Nucleotide sequence analyses of peanut stunt cucumovirus RNAs 1 and 2

Akira Karasawa,1* Kazuhiro Nakahō,1† Tetsuji Kakutani,2 Yuzo Minobe2 and Yoshio Ehara1

1Faculty of Agriculture, Tohoku University, Sendai 981 and 2Department of Molecular Biology, National Institute of Agrobiological Resources, Tsukuba 305, Japan

The nucleotide sequences of the RNAs 1 and 2 of the peanut stunt virus strain J (PSV-J) were determined and compared with those of the cucumber mosaic virus strain Y (CMV-Y, subgroup I), strain Q (CMV-Q, subgroup II) and the tomato aspermy virus strain V (TAV-V) at both the nucleotide and protein levels. RNA 1 of PSV-J consists of 3355 nucleotides (nt) and has one large open reading frame (ORF) which can encode the putative la protein of Mr 112025. PSV-J RNA 1 and the la protein are 65 to 73% identical to those of CMV-Y and -Q, and 65 to 69% to those of TAV-V. RNA 2 of PSV-J contains 2946 nt and also has one large ORF which can encode the putative 2a protein of Mr 93575. For RNA 2 and the 2a protein, identities between PSV-J and two strains of CMV are calculated to be 53 to 61%. When compared with TAV-V, the same degree of similarity as seen with CMVs is observed. The la protein has the consensus sequences found in some helicases and methyltransferases and the 2a protein includes a sequence which exists in several RNA-dependent RNA polymerases.

Introduction

The cucumoviruses are a group of viruses having tripartite, positive-sense ssRNA genomes (Peden & Symons, 1973). Monocistronic RNAs 1 and 2 encode the 1a and 2a proteins involved in virus replication (Nitta et al., 1988b; Hayes & Buck, 1990). RNA 3 has two cistrons, for the 3a protein and the coat protein; the latter is expressed from subgenomic RNA 4 (Schwinghammer & Symons, 1977). Cucumber mosaic virus (CMV), the type member of the group, has many strains which differ in pathogenicity and several biological properties. They have been divided into two subgroups (Piazzola et al., 1979) based on nucleic acid hybridization studies (Piazzola et al., 1979; Gonda & Symons, 1978), serology (Devergne & Cardin, 1973) and peptide mapping of the coat protein (Edwards & Gonsalves, 1983). The entire nucleotide sequences of several strains of CMV have been determined (for subgroup I, strain Y (Nitta et al., 1988a; Kataoka et al., 1990a, b); for subgroup II, strain Q (Davies & Symons, 1988; Rezaian et al., 1984, 1985) and the subgrouping has been confirmed at the nucleotide level. Recently, sequences of parts of the genomes of tomato aspermy virus (TAV) strains C and V have been reported (O'Reilly et al., 1991; Moriones et al., 1991; Bernal et al., 1991). Furthermore we have determined the complete sequence of RNA 3 of peanut stunt virus strain J (PSV-J) (Karasawa et al., 1991), another member of the cucumovirus group. In this paper, we determined the nucleotide sequences of RNAs 1 and 2 of PSV-J and compared them with those of CMV-Y, -Q and TAV-V. As a result, the entire genome and other information on PSV-J has been studied and the relationships of PSV to other members of the group have been established at the nucleotide level.

Methods

cDNA cloning and sequencing. Double-stranded cDNA of PSV-J RNAs was synthesized and inserted into the EcoRV site of the pBluescript II SK(+) vector (Stratagene) as described (Karasawa et al., 1991). Clones specific to RNAs 1 and 2 were selected by dot blot hybridization using RNA probes purified from low melting point agarose and labelled with [γ-32P]ATP by T4 polynucleotide kinase. Clones pPJ11 and pPJ21 specific to RNAs 1 and 2, respectively, were obtained. They were almost full-length clones of RNAs 1 or 2 except for 40 and 20 nucleotides of their 5' ends, respectively. Both pPJ11 and pPJ21 were digested with appropriate restriction endonucleases and deletions were made by using exonuclease III and mungbean nuclease. The resultant ordered sets of deletion clones were used as templates for sequencing by the dideoxynucleotide chain termination method (Sanger et al., 1977). For clone pPJ11, Taq DNA polymerase (Perkin-
shown in Fig. 2 with the deduced amino acid sequence of the predicted translation product, the 2a protein. RNA 2 contains 2946 nt. Following the 5' non-coding region of 90 nt, there exists one large ORF of 2505 nt, giving a predicted Mr for the 2a protein of 93 575. The length of the 3' non-coding region is 79 nt, there exists one large ORF of 2505 nt, giving a predicted Mr of 112025.

Results and Discussion

Nucleotide sequences of RNAs 1 and 2 and comparisons with other cucumoviruses

Fig. 1 shows the complete nucleotide sequence and the deduced amino acid sequence of the predicted translation product of PSV-J RNA 1. RNA 1 is 3355 nucleotides (nt) long. Following the 5' non-coding region of 90 nt, there exists one large open reading frame (ORF) of 3018 nt. The length of the 3' non-coding region is 247 nt. The predicted la protein translated from RNA 1 is a polypeptide of 1006 amino acids and has an $M_r$ of 112025.

The complete nucleotide sequence of PSV-J RNA 2 is shown in Fig. 2 with the deduced amino acid sequence of the predicted translation product, the 2a protein. RNA 2 contains 2946 nt. Following the 5' non-coding region of 79 nt, there exists one large ORF of 2505 nt, giving a predicted $M_r$ for the 2a protein of 93 575. The length of the 3' non-coding region is 362 nt.

We compared the nucleotide and the predicted amino acid sequences of RNAs 1 and 2 of PSV-J with those of two strains of CMV, CMV-Y (subgroup I; Kataoka et al., 1990a, b) and CMV-Q (subgroup II; Rezaian et al., 1984, 1985), and with TAV-V (Moriones et al., 1991; Bernal et al., 1991) (Table 1). RNA 1 of PSV-J shows significant similarity to other cucumoviruses at both the nucleotide and amino acid levels. The resemblance is closest between PSV-J and CMV-Q, and lowest between PSV-J and TAV-V. The values for the proteins are slightly higher than those for the RNAs. The regions showing homology are distributed over the entire molecule in every case (data not shown). On the other hand, homologies between RNAs 2 of PSV-J and other cucumoviruses are very few. At the nucleotide level, the identity values are approximately 60%, whereas the amino acid sequences differ more. When a dot matrix comparison program was used to show homologous regions between RNAs 2 of PSV-J and CMV-Y, related sequences were observed in the region from 1000 to 2500 nt, but more distal regions 5' and 3' of this sequence showed lower identity to CMV-Y RNA 2. The 2a proteins of PSV-J and CMV-Y were compared by using the same program. Although approximately 250 amino acid residues from the N terminus are somewhat different from the CMV-Y 2a protein, the remainder of the molecule is similar to the corresponding region of the CMV-Y 2a protein except for the region near the C terminus. Similar results are obtained in all cases between PSV-J and other cucumoviruses (data not shown).

By determining the entire RNA organization of PSV-J, the genetic position of PSV-J in the cucumovirus group is established (Table 1). Based on RNA 3 including the 3a protein and the coat protein, PSV-J resembles TAV-C more closely than the two CMVs. A striking difference is seen when the coat protein of PSV-J is compared with those of TAV-C and CMV-Y (Karasawa et al., 1991, Table 1). The similarity between the coat proteins of PSV-J and TAV-C is 60-4%, whereas the values between PSV-J and CMV-Y or TAV-C and CMV-Fny (Bernal et al., 1991) are lower than 50%. However, RNAs 3 and the 3a proteins are more similar among the three viruses. On the other hand, the three viruses revealed approximately equal relationships at the levels of their RNAs 2 and the 2a proteins. As mentioned above, the degree of conservation of RNA 2 among cucumoviruses is low. These lower values have been noticed also between CMV-Fny and -Q and between TAV-V and two CMVs (Bernal et al., 1991). It is unclear whether or not this lower conservation correlates with any properties of the virus. Based on the sequences of RNAs 1 of cucumoviruses, genetic distances between any two viruses are equivalent. Whereas the percentage identity is much higher (approx. 85%) between two CMV subgroups, the similarity values between any two cucumoviruses is about 70% (Bernal et al., 1991; Table 1). Although the degrees of conservation of sequences are different for each RNA, the values between any two viruses are almost equal and that within subgroups is higher than those between viruses of different subgroups, indicating definitely that these three cucumoviruses share the same origin and that the subgrouping in CMV occurred after the establishment of CMV. The reason why degrees of preservation are different for each RNA remains unclear.

Predicted function(s) of the 1a and 2a proteins

Brome mosaic virus (BMV) and alfalfa mosaic virus (AlMV), which have a tripartite genome like CMV and PSV, have been assigned to a superfamily also comprising tobacco mosaic virus (TMV) and the animal Sindbis virus (Goldbach, 1986). In this superfamily, the 1a proteins of BMV and AlMV, the 126K protein of TMV and the non-structural (NS) proteins 1 and 2 of Sindbis virus are similar in sequence, and the 2a proteins of BMV and AlMV, the readthrough region of the TMV 183K...
sequence and the underlined sequence are the conserved helicase domain and the methyltransferase domain, respectively (see text for details).

**Fig. 1. Complete nucIeotide sequence of PSV-1 RNA 1 and deduced amino acid sequence of the predicted L protein.** The boxed and the underlined sequence are the conserved helicase domain and the methyltransferase domain, respectively (see text for details).
underlined is the conserved polymerase core domain (see text for details).

**Fig. 2** Complete nucleotide sequence of PSV-J RNA 2 and deduced amino acid sequence of the predicted 2a protein. The sequence underlined is the conserved polymerase core domain (see text for details).
protein and NS protein 4 of Sindbis virus also have conserved sequences. By analogy to other viruses and cellular proteins, it is presumed that the 1a and 2a proteins act as a helicase and an RNA-dependent RNA polymerase, respectively (Gorbalenya & Koonin, 1989; Habili & Symons, 1989). The 1a proteins of A1MV, BMV, CMV and TAV also have the conserved domain in their N-terminal portions which is maintained in the TMV 126K protein and NS protein 1 of Sindbis virus (Goldbach, 1986; Bernal et al., 1991) and is expected to act as a methyltransferase for viral RNA capping (Mi et al., 1989; Scheidel et al., 1989). Furthermore, evidence that the 1a and 2a proteins are involved in the replication complex with a 50K host protein has been proposed in the CMV–tobacco system (Hayes & Buck, 1990).

The 1a and 2a proteins of PSV-J, as mentioned above, show homology to those of CMV. Although the 1a proteins are conserved between two viruses to a greater degree than are the 2a proteins, the identity is much higher (73.7%) when the sequences from amino acid 700 to the C terminus of the 1a protein are compared; this region is where the conserved helicase domain is located (Fig. 1, boxed region). In addition, the predicted methyltransferase domain is present at amino acids 81 to 174 (Fig. 1, underlined region). When these conserved domains are aligned with the 1a proteins of other tripartite genome plant viruses of the Sindbis superfamily, the amino acid sequences are very similar (data not shown). The underlined amino acid sequences in Fig. 2 show the conserved core domain of the polymerase. This sequence is compared with corresponding regions of other 2a proteins as well as of the 1a protein and are proved to be conserved almost completely among several tripartite plant viruses including CMV (data not shown). Thus, the 1a and 2a proteins may be involved in the replication of PSV as seen in CMV.

**Non-coding regions (5’ and 3’) of RNAs**

The determination of the entire genomic organization makes it possible to compare the 5’ and 3’ non-coding regions of each RNA. Fig. 3 shows the alignment of the 5’ non-coding regions. The sequences of RNAs 1 and 2 are very similar, whereas RNA 3 is somewhat different from RNAs 1 and 2. This sequence diversity in the 5’ non-coding region is exhibited in CMV (for CMV-Y; Kataoka et al., 1990b). For RNAs 1 and 2, the conserved sequence (GGUUCGANUCC) in the B box of the internal control regions of tRNA promoters (ICR2), which acts as a promoter for plus-strand synthesis in viral replication (Pogue et al., 1990), is found as reported for BMV and CMV (Marsh et al., 1989), TAV (Moriones et al., 1991) and PSV (Karasawa et al., 1991). For CMV, this sequence has been proposed to interact with the antisense sequence of CMV satellite RNA (satRNA), thereby regulating viral and satRNA replication (Rezaian et al., 1985). Molecular analysis of the interaction of PSV (isolate ER) genomic RNAs with satRNAs has been reported recently (Naidu et al., 1991). In this report, some satRNAs diminished the accumulation of RNAs 1...
and 2 in the leaves, thus attenuating the symptoms. We tried to find homologies between the sequences of the PSV-J genomic RNAs and satRNAs but failed to detect sequences common to RNAs 1 and 2 likely to interact with satRNA (data not shown). Although we do not know what similarities exist at the nucleotide level between strain J and isolate ER, the helper strain described in the report of Naidu et al. (1991), it will be very interesting to determine whether isolate ER possesses a sequence that can interact with its satRNA.

On the other hand, the 3’ non-coding regions of each RNA are very similar. Particularly, in the 44 nt from the 3’ end, each RNA shows a perfect match. Although the regions more 5’ are somewhat different, only six bases for RNA 2 and five for RNA 3 are unmatched to RNA 1 in the 150 nt region from the 3’ end. Identity values are 46-1% for CMV-Y and 45-5% for TAV-V in the corresponding regions. This region of CMV can form a tRNA-like conformation (Joshi et al., 1983) and can be aminoaacylated by tyrosine (Kohl & Hall, 1974). For BMV, it has been assumed to be involved in the initiation of minus-sense RNA replication (Dreher & Hall, 1988). We tried to adapt the secondary structure model for RNA 1, but the result obtained is somewhat different, particularly in requiring a new loop at its 5’ region (Fig. 4). Whether or not this result reflects the correct conformation of PSV-J RNAs, details of the function(s) of this region remain to be tested.

We thank Drs S. Kashiwazaki and N. Suzuki for their kind suggestions for experiments and Ms Hiromi Higo for synthetic oligonucleotides. We also thank Dr Ilkka Havukkala for critical reading of the manuscript and Drs T. Hashiba and H. Takahashi for their support in experiments.

References


Nucleotide sequences of PSV RNAs 1 and 2


(Received 9 September 1991; Accepted 25 November 1991)