by an open reading frame (ORF) which is the only potential ORF in the sequence. The encoded protein was identified as the coat protein of IPCV because the M_r of 23017 is close to the value of 24500 estimated by gel electrophoresis [16] and because the amino acid composition was a significant match (Chi squared = 18.2, with 15 degrees of freedom) to that reported for an Indian isolate of PCV (of unknown serotype) [6]. The only pronounced mis-match was between the 23 serine residues deduced from the sequence (Fig. 1) and the 12 found experimentally [6]. The putative sequence of IPCV coat protein also had strong homologies with the sequence of the coat protein of PCV [14].

Table 1 shows the result of comparisons made by using GAP among the sequences of the coat proteins of IPCV, PCV, SBWMV, BNYVV, NVMV, BSMV and TMV. The only marked similarities were 61% identity between IPCV and PCV coat proteins, and 37% and 36% identity between BSMV coat protein and those of IPCV and PCV, respectively. Otherwise there were no greater similarities among furovirus coat proteins than any had with TMV coat protein. Similarities between furovirus coat proteins and those of other

```

10          30          50          70          90          110
GCATTCTTGC TCGACTTTCGTGTAGCGAAAAGCGTATCAGCAGACAGCGGTTTACCATCCCTTGATCCAAAACCGGTTTACCTACACTCTATCTTTGGGTTGTATCATTGCGACTTTGTTA
130          150          170          190          210          230
TTGTGGAGTTCGATTTCGTTTCGACTGAGCTAGCCCTCGCCCTCCAGCTTAAACGACTTCGCAGGTTACCGGCTGTATCTTTAGCAGAGGGTAACGGGAATTACCATCGGTTGGCGGTAITCTGTTC
250          270          290          310          330          350
TATTACCGTGGTTTCCCAACCCGAGAAGTCTTATAGACTGAGTTTTTAAAACTCTGCCTCCGCGGTTTCGGAAAGTTCAGCTGAGAAAGCTGAAGACAGGGGCTGACCTTATCGGTTAC
370          390          410          430          450          470
TAGTCTTAGTTCGCGGTGCCATCCGCCCTCTTGATTTTGTATTGTGGAGTTCGATTTCGACTGAGCTAGCCCTCGCCTCAGCTTAAAGGATTCGCAAGCTTTGGCGGTGAGAC
490          510          530          550          570          590
TAGTCTTTTATTTTCCACCCATGCTAATTTGCTGAACTTACTCGTGGTGGTGGTCACTATGGAATAGAGCGCGTGGAGGAATCAGCTTGTGAGTGAAGGCTCAACTCTGACTGGTGG
610          630          650          670          690          710
      M S N L S E V T R G G G H Y G I E A W R N H V V S E R L N S D W W
ATAAAAAGTGATCGCTGGGACCAGCTCCTAGCGGATTTGAGAGCCGTGAACCTTTGAAAGTCAATAGTCTCGCTCCGAGGTCGCGCTATATTAAACGTTGCGCTAAAGATTTACCCGCT
I K S D R W D Q L L A D L R A V N F E V N S S R S E V A S I I N R V I K D L P A
730          750          770          790          810          830
GCTGTTTCAGCAGCTTTTCCGTTGCCAAGGGTGTTTGGGTAGCGATCTTATACCACCTTATATTACGTGCACGTGAAATCTGATCTTAAAGCAGAAGTCTACAGGCTTATTGCTGCT
A V S A R F P G A K G V L G S D T Y T T L Y Y V H V K S D L K Q K F Y R L I A A
850          870          890          910          930          950
GCTGACCAAGGAAGAATCGAGACTTGAAGTGGGGAGATCCTCTGTCCCGCCAGCCACAAGCAGCTCATCAAACAATCAAGCGGATGTACCTAGCAAGGGGATCAACCGAATAAGAGAT
A D Q G K N R D L E V G R S S V P A A T S S S S N N Q A I V P S K G I N A I R D
970          990          1010          1030          1050
CAGCAACTCTCAGAGCGGATCTCTTTTATCGGTACGATCTGTGGACATCGAACTTACTACTGTGAAACAGTACGATCAGTTCTGTTTGAATCGACGTTTAGTGTAACTGGATC
Q Q P L R D G S L S Y R Y D L L D I E L T T V K Q Y D Q F L F E S T F S V N W I
1090          1110
CCTAATCCAGCGGCCACTCCCAACCGCGCTGCCCGGGGGTGAATTA
P N P A A T P N A A A P G V N *

```

Fig. 1. Nucleotide sequence of the 5' 1128 nucleotides of IPCV RNA-2. The sequence is shown with that of the encoded coat protein. * signifies the termination codon

Table 1. Results of pairwise GAP comparisons between coat proteins

	IPCV	PCV	SBWMV	% identity BNYVV	NVMV	BSMV	TMV
IPCV		61	18	18	23	37	22
PCV	80		18	18	26	36	23
SBWMV	37	37		21	19	20	24
BNYVV	44	38	46		22	15	17
NVMV	41	40	40	42		22	16
BSMV	56	59	37	38	44		20
TMV	39	46	44	38	34	41	

% similarity

Values shown are the percent of identical residues (% identity) or the percent of matches of chemically similar amino acids (% similarity) obtained when comparing coat proteins of the viruses indicated. Default values of gap weight = 3 and length weight = 0.1 were used to obtain the alignments

tobamoviruses or tobacco rattle and pepper ringspot tobamoviruses were essentially the same as those obtained with TMV coat protein (data not shown).

The strong similarities among coat proteins of IPCV, PCV and BSMV (Table 1) were also obvious when the sequences were compared by using the NJTREE option in CLUSTALV. Figure 2 shows the 'unrooted tree' obtained by comparing coat proteins of IPCV, PCV, BSMV, SBWMV, NVMV and

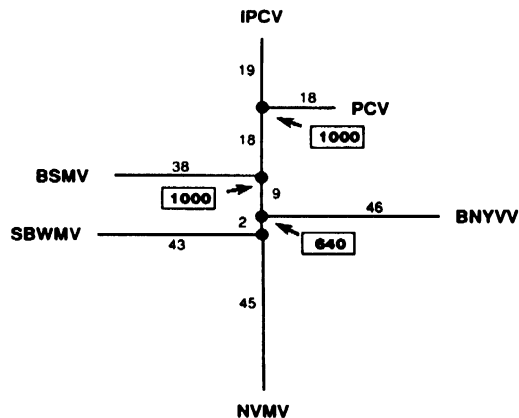


Fig. 2. An unrooted tree generated by CLUSTALV illustrating the similarity distances among the coat proteins of *IPCV* (Indian peanut clump furovirus), *PCV* (peanut clump virus), *BSMV* (barley stripe mosaic hordecivirus), *SBWMV* (soil-borne wheat mosaic furovirus), *NVMV* (Nicotiana velutina mosaic furovirus), and *BNYVV* (beet necrotic yellow vein furovirus). Unboxed numbers indicate the distances between the nodes; the boxed numbers are the numbers out of a bootstrapped sample of 1000 which gave the same branching pattern as that shown

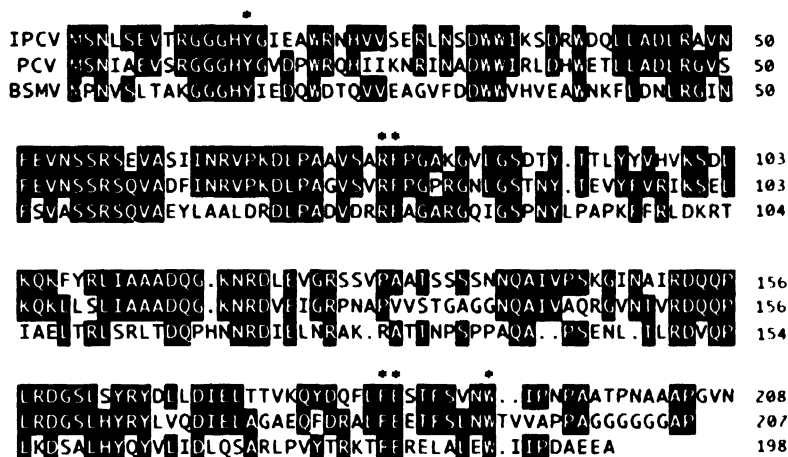


Fig. 3. Alignment of the coat proteins sequences of *IPCV* (Indian peanut clump furovirus), *PCV* (peanut clump virus), and *BSMV* (barley stripe mosaic hordeivirus). Amino acid matches are shown in reverse contrast. Numbering excludes gaps inserted by CLUSTALV. * indicates the residues common to coat proteins from viruses with rod-shaped particles and mentioned in the text

BNYVV. The only clustering detected was of IPCV with PCV and to a lesser extent of IPCV and PCV with BSMV; these clusters were judged significant because they occurred in all the bootstrapped samples (Fig. 2). 'Trees' which included tobamovirus and tobnavirus coat proteins showed essentially the same relationship among furovirus sequences as that shown in Fig. 2. Except for IPCV and PCV, the different furoviruses were as dissimilar from each other as they were from tobamoviruses or tobnaviruses.

Figure 3 illustrates the alignment obtained between the coat proteins of IPCV, PCV and BSMV and the extent and location of the amino acids identical in two of the three proteins. Some of the residues (asterisked in Fig. 3) which were the same in all three sequences (Y 14, R 78, F 79, F 186, E 187, W 194) correspond to those identified [10] as being present in analogous positions in the coat proteins of both tobnaviruses and tobamoviruses (Fig. 3). Most of these residues are also present in the coat protein sequences of other furoviruses.

Although the sequence alignments do not suggest any feature distinguishing furovirus coat proteins from others, Fauquet et al. [6] were able, by a multi-dimensional classification of the amino acid compositions of plant virus coat proteins, to detect a clustering of furoviruses, and a similarity of the whole cluster with BSMV.

Discussion

Some authors have regarded IPCV and PCV as being strains of one virus [4] whereas others have considered the viruses to be separate on the grounds of a lack of serological relationship and only limited nucleotide sequence homo-

