Arch Virol (2003) 148: 83–98
DOI 10.1007/s00705-002-0900-9

Archives of Virology
Printed in Austria

Molecular diversity of RNA-2 genome segments in pecluviruses causing peanut clump disease in West Africa and India*

R. A. Naidu[1,2], **S. Sawyer**[3], and **C. M. Deom**[1]

[1]Department of Plant Pathology, University of Georgia, Athens, Georgia, U.S.A.
[2]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India
[3]Departments of Mathematics and Genetics, Washington University, St. Louis, Missouri, U.S.A.

Received May 8, 2002; accepted August 9, 2002
Published online October 24, 2002

Summary. The complete nucleotide sequence of RNA-2 genome segments of four isolates of *Peanut clump virus* (PCV) and two isolates of *Indian peanut clump virus* (IPCV) were determined. Comparisons among the complete RNA-2 sequences of six isolates from this study and two published earlier, revealed a high degree of variability in size (between 4290 and 4652 nucleotides) and nucleotide sequence identities (between 58 % and 79 %). Amino acid sequence alignments of the five open reading frames (ORF) showed that ORF 4, which encodes the second of the triple gene block proteins, is highly conserved (90 to 98 % identical) whereas the protein encoded by ORF 2, whose function is unknown, is less conserved (25 to 60 % identical). The coat protein of the eight isolates showed amino acid identities between 37 % and 89 % and contained several conserved residues. Phylogenetic comparisons, based on complete RNA-2 sequences, revealed that the eight isolates grouped into two distinct clusters with no geographical distinction between PCV and IPCV isolates. Phylogenetic tree topologies for individual ORFs showed an overall similarity with that obtained from entire RNA-2 sequences, although the relative positions of individual isolates vary within each cluster. The results indicate that there is substantial divergence among the RNA-2 genomes of pecluviruses and suggest that different proteins have evolved differently, possibly due to different selection pressures.

*The nucleotide sequence reported in this paper has been deposited at GenBank as accession numbers AF447396, AF447397, AF447398, AF447399, AF447400, AF447401.

Introduction

The genus *Pecluvirus* contains two species: *Peanut clump virus* (PCV) and *Indian peanut clump virus* (IPCV) [8]. PCV occurs in West Africa and IPCV in the Indian subcontinent. Both viruses induce similar, though not identical, disease symptoms in peanut (= groundnut, *Arachis hypogaea* L.) called peanut clump [30, 35]. Diseased plants typically show severe stunting due to shortened internodes and the leaves appear dark green with reduced size resulting in plants appearing bushy or clumped. In some cases, the diseased plants show variable symptoms without clumping [18]. Peanut clump disease causes considerable economic impact on peanuts in certain regions of Africa and the Indian subcontinent [29]. PCV and IPCV are soil-borne and reported to be transmitted by *Polymyxa* sp. [30, 35]. Both viruses are also transmitted by mechanical sap inoculation and through seed [15, 28, 30, 35]. PCV and IPCV have extremely wide host ranges, which include monocotyledonous and dicotyledonous plants such as peanuts, pearl millet, finger millet, sorghum, barley, wheat and sugarcane [6, 18].

PCV and IPCV contain bipartite genomes encapsidated separately in rod-shaped particles of bimodal length [30, 35]. The RNA-1 and RNA-2 of PCV (isolate from Senegal [34], designated PCV-S in this study) are 5897 and 4504 nucleotides (nt) in length [11, 17]. The RNA-1 of an isolate of IPCV (designated IPCV-H) and RNA-2 of another isolate (designated IPCV-L) are 5841 and 4290 nts in length, respectively [23, 25]. Sequence comparisons showed that the RNA-1 of PCV and IPCV is more conserved than RNA-2 [23, 25].

Based on the serological relationships, using polyclonal antibodies, IPCV isolates have been distinguished into three serotypes [26]. A comparative analysis of their coat protein (CP) sequences revealed significant diversity and indicated that they are as distantly related to each other as they are individually to an isolate of PCV [25, 37]. Huguenot et al. [14] and Manohar et al. [18] reported, using a panel of monoclonal antibodies, antigenic diversity among several isolates of PCV in West Africa. However, information on the variability in CP sequences of PCV isolates is not available to determine the extent of nucleotide and amino acid diversity among themselves and with the IPCV serotypes.

In this study, we determined the complete nucleotide sequence of RNA-2 of four different isolates of PCV from three different countries in West Africa. We also obtained sequence information for the previously unsequenced portions of the RNA-2 of two IPCV isolates studied earlier [25, 37]. Comparisons among these sequences, and those published earlier [11, 17, 25] revealed a high degree of variability in RNA-2 genomes of PCV and IPCV isolates.

Materials and methods

Virus isolates

The origin of different isolates of PCV and IPCV used in this study is summarized in Table 1. The IPCV-L, -D, and -H isolates, belonging respectively to the L, T, and H serotypes of IPCV [26], were from the same source described in previous studies [25, 37]. The four PCV isolates

Table 1. Origin and lengths of the coding and noncoding regions of RNA-2 molecules of different IPCV and PCV isolates

Geographical origin			Length (in nt) of genomic regions									
Isolate	Place	Country	Full length	5′ NCR	ORF 1	ORF 2	IGR	Triple gene block			3′ NCR	Accession number
								ORF 3	ORF 4	ORF 5		
India												
IPCV-D	Durgapura	India	4310	394	663	909	94	1329	369	465	280	AF239731, AF447396
IPCV-H	Hyderabad	India	4507	502	627	1020	117	1323	369	468	275	X76658, AF447397
IPCV-L	Ludhiana	India	4290	388	603	1065	84	1230	369	465	280	AF239729
West Africa												
PCV-S	Sine Saloum	Senegal	4504	390	624	1104	135	1341	369	462	276	L07269
PCV-Ni	Niamey	Niger	4305	401	624	960	173	1236	369	462	277	AF447398
PCV-N	Maradi	Niger	4553	422	702	1059	153	1302	369	465	279	AF447399
PCV-M	Sanankaroba	Mali	4658	563	678	1071	134	1296	366	462	279	AF447400
PCV-B	Bobo Dioulasso	Burkina Faso	4466	392	678	1020	161	1296	369	465	279	AF447401

NCR, non-coding region; *ORF*, open reading frame; *IGR*, intergenic region

(PCV-N, PCV-Ni, PCV-M, and PCV-B) were obtained from infected peanut plants collected during a disease survey in 1996 [24]. In preliminary studies, PCV-Ni, but not PCV-N, -M, and -B, showed a positive serological reaction with polyclonal antibodies to PCV-S from Senegal [34]. All four PCV isolates, like PCV-S, did not cross react with polyclonal antibodies to the three IPCV serotypes.

RNA extraction and RT-PCR amplification

Nicotiana benthamiana plants were inoculated separately with PCV-Ni, PCV-N, PCV-M, and PCV-B isolates using extracts from infected peanut leaves. After a single mechanical inoculation to *N. benthamiana*, total RNA was extracted from infected leaf tissue using a RNAeasy Plant Mini Kit (Qiagen Inc., USA). In the case of the IPCV-H and IPCV-D isolates, viral RNA preparations extracted earlier from purified particles were used [25].

For RT-PCR, two sets of primers were used to amplify RNA-2 specific sequences. Primers were designed based on conserved sequences identified by aligning the RNA-2 sequences of IPCV-L (accession number AF239729) and PCV (accession number L07269). Primers PCW1 (5′ACCATCCCTTGTATCCAA) and PCW2 (5′GCAAGATTTCAGCACTTTG) correspond, respectively, to nt 54–71 (sense) and to nt 2802–2820 (complementary) in IPCV-L RNA-2 [25], and amplify a fragment of about 2.8 kb representing most of the 5′ non-coding region (NCR), ORFs 1 and 2, and part of ORF 3. Primers PCW3 (5′CAAAGTGCTGAAATCTTGC) and PCW4 (5′TGGGATGGATATCGCTCCG) correspond, respectively, to nt 2802–2820 (sense) and to nt 4272–4290 (complementary), and amplify a fragment of about 1.4 kb representing part of ORF 3, ORF 4, ORF 5, and the 3′ NCR. The two amplified fragments overlap between nt 2802–2820 in ORF 3 and together represent most of RNA-2 except for approximately 50 nt at the 5′ end (Fig. 1).

Reverse transcription was performed on either total RNA or viral RNA with primers PCW2 and PCW4 at 42 °C for 1 hr using SuperScript Preamplification System (Life Technologies, Grand Island, NY). After denaturation at 95 °C for 10 min, the cDNA was directly used as a template for PCR amplification in a Perkin-Elmer 9600 Thermal Cycler. A PCR program of 94 °C (30 s), 55 °C (30 s), 72 °C (2 min) for 35 cycles with a final extension at 72 °C for 10 min was performed. The 5′ ends of RNA-2 were amplified using the GeneRacer™ Kit (Version D, Invitrogen, Carlsbad, CA). A GeneRacer 5′ nested primer (supplied by the manufacturer) and an internal primer specific to each isolate were used for amplifying a fragment of about 500 nt.

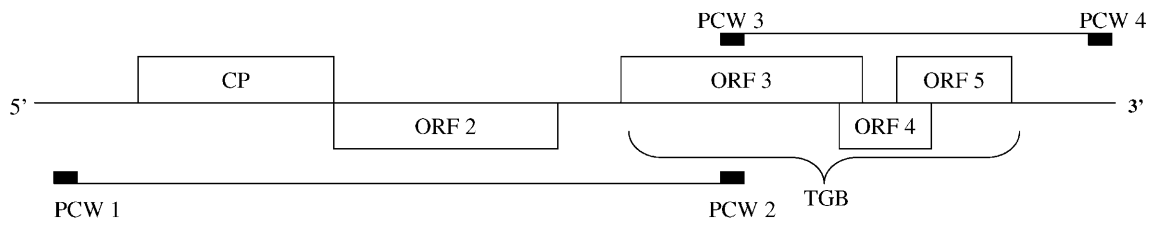

Fig. 1. Diagram showing genome organization of RNA-2 of pecluviruses. The position of open reading frames encoded by RNA-2 is indicated as open boxes above and below the line. Position of different primers (solid boxes) used for RT-PCR represent the conserved sequences in the RNA-2 genome. Lines connecting the primers represent the length of amplified fragments. Primers PCW2 and PCW3 are complementary. Primers PCW1 and PCW2 amplify a fragment of ∼2.8 kbp, and primers PCW3 and PCW4 amplify a fragment of about ∼1.4 kbp. *TGB* = Triple gene block

Cloning and nucleotide sequencing

The amplified DNA fragments were gel purified, cloned into pGEM-T Easy vector (Promega Corporation, U.S.A.), and transformed into *E. coli* DH5α. After blue-white selection, plasmid DNA was prepared from single isolated white colonies grown in LB broth in the presence of ampicillin at 37 °C overnight using a Qiagen plasmid kit. For each isolate, a directed sequencing protocol ("DNA walking") was followed using progressive isolate-specific oligonucleotide primers designed based on the partial nt sequence obtained. In all cases, at least two independent clones were sequenced in both directions. Whenever necessary, additional clones were sequenced to confirm regions of ambiguity and a consensus sequence was obtained. In some cases, the gel-purified RT-PCR fragments were sequenced directly in both directions using specific primers. Nucleotide sequences were edited, assembled and analyzed using DNASIS (Hitachi, San Bruno, CA). Pairwise comparisons of nucleotide and amino acid sequences were made using GAP program within the UWGCG package.

Phylogenetic analyses

The RNA-2 nucleotide sequence data generated from this study has been deposited in the EMBL Database under accession numbers shown in Table 1. The partial sequences reported earlier for RNA-2 of IPCV-H (Accession nos. X76658 and AF239730) and IPCV-D (Accession no. AF239731) matched with the respective RNA-2 sequences obtained in this study, except for the C-terminal sequence reported earlier for ORF 2 of IPCV-D (Accession no. AF239732). Therefore, RNA-2 sequences for IPCV-H and -D isolates obtained in this study were included for sequence alignments and phylogenetic analyses. The published sequences for RNA-2 of IPCV-L (Accession no. AF239729) and PCV-S (Accession no. L07269) were used in this study.

Multiple alignments of nucleotide sequences and predicted amino acid sequences of the proteins encoded by different ORFs were obtained using CLUSTAL W from Network Protein Sequence Analysis (NPSA, http://npsa-pbil.ibcp.fr) with the suggested default settings. Phylogenetic analyses were performed by using the neighbor-joining method with the Kimura two parameter model (using programs BTEST and VTRSHOWDAT written by S. Sawyer). Phylogenetic trees were also derived by Maximum Likelihood settings (50 % majority-rule consensus tree with 100 bootstrap replicates) with a heuristic search method (stepwise addition: random, initial 'MaxTrees' setting: 100, branch-swapping algorithm: tree-bisection-reconnection) using PAUP* 4.0 beta 8 (Sinauer Associates, Inc., Sunderland, MA).

Results

RT-PCR

In RT-PCR, primer pairs PCW1 and PCW2, and PCW3 and PCW4 amplified fragments of the expected size from RNA 2 of PCV-B, PCV-M, PCV-N, PCV-Ni, IPCV-D, and IPCV-H isolates. In all cases, the RT-PCR amplified fragments were successfully cloned except the PCW1 and PCW2 amplified fragment of IPCV-H RNA-2. Repeated attempts to clone this fragment were unsuccessful and, therefore, the PCR amplified fragments were directly used for sequence analysis.

Diversity in RNA-2 sequences among isolates of PCV and IPCV

The complete RNA-2 of four PCV (PCV-B, -M, -N, and -Ni) and two IPCV (IPCV-H and -D) isolates were sequenced to gain a better understanding of the diversity

among different PCV and IPCV isolates. These sequences were then compared with published PCV-S and IPCV-L RNA-2 sequences [11, 25]. Variation in the length of the RNA-2 molecules of all eight isolates was observed (Table 1). IPCV-L was the shortest with 4290 nt and PCV-M was the longest with 4652 nt. Five ORFs were identified in each of the RNA-2 molecules. They are arranged in two blocks separated by an intergenic region (IGR) of varying length similar to those in RNA-2 of PCV-S and IPCV-L (Fig. 1), suggesting that the overall genome organization of RNA-2 was identical in all PCV and IPCV isolates. Multiple sequence alignments showed that the sequences in all eight isolates are noticeably more divergent in CP and ORF-2 than in ORFs 3, 4, and 5 (data not shown). In pairwise comparisons, the overall nt identity among the RNA-2 molecules ranged between 58 and 79 %. Based on the percent identity, the five PCV isolates fall into two groups, PCV-S and -Ni as one group with 79 % identity and PCV-B, -N, and -M as a second group with 72 % identity. Nucleotide identities between these two groups varied from 58 to 64 % and either of them showed similar identity when compared with the three IPCV isolates, suggesting that they are as distantly related to each other as they are to the IPCV isolates. Among the three IPCV isolates, the greatest identity of 77 % was observed between IPCV-H and -D, whereas IPCV-L showed 65 and 68 % identity with IPCV-D and IPCV-H isolates, respectively. These results indicate that IPCV isolates fall into two groups with IPCV-H and -D as one group distinct from IPCV-L. The nt identity also suggests that IPCV-L is as distantly related to either of the two other IPCV isolates as to the five PCV isolates.

Diversity in RNA-2 encoded proteins among isolates of PCV and IPCV

The CP ORF of the five PCV and three IPCV isolates vary from 603 nt (IPCV-L) to 702 nt (PCV-N) in length (Table 1). The CPs of PCV-Ni and PCV-S are the most closely related with 79 % and 89 % identity at the nucleotide and amino acid level, respectively. This value, together with their serological relatedness, suggests that PCV-Ni might be a strain of PCV-S. The CPs of PCV-B, -N, and -M showed higher nucleotide (65–67 %) and amino acid (74–77 %) identities among themselves than with the CPs of the other two African or three Indian isolates. The CPs of IPCV-D, -H, and -L, on the other hand, showed nucleotide and amino acid sequence identities between 58–65 % and 61–66 %, respectively, among themselves and with PCV-S and -Ni. These results indicate that the three Indian isolates (IPCV-D, -H, and -L) and the two African isolates (PCV-S and Ni) form one group whereas the other three Africa isolates (PCV-B, -N, and -M) form a separate group.

Multiple alignments of CP sequences between the five PCV and three IPCV isolates showed conservation of several amino acid residues, with a preponderance of hydrophobic residues, at different positions across the entire CP (Fig. 2). Several important residues, including arginine and glutamic acid in the central and C-terminal parts (indicated by asterisks), respectively, that have proposed structural and functional roles in the CPs of rod-shaped viruses like tobamo- and tobraviruses [7, 10] are well conserved among all isolates. The C-terminus,

```
IPCV-D   M-ANISEVRRGGGHYSIASWRNHVIRQNLNHDWWIRSDRWAQLLADLRAVNYEVNSSRSE   59
IPCV-H   M-SNLSEVTRGGGHYGIEAWRNHVVSERLNSDWWIKSDRWDQLLADLRAVNFEVNSSRSE   59
IPCV-L   MTTNISEVARGNGHYNVEGWRNHIVRSRANADWWIRSDLYDQLLADLRAVNFEVNTSRSE   60
 PCV-S   M-SNIAEVSRGGGHYGVDPWRQHIIKNRINADWWIRLDHWETLLADLRGVSFEVNSSRSQ   59
 PCV-Ni  M-TNIAEVSRGGGHYGVEPWRQHIVKNGINTDWWIRLDHWANLLASLRGVNFEVNSSRSQ   59
 PCV-N   M-PNINLGHLGNGHYSEDVFEEKIYEAGVDQDWWIEQSNYDKLLSGLRAVNFEVNSSRQQ   59
 PCV-M   M-PNINLGHLGGGHYFEDVFEQKIYDAGIDSDWWIESEAYDKLLAGLRAINFEVNSSRQQ   59
 PCV-B   M-PNINLGALGGGHYPEEVFEQKIYDTGVDSDWWIEVEAYDKLLSGLRSINFEVNSSRQQ   59

IPCV-D   VANIINRVPKDLPADPSARFPGVVGTPGETNYSLVYYVRVEAPVREKFLGIIAAADQGKS  119
IPCV-H   VASIINRVPKDLPAAVSARFPGAKGVLGSDTYTTLYYVHVKSDLKQKFYRLIAAADQGKN  119
IPCV-L   VSAIINRVPKDLPAAVSARFPGARGVLGSDTYKEIYFVRIKPELKQRFLNLIAAADQGKN  120
 PCV-S   VADFINRVPKDLPAGVSVRFPGPRGNLGSTNYTEVYFVRIKSELKQKLLSLIAAADQGKN  119
 PCV-Ni  VADFINRVPKDLPAGVSARFPGPRGTLGSTNYNEVYFVRVKPDLKQKLLSLIAAADQGKN  119
 PCV-N   VAKLINNVDKDLPANPGVRFPASRGTLGSADYKPYRFRRIDSTNRGYFEDLVRISDQGKS  119
 PCV-M   VAQLINHVDKDLPANPGVRFPKHRGTLGHADYRPLRFRRIDAQTRAYFEALVRISDQGKS  119
 PCV-B   VAKLINNVDKDLPANPGVRFPGSRGTLGHSDYRPFRFRRIDATTRGFFEDLVRISDQGKS  119
                          *

IPCV-D   RDVEVGRPSAPSVASGAG-NQ-AIVPVRGVNAIRDQQPLRDGSLSFRYKLVDVELASVTQ  178
IPCV-H   RDLEVGRSSVPAATSSSSNNQ-AIVPSKGINAIRDQQPLRDGSLSYRYDLLDIELTTVKQ  179
IPCV-L   RDIEVGRPTAPTVTNSAGGNQ-AIVAQRGVNAVRDTQPMRDGALHFRYELRDIELAVVDQ  180
 PCV-S   RDVEIGRPNAPVVSTGAGGNQ-AIVAQRGVNTVRDQQPLRDGSLHYRYLVQDIELAGAEQ  179
 PCV-Ni  RDVEIGRPNAPVVTTGAGGNQ-AIVAQRGVNTVRDQQPLRDGSLHYRFLLQDIEVFSVEQ  179
 PCV-N   RDTEINRPAAPTVGSNSQGTSLANN--RGLSVIRDEQPLRDGAQNFGHTFTALSLSNIPT  178
 PCV-M   RDAEVNRPTAPTIASNSAGTSLANN--RGVNVVRDDQPLRDGAQNFGHTFGYIALANPPS  178
 PCV-B   RDTEINRPAAPTVGSNAQGTSLANN--RGLSVIRDEQPLRDGAQNYGHTFTTISLSNPVT  178

IPCV-D   FDQVLFEETFQITWIEKDASTKTTPTASTNTSPTGVAPGDPSN                   220
IPCV-H   YDQFLFESTFSVNWIPNPAATPNAAAPGVN                                208
IPCV-L   FDQVIFEDVFKLTWTPAQPGP                                         200
 PCV-S   FDRALFEETFSLNWTVVAPPAGGGGGGAP                                 207
 PCV-Ni  FDRALFEETFSLNWTVVAPAAGGGGGGAP                                 207
 PCV-N   YNQRKFEREFSLSWTAASGHHTTGTGAGTSGSGTGSGTGTGGARPTTTGGNTAGNP      233
 PCV-M   YNQKKFESDFTLTWTATGTSGTGSTTGTGTGTGSGARPATTGNTGGAP              225
 PCV-B   YSRRKFESEFNLVWSASGGTHPTGTGTSVATGTGTTGTGRTGAGGNAP              225
               *       ↑
```

Fig. 2. Amino acid sequence alignment of coat proteins of PCV and IPCV isolates. Conserved residues in all isolates are shown in bold. The residues, Arginine and Glutamic Acid, involved in salt bridge formation in the CPs of rod-shaped viruses like *Tobacco mosaic virus* (TMV) and *Tobacco rattle virus* (TRV), are indicated by asterisks. The sequences at the C-terminus, downstream of the conserved Tryptophan residue (indicated by an arrow) are highly variable in PCV and IPCV isolates

downstream of the conserved tryptophan residue (indicated by arrowhead), is highly variable among the isolates; with PCV-N having the longest and IPCV-L the shortest. This region is similar in size in PCV-S and -Ni, and PCV-M and -B; however, the amino acid composition is identical (except at one position) between PCV-S and -Ni and different between PCV-M and -B.

ORF-2 is highly variable in length among all PCV and IPCV isolates, ranging between 909 and 1104 nt (Table 1). In pairwise comparisons, the amino acid sequence identity values for the ORF-2 encoded protein ranged between 25 % and 60 % depending on the viruses compared. Multiple alignment of the ORF-2 translation product showed a marked discontinuity in the degree of similarity among N-terminal and among C-terminal portions. In the N-terminal 180 amino acids, a total of 22 identical and 29 similar residues were found among all isolates, whereas no identical or similar residues were found in the remaining C-terminal portion of these isolates. It is also interesting to note that the "leucine zipper" (L-x_6-L-x_6-L-x_6-L) reported earlier in IPCV-L between residues 200 and 221 [25], is present only in IPCV-D (residues 70 and 91) and PCV-Ni (residues 265–286).

Of the three ORFs in the triple gene block (TGB), ORF 4 and 5 are most similar in size among the eight isolates (Table 1). However, ORF 4 is highly conserved with 81–92 % and 90–98 % identity, whereas ORF 5 showed 64–89 % and 54–81 % identity at the nucleotide and amino acid level, respectively, among these isolates. The nucleotide and amino acid identities of ORF 3 ranged, respectively, between 63–88 % and 66–92 % depending on the viruses compared. Of the three TGB proteins, multiple alignments of the protein encoded by ORF 3 of all eight isolates showed marked discontinuity in the degree of similarity among N-terminal and among C-terminal portions. Whereas in the C-terminal 345 amino acids that showed good alignment, 233 (67 %) residues were identical in all isolates and in the remaining N-terminal portions of the proteins, no amino acid residues could be aligned in all eight sequences.

Previous studies have shown that the TGB proteins of IPCV-L and PCV-S have greater similarities with *Barley stripe mosaic virus* (BSMV) and *Potato mop top virus* (PMTV) than with other TGB-containing viruses [11, 25]. The sequence data for the PCV and IPCV isolates obtained in this study also shows similarities with BSMV and PMTV and contain several conserved domains found in respective proteins of hordeiviruses [33]. Thus, ORF 3 of all PCV and IPCV isolates contain the putative NTPase/helicase motifs, ORF 4 contains a strongly conserved central region flanked by potentially membrane-spanning or -associated hydrophobic domains, and ORF 5 also contains hydrophobic domains [11, 33] (data not shown).

The 5′ and 3′ NCRs show distinct patterns of similarity among PCV and IPCV isolates

Comparative analyses of the NCRs among all eight isolates showed greater variability in length in the 5′ NCR than in the 3′ NCR (Table 1). Pairwise comparisons of the 5′ NCRs revealed that the sequence in IPCV-H is more similar to the 5′ NCR of IPCV-D (76 % identity). The 5′ NCR is more similar among the five PCV isolates (68–81 % identity). However, IPCV-H and -D and the five PCV isolates showed 41–57 % identity with the corresponding region in IPCV-L. The first 70 nt are 99 % identical between IPCV-H and -D, and 100 % identical among the five PCV isolates. The 5′ 70 nt in IPCV-L, on the other hand, are significantly

less identical with IPCV-H (67 %), -D (71 %) and the five PCV isolates (89 %). Curiously, the 5′ NCR of PCV-M contains a 36 nt repeat (between nt 313–348 and 440–475) with one mismatch. It is unlikely that this insert is an artifact of cloning, RT-PCR or sequencing, because the 5-terminal sequence of PCV-M was determined and verified from six independent clones.

In contrast, the 3′ NCR is well conserved among the eight isolates with greater than 85 % nucleotide identity. The sequences of the tRNA-like structures, reported earlier at the 3′ end of PCV-S RNA-2 [9], are highly conserved among the eight isolates. A few nucleotide differences in the tRNA-like structure were observed in IPCV-L, -H and PCV-B, -N, and -M. These changes, however, appear to have no significant effect on the overall tRNA-like structure (data not shown). In contrast to an incomplete anticodon (GΔC) present in PCV-S, the intact anticodon for valine (GAC) is conserved in other isolates.

Phylogenetic relationships

Phylogenetic relationships of the five PCV and three IPCV isolates were evaluated using entire RNA-2 sequences as well as different coding and non-coding regions. Trees generated using the neighbor-joining method distinguished the eight isolates into two clusters (Fig. 3a). Maximum likelihood trees had similar topologies and generally showed very similar results (data not shown). In both analyses, two of the five West African isolates (PCV-S and -Ni) shared a common node with the three IPCV isolates to form one cluster, while the other three African isolates (PCV-B, -N, and -M) formed a second cluster. This indicates that there is no obvious relationship between the phylogeny of these viruses and their geographical origin. Within the IPCV cluster, IPCV-H and -D formed one group suggesting that they diverged from one another after IPCV-L had diverged from their common ancestor. Similarly, within the PCV-B, -N, and -M cluster, PCV-B and -M grouped together and diverged from PCV-N. Phylogenetic tree topologies for individual ORFs of all PCV and IPCV isolates showed an overall similarity with that obtained from the entire RNA-2 sequences (Fig. 3b–f). However, the relative positions of individual isolates within each cluster vary when different ORFs were used for generating the phylograms. On the otherhand, the phylograms derived from 5′ NCR and 3′ NCR sequences displayed different topologies (Fig. 3g, h). Such a phylogenetic incongruence between PCV and IPCV isolates suggests possible recombination in the RNA-2 sequences.

A preliminary analysis of the RNA-2 sequences using the program GENECONV [27, 31] revealed a total of 22 putative recombination events among the PCV and IPCV isolates that were significant using a multiple-comparison-corrected permutation test ($P < 0.001$) (data not shown). The most significant of these putative events (1431 nt long) was observed between IPCV-H and -D sequences spanning from the first quarter of ORF 3 through the very end of ORF 5. The remaining 21 events were mapped within the 5′ and 3′ NCRs with recombination fragments ranging in size from 50 to 321 nt. Fourteen of these recombinant events were observed in the 3′ NCR between all PCV and IPCV isolates and the remaining seven were observed in the 5′ NCR only between PCV isolates.

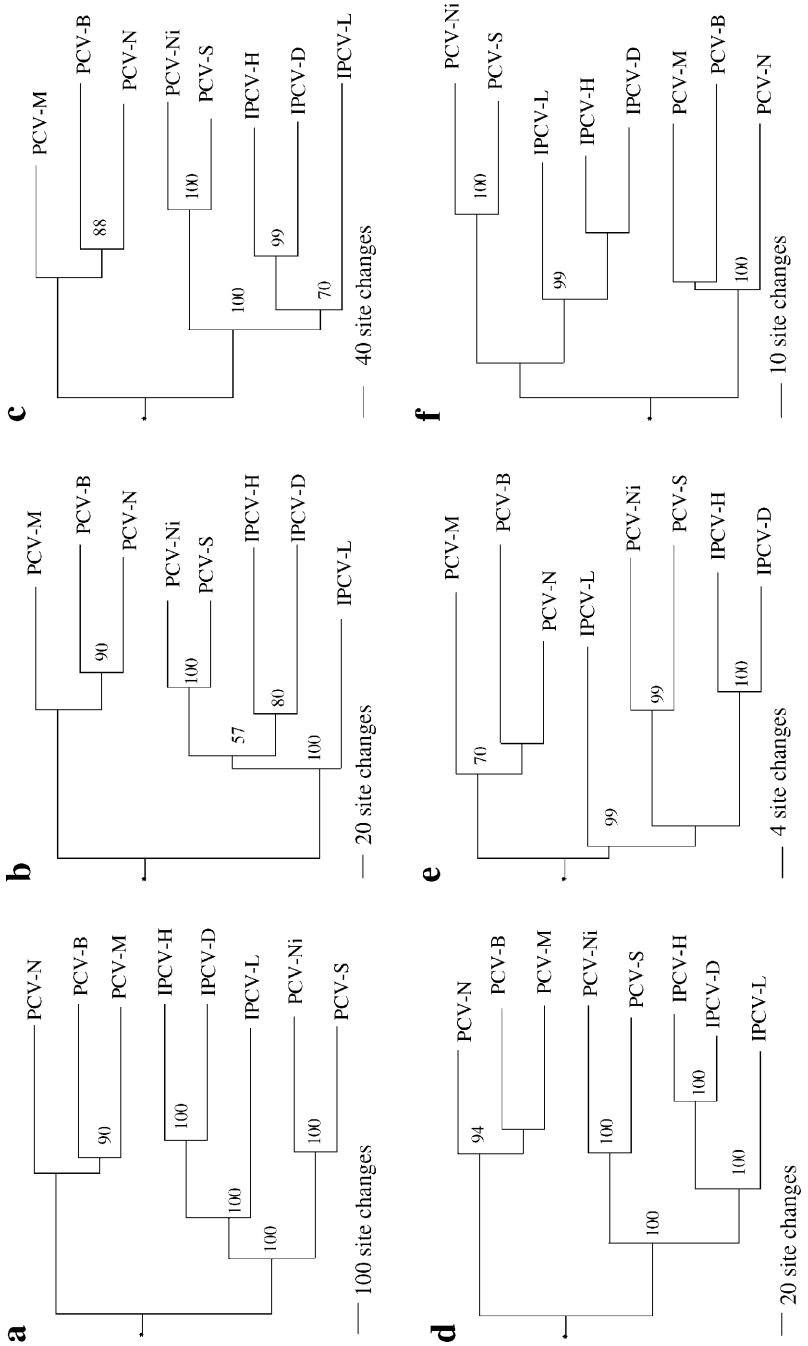

Fig. 3 (*continued*)

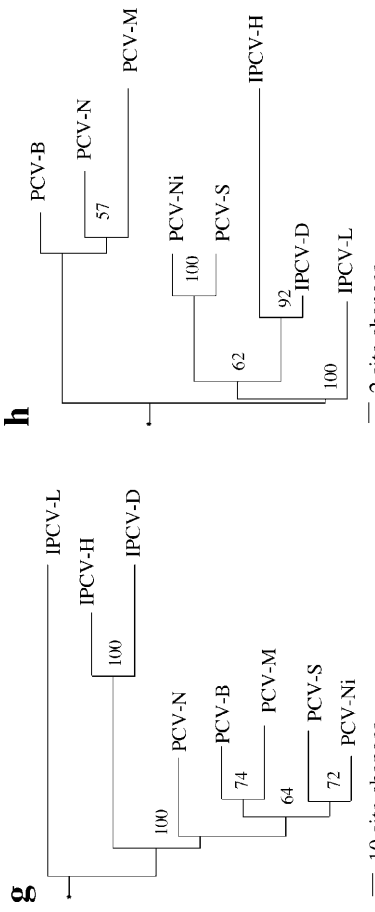

Fig. 3. Neighbor-joining phylogenetic trees indicating the relationship between PCV and IPCV RNA-2 genome sequences (**a** complete RNA-2 sequence, **b** CP, **c** ORF-2, **d** ORF-3, **e** ORF-4, **f** ORF-5, **g** 5′ NCR, **h** 3′ NCR) using Kimura two-parameter distance corrections. The estimated trees were arbitrarily rooted at that interior point of the graph whose maximum distance to the nodes is minimized. Numbers attached to nodes are the percentages of 1000 boostrap replicated trees in which that node exists. Bootstrap percentages less than 50 are not shown

Discussion

The present study provides a comprehensive analysis of the molecular diversity among RNA-2 genomic segments of PCV and IPCV isolates infecting peanuts, respectively in West Africa and India. Although the genome organization among all isolates is similar, the comparative data clearly showed that there is substantial divergence among the RNA-2 genomes of PCV and IPCV isolates. This diversity is not random and is more structured in that the CP and ORF-2 translation products, located in the 5′ part of RNA-2 genomic segments, are highly variable whereas the TGB proteins, located in the 3′ part, are more conserved. Such diversity suggests that different proteins encoded by RNA-2 genomes of pecluviruses have evolved differently, possibly due to different selection pressures. Thus, the TGB proteins, involved in viral spread through plants [13], seem to be under strong selection pressure to maintain sequence integrity and function, and the CP and ORF-2 are under divergent selection pressure to allow adaptation to different host-vector niches.

PCV and IPCV were classified into two species in the genus *Pecluvirus* largely based on their geographical origin and serological specificity [8]. The genomic sequence comparisons between PCV and IPCV isolates revealed that RNA-2 gene products vary more than RNA-1 products [11, 17, 23, 25]. This pattern of genome variation in PCV and IPCV is similar to that reported in TRV; RNA-1 molecules of TRV strains are relatively similar whereas RNA-2 molecules are distinct [22]. Based on the guidelines proposed for virus classification [20, 36], an overall RNA-2 sequence identity between 58 and 79 % among the five PCV and three IPCV isolates would indicate their separation into distinct species. Nevertheless, nucleotide sequences of RNAs 1 of PCV and IPCV isolates and compatibility studies involving their genomic RNAs would provide further information in determining whether they should be considered either as distinct species along the lines suggested earlier [25] or genetically divergent strains of the same species as in the case of TRV.

Similar to TRV, there is a lack of correlation between the serological relatedness of PCV and IPCV isolates and their geographic distribution. For example, the two serologically related PCV-S and PCV-Ni isolates occur in Senegal and Niger, respectively, and the three serologically distinct isolates of IPCV occur in geographically widely separated locations in India [26]. In contrast, serologically unrelated isolates of PCV were found to occur within the same field, for example in Senegal [14, 18]. No information, however, is available on whether antigenically distinct isolates of IPCV occur within the same field. Since pecluviruses are seed-borne in peanut and several monocotyledonous species [6, 15, 28, 35], it is also unknown whether seed exchange contributes to introduction of PCV and IPCV isolates to new areas. Due to the limited value of serology in diagnosing diverse isolates of PCV and IPCV, RT-PCR assays, using sequence information reported in this study, will be useful in rapid evaluations of the variability among pecluviruses and permit the generation of data for detailed molecular epidemiological studies of peanut clump disease both in West Africa and India.

The CPs of all PCV and IPCV isolates showed conservation of several residues (Fig. 2) that were found in tobraviruses and tobamoviruses, which have similar

rod-shaped particle structure [10]. Thus, in analogy with the structure predictions for the CPs of TMV and TRV [2, 10], all PCV and IPCV isolates have structural features similar to those of tobraviruses and tobamoviruses with their N and C termini located on the external surface of the virus particle. The C-terminal domain of the CPs of PCV and IPCV isolates is highly variable, and by analogy with previous studies on tobraviruses [5, 16, 21], this domain could presumably be exposed on the surface of the virus particles and have a high mobility. In depth studies on the three-dimensional molecular structure of the CP by epitope mapping [16] and Raman optical activity [2] would provide insights into pecluvirus particle structure and function at the molecular level.

Circumstantial evidence indicates that the protein encoded by ORF-2 might be involved in transmission by the fungal vector as deletions in ORF-2 in isolates that were maintained by mechanical inoculations have frequently been observed [17]. This protein is expressed by a leaky scanning mechanism *in vitro* [12] and seems to be unique to PCV and IPCV since a protein(s) analogous to ORF 2 is not present in the genomes of other characterized fungus-transmitted viruses [1]. Thus the role of ORF-2 encoded protein in vector transmission and whether variability in this protein as well as CP reflects specificity in interactions between different PCV and IPCV isolates and the vector remains to be studied.

Considering the wide diversity in RNA-2 genomes of PCV and IPCV isolates infecting peanuts in Africa and India, we speculate that these isolates have evolved independently, and were already widely perpetuating, possibly on indigenous host plants, viz. perennial grass species. It is likely that PCV and IPCV isolates cause symptomless infections in indigenous reservoir host species in which they have co-evolved and to which they are well-adapted but cause severe damage to introduced susceptible species to which these viruses are not adapted. This is plausible since peanut, a native of South America, is an introduced crop to Africa and Asia and therefore has become a 'new-encounter' host [4]. This argument is supported by the lack of resistance in cultivated peanut to clump disease [19, 30] indicating the absence of a co-evolution of peanut-PCV/IPCV pathosystem. Deploying CP-mediated transgenic resistance for controlling clump disease in peanuts [32] might be of less practical possibility due to low levels of identity in the CP sequences of different PCV and IPCV isolates [25, present study] and the occurrence of heteroencapsidation when transgenic tobacco plants expressing IPCV-H CP were challenge inoculated with IPCV-L [3]. Nevertheless, RNA-2 sequences obtained in this study provide useful information in devising alternative strategies, perhaps utilizing the highly conserved sequences in the TGB proteins, in developing resistance to a diverse range of pecluviruses causing clump disease in India and West Africa.

Acknowledgements

R. A. Naidu is grateful to Dr. M. A. Mayo, Scottish Crop Research Institute, Dundee DD2 5DA, United Kingdom, for providing facilities to maintain PCV isolates and prepare cDNA clones, and to Mr. I. M. Roberts, Scottish Crop Research Institute, for help with serology. Suggestions and comments made by John L. Sherwood, Russell L. Malmberg and U. Melcher

are gratefully acknowledged. Part of this work was supported by funds from the Peanut Collaborative Research Support Program (Peanut CRSP) of the United States Agency for International Development (Grant LAG-G-00-96-90013-00).

References

1. Adams MJ, Antoniw JF, Mullins JGL (2001) Plant virus transmission by plasmodiophorid fungi is associated with distinctive transmembrane regions of virus-encoded proteins. Arch Virol 146: 1139–1153
2. Blanch EW, Robinson DJ, Hecht L, Barron LD (2001) A comparison of the solution structures of tobacco rattle and tobacco mosaic viruses from Raman optical activity. J Gen Virol 82: 1499–1502
3. Bragard C, Duncan GH, Wesley SV, Naidu RA, Mayo MA (2000) Virus-like particles assemble in plants and bacteria expressing the coat protein gene of Indian peanut clump virus. J Gen Virol 81: 267–272
4. Buddenhagen IW (1983) Crop improvement in relation to virus diseases and their epidemiology. In: Plumb RT, Thresh JM (eds) Plant virus epidemiology. Blackwell Scientific Publications, Oxford, pp 25–37
5. Brierley KM, Goodman BA, Mayo MA (1993) A mobile element on a virus particle surface identified by nuclear magnetic resonance spectroscopy. Biochem J 293: 657–659
6. Delfosse P, Reddy AS, Legrève A, Devi PS, Thirumala Devi K, Maraite H, Reddy DVR (1999) Indian peanut clump virus (IPCV) infection on wheat and barley: symptoms, yield loss and transmission through seed. Plant Pathol 48: 273–282
7. Dolja VV, Boyko V, Agranovsky AA, Koonin EV (1991) Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: two families with distinct patterns of sequence and probably structure conservation. Virology 184: 79–86
8. Fritsch C, Dollet M (2000) Genus: *Pecluvirus*. In: van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens E, Estes M, Lemon S, Maniloff J, Mayo MSA, McGeoch D, Pringle CR, Wicker RB (eds) Virus taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, pp 913–917
9. Goodwin JB, Dreher TW (1998) Transfer RNA mimicry in a new group of positive-strand RNA plant viruses, the furoviruses: differential aminoacylation between the RNA components of one genome. Virology 246: 170–178
10. Goulden MG, Davies JW, Wood KR, Lomonossoff GP (1992) Structure of tobraviral particles: a model suggested from sequence conservation in tobraviral and tobamoviral coat proteins. J Mol Biol 227: 1–8
11. Herzog E, Guilley H, Manohar SK, Dollet M, Richards K, Fritsch C, Jonard G (1994) Complete nucleotide sequence of peanut clump virus RNA 1 and relationships with other fungus-transmitted rod-shaped viruses. J Gen Virol 75: 3147–3155
12. Herzog E, Guilley H, Fritsch C (1995) Translation of the second gene of peanut clump virus RNA-2 occurs by leaky scanning *in vitro*. Virology 208: 215–225
13. Herzog E, Hemmer O, Hauser S, Meyer G, Bouzoubaa S, Fritsch C (1998) Identification of genes involved in replication and movement of peanut clump virus. Virology 248: 312–322
14. Huguenot C, Givord L, Sommermeyer G, Van Regenmortel MHV (1989) Differentiation of peanut clump virus serotypes by monoclonal antibodies. Res Virol 140: 87–102
15. Konaté G, Barro N (1993) Dissemination and detection of peanut clump virus in groundnut seed. Ann Appl Biol 123: 623–627

16. Legorburu FJ, Robinson DJ, Torrance L (1996) Features on the surface of the tobacco rattle tobravirus particle that are antigenic and sensitive to proteolytic digestion. J Gen Virol 77: 855–859
17. Manohar SK, Guilley H, Dollet M, Richards K, Jonard G (1993) Nucleotide sequence and genetic organization of peanut clump virus RNA 2 and partial characterization of deleted forms. Virology 195: 33–41
18. Manohar SK, Dollet M, Dubern J, Gargani D (1995) Studies on variability of peanut clump virus: symptomatology and serology. J Phytopathol 143: 233–238
19. Mayo MA, Naidu RA (1997) Peanut clump disease: an overview. In: Reddy DVR, Delfosse P, Lenné JM, Subrahmanyam P (eds) Groundnut virus diseases in Africa: Summary and recommendations of the Sixth Meeting of the International Working Group. ICRISAT, Andhra Pradesh, India and Belgian Administration for Development Cooperation, 1000 Brussels, Belgium, pp 22–23
20. Mayo MA, Pringle CR (1997) Virus taxonomy – 1997. J Gen Virol 79: 649–657
21. Mayo MA, Robertson WM, Legorburu FJ, Brierley KM (1994) Molecular approaches to an understanding of the transmission of plant viruses by nematodes. In: Lamberti F, De Giorgi C, Bird DM (eds) Advances in molecular nematology. Plenum Press, New York, pp 277–293
22. MacFarlane SA (1999) Molecular biology of the tobraviruses. J Gen Virol 80: 2799–2807
23. Miller JS, Wesley SV, Naidu RA, Reddy DVR, Mayo MA (1996) The nucleotide sequence of RNA-1 of Indian peanut clump furovirus. Arch Virol 141: 2301–2312
24. Naidu RA, Kimmins FM, Delfosse P, Alegbejo MD, Waliyar F (1997) Groundnut rosette and peanut clump diseases in West Africa. In: Reddy DVR, Delfosse P, Lenné JM, Subrahmanyam P (eds) Groundnut virus diseases in Africa: Summary and recommendations of the Sixth Meeting of the International Working Group. ICRISAT, Andhra Pradesh, India and Belgian Administration for Development Cooperation, 1000 Brussels, Belgium, pp 49–51
25. Naidu RA, Miller JS, Mayo MA, Wesley SV, Reddy AS (2000) The nucleotide sequence of *Indian peanut clump virus* RNA 2: sequence comparisons among pecluviruses. Arch Virol 145: 1857–1866
26. Nolt BL, Rajeshwari R, Reddy DVR, Bharatan N, Manohar SK (1998) Indian peanut clump virus isolates: host range, symptomatology, serological relationships, and some physical properties. Phytopathology 78: 310–313
27. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265: 218–225
28. Reddy AS, Hobbs HA, Delfosse P, Murthy AK, Reddy DVR (1998) Seed transmission of Indian peanut clump virus (IPCV) in peanut and millets. Plant Dis 82: 343–346
29. Reddy DVR, Delfosse P, Mayo MA (1999) Pecluviruses. In: Webster R, Granoff A (eds) Encyclopedia of virology. Academic Press, San Diego, pp 1196–1200
30. Reddy DVR, Nolt BL, Hobbs HA, Reddy AS, Rajeshwari R, Rao AS, Reddy DDR, McDonald D (1988) Clump virus in India: isolates, host range, transmission and management. In: Cooper JI, Asher MJC (eds) Developments in applied biology 2, Viruses with fungal vectors. AAB, Wellesbourne, Warwick, UK, pp 239–246
31. Sawyer SA (1998) GENECONV: A computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis, available at: http://www.math.wustl.edu/∼sawyer/mbprogs
32. Sharma KK, Anjaiah V (2000) An efficient method for the production of transgenic plants of peanut (*Arachis hypogaea* L.) through *Agrobacterium tumefaciens*-mediated genetic transformation. Plant Sci 159: 7–19

33. Solovyev AG, Savenkov EI, Agranovsky AA, Morozov SYU (1996) Comparisons of the genomic *cis*-elements and coding regions in RNAβ components of the hordeiviruses barley stripe mosaic virus, lychnis ringspot virus, and poa semilatent virus. Virology 219: 9–18
34. Thouvenel JC, Dollet M, Fauquet C (1976) Some properties of peanut clump, a newly discovered virus. Ann Appl Biol 84: 311–320
35. Thouvenel J-C, Fauquet C, Fargette D, Fishpool LDC (1988) Peanut clump virus in West Africa. In: Cooper JI, Asher MJC (eds) Developments in applied biology 2, Viruses with fungal vectors. AAB, Wellesbourne, Warwick, UK, pp 247–254
36. Van Regenmortel MHV, Bishop DHL, Fauquet CM, Mayo MA, Maniloff J, Calisher CH (1997) Guidelines to the demarcation of virus species. Arch Virol 142: 1505–1518
37. Wesley SV, Mayo MA, Jolly CA, Naidu RA, Reddy DVR, Jana MK, Parnaik VK (1994) The coat protein of Indian peanut clump virus: relationships with other furoviruses and with barley stripe mosaic virus. Arch Virol 134: 271–278

Author's address: R. A. Naidu, Department of Plant Pathology, University of Georgia, Athens, GA 30602, USA; e-mail: naidu@arches.uga.edu; Fax: 706-542-1262.