Evolutionary relationships in the cucumoviruses: nucleotide sequence of tomato aspermy virus RNA 1

Juan J. Bernal, Enrique Moriones and Fernando García-Arenal*

Departmento de Patologia Vegetal, E.T.S.I. Agrónomos, Ciudad Universitaria, 28040 Madrid, Spain

RNA 1 of the V strain of tomato aspermy virus (TAV) consists of 3410 nucleotides and contains one open reading frame (ORF) of 2982 nucleotides, resembling RNA 1 of cucumber mosaic virus (CMV) strains Q and Fny (68% and 66% identical, respectively) and of brome mosaic virus (BMV) (41% identical). In comparisons between amino acid sequences, three conserved regions (N-terminal, C-terminal and central) between TAV and each CMV were found. The N- and C-terminal regions were also conserved with BMV, and contained, respectively, consensus motifs for methyltransferases and for nucleic acid helicases. The 5' and 3' non-coding sequences were highly similar to those of TAV RNA 2. When the sequences for the genomic RNAs of the V and C strains of TAV, and of their encoded products, are compared with those reported for CMV strains representing either subgroup I (Fny-CMV) or subgroup II (Q-CMV) of CMV, it was found that the different virus-encoded proteins are conserved differently between these three viruses. Also, the divergence between TAV and both CMV subgroups has proceeded at different rates for the different ORFs. On the whole, the divergence between TAV and CMV is of the same order as that found between CMV subgroups I and II, which suggests that TAV, Q-CMV and Fny-CMV could be considered as representing three equivalent subgroups of a taxonomic entity.

Introduction

Cucumber mosaic virus (CMV), the type member of the cucumovirus plant virus group, has been characterized extensively and the nucleotide sequence of the genomic RNAs of several strains has been completely (Davies & Symons, 1988; Owen et al., 1990; Rezaian et al., 1984, 1985; Rizzo & Palukaitis, 1988, 1989) or partially (Hayakawa et al., 1989a, b; Noel & Ben Tahar, 1989; Quemada et al., 1989) determined. A less extensively studied cucumovirus is tomato aspermy virus (TAV), which can be differentiated from CMV (Habili & Francki, 1974a, b; Marrou et al., 1975; Devergne & Cardin, 1975) although both viruses are closely related in several properties (Habili & Francki, 1974a; Mossop & Francki, 1979; Jaegle et al., 1990). Recently, the nucleotide sequence for the RNA 3 of a British TAV strain (C-TAV), and for the RNA 2 of an Australian strain (V-TAV) have been reported (O'Reilly et al., 1991; Moriones et al., 1991). Here we report the nucleotide sequence for the RNA 1 of V-TAV and, on the basis of the complete nucleotide sequence for the genomic RNAs of TAV, discuss the relationships between TAV and CMV.

Methods

Virus. The V strain of TAV, and the purification of V-TAV virions and RNA, have been described (Habili & Francki, 1974a).

cDNA synthesis and cloning. Preparations of dsDNA, cloning into pUC18, preparation of plasmid DNA, and subcloning in M13mp18 and M13mp19 were as described (Moriones et al., 1991). To obtain clones representing all of RNA 1, two different primers were used, complementary to the 5' extreme 18 nucleotides (nt) and to nt 1335 to 1352. dsDNA to the 5' region was prepared by the polymerase chain reaction (Saiki et al., 1988) using primers complementary to nt 404 to 418, and identical to 1 to 10.

Sequence analysis. From the clones in M13mp18 and M13mp19, sets of overlapping clones containing progressive unidirectional deletions of inserted DNA were obtained using the exonuclease III method of the Erase-a-Base system (Promega). Single-stranded DNA from these clones was sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977) using the Sequenase kit (United States Biochemical). Analysis and comparisons of the TAV RNA 1 sequence, and of the predicted sequence for the encoded protein, were done with the Microgenie Sequence Analysis Program (Beckman).

The RNA sequence data reported in this paper have been submitted to the DDBJ, EMBL and GenBank nucleotide sequence databases and assigned the accession number D01101.

0001-0298 © 1991 SGM
Fig. 1. Nucleotide sequence of V-TAV RNA and encoded amino acid sequence for the long ORF. The initiation and stop codons are underlined.
Results and Discussion

Genetic organization

The complete nucleotide sequence of both the viral encapsidated plus strand, and of the minus strand, of the genomic RNA 1 of V-TAV was determined from a collection of overlapping clones in M13. V-TAV RNA 1 (Fig. 1) was 3410 nt long, with a base composition of 27.6% A, 27.9% U, 21.8% C and 22.7% G. A long open reading frame (ORF) was found in the plus strand, starting at the first AUG codon (positions 95 to 97) and extending to position 3076. Its putative translation product (Pla), 993 amino acid residues long, had a calculated Mr of 112074. The next longest ORF in the plus strand was 375 nt long. V-TAV RNA 1 has a size and a genetic organization similar to those of the RNA 1 of other tripartite plant viruses. Sequence similarity over the whole RNA molecule was 68% with Q-CMV RNA 1 (Rezaian et al., 1985), 66% with Fny-CMV RNA 1 (Rizzo & Palukaitis, 1989) and 41% with brome mosaic virus (BMV) RNA 1 (Ahlquist et al., 1984).

Non-coding regions

Experiments on 5' end labelling with [γ-32P]ATP show V-TAV RNAs to be capped at the 5' end (Moriones et al., 1991). The 5' leader sequence of V-TAV RNA 1 extended 94 nt up to the beginning of the Pla ORF. The sequence is similar in length to the leader sequences of V-TAV RNA 2, and of Q- and Fny-CMV RNAs 1 and 2. The 5' non-coding regions of V-TAV RNAs 1 and 2 are 78% similar, and are identical for the 5' extreme 49 nt. Thus, putative regulatory signals described for V-TAV RNA 2 (Moriones et al., 1991) are also present in RNA 1. The 3' non-coding region of V-TAV RNA 1 is 334 nt long, much shorter than that of V-TAV RNA 2 (500 nt), and about the size of that of C-TAV RNA 3 (306 nt). For the 3' extreme 318 nt, V-TAV RNAs 1 and 2 are 98% similar, and show 94% similarity with C-TAV RNA 3. Thus, structural and putative regulatory features at the 3' end of V-TAV RNA 2 (Moriones et al., 1991) are also found for V-TAV RNA 1. At odds with RNA 2, sequence similarities for the 3' non-coding regions, upstream to the 3' extreme 149 nt involved in the tRNA-like structure, are much higher between TAV and Q-CMV (71%) than between TAV and Fny-CMV (36%) or Q-CMV and Fny-CMV (37%).

TAV RNA 1 translation product

The expected size for the putative protein encoded by V-TAV RNA 1 corresponds to that of an in vitro translation product of V-TAV RNAs 1 and 2 (A. Fraile, unpublished results). The context for the initiation codon of this putative translation product is the same as for those of Q-CMV and Fny-CMV RNA 1, and differs from the optimal sequence context for plant mRNAs (Lütcke et al., 1987), although the putatively important G at position +4 is present. Codon usage in the Pla ORF deviated from random: except for Ser, codons with U or A at the third position were preferred, as for the P2a ORF (Moriones et al., 1991).

V-TAV Pla was the same size as that of Fny-CMV (Fny-CMV Pla), two amino acids longer than that of Q-CMV (Q-CMV Pla), and 32 amino acids longer than that of BMV (BMV Pla). Its sequence (Fig. 1) has a high similarity (73%) with both CMVs Pla, whereas similarity with BMV Pla (26%) was not significant. Matrix analysis of the amino acid sequences showed that similarity among the cucumoviral Plas is distributed in three highly similar blocks (N-terminal, central and C-terminal; I, III and V in Table 1) separated by two shorter regions of low similarity (II and IV in Table 1). Hydrophobicity patterns, and the distribution of basic and acidic amino acids, were very similar for V-TAV, Q-CMV and Fny-CMV Pla: basic amino acids predominate in the 150 C-terminal amino acids, acidic ones predominate between residues 500 and 600, and neither type predominates in the rest of the molecule. Regions I and V also showed similarity with equivalent regions of BMV Pla (Table 1). In the large region V, six conserved motifs of nucleic acid helicases (Hodgman, 1988; Gorbalenya et al., 1988, 1989) are found, including two motifs involved in nucleotide binding (Walker et al., 1982; Gorbalenya et al., 1988) (VDGVAGCGKTTAIK, nt 2228 to 2269 and RVLDEVVLLH, nt 2429 to 2461). In the highly conserved region I, a motif (Candresse et al., 1990) (nt 335 to 616) conserved among RNA viruses of the Sindbis virus-like supergroup (Goldbach &

Table 1. Similarities between V-TAV RNA 1-encoded protein and Q-CMV-, Fny-CMV- and BMV RNA 1-encoded proteins

<table>
<thead>
<tr>
<th>Region</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid positions for V-TAV</td>
<td>1-220</td>
<td>221-262</td>
<td>263-494</td>
<td>495-569</td>
<td>570-993</td>
</tr>
<tr>
<td>V-TAV</td>
<td>Q-CMV</td>
<td>1-220</td>
<td>221-263</td>
<td>264-497</td>
<td>498-568</td>
</tr>
<tr>
<td>Fny-CMV</td>
<td>1-220</td>
<td>221-262</td>
<td>263-495</td>
<td>496-569</td>
<td>570-993</td>
</tr>
<tr>
<td>BMV</td>
<td>1-214</td>
<td>215-257</td>
<td>258-476</td>
<td>477-554</td>
<td>555-961</td>
</tr>
<tr>
<td>Similarity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q-CMV</td>
<td>85</td>
<td>31</td>
<td>79</td>
<td>31</td>
<td>74</td>
</tr>
<tr>
<td>Fny-CMV</td>
<td>84</td>
<td>33</td>
<td>81</td>
<td>33</td>
<td>73</td>
</tr>
<tr>
<td>BMV</td>
<td>52</td>
<td>23</td>
<td>18</td>
<td>26</td>
<td>50</td>
</tr>
</tbody>
</table>
Wellink, 1988) is found. In Sindbis virus this motif has been associated with an RNA methyltransferase activity (Mi et al., 1989; Scheidel et al., 1989). In addition to the presence of the above conserved motifs, a role for CMV and BMV Pla in viral RNA replication is substantiated by the ability of BMV and CMV RNAs 1 and 2 to replicate in protoplasts (Kiberstis et al., 1981; Nitta et al., 1988), by the presence of the Pla in the RNA polymerase complexes, and by the inhibition of the activity of these complexes by antibodies to the Pla (Hayes & Buck, 1990; Horikoshi et al., 1988; Quadt et al., 1988). For BMV Pla both the N- and C-terminal conserved domains have been shown to be involved in the synthesis of all types of viral RNAs (Kroner et al., 1990). The conservation of Pla proteins may extend these roles to TAV Pla.

Relationships between CMV and TAV

TAV differs from CMV in its host range (TAV not infecting cucurbit; Marrou et al., 1975) and, mainly, serologically (Habili & Francki, 1974a; Devergne & Cardin, 1975). Other reported differences between TAV and CMV, e.g. the physicochemical characteristics of the viral particle, are not consistent for all strains (Habili & Francki, 1974a; Savithri et al., 1984). Although it has been determined for two Chrysanthemum isolates of different origins, the complete nucleotide sequence for the genomic RNAs of TAV is available with this report, and may be compared to those of CMV strains representing both subgroup I (CMV-I) (i.e. that of Fny-CMV; Owen et al., 1990; Rizzo & Palukaitis, 1988, 1989), and subgroup II (CMV-II) (i.e. that of Q-CMV; Davies & Symons, 1988; Rezaian et al., 1984, 1985), providing a new basis for establishing the relationships between TAV and CMV. Partial sequences from other CMV strains (Hayakawa et al., 1989a, b; Noel & Ben Tahar, 1989; Quemada et al., 1989) show intragroup dissimilarity for protein sequences not to exceed 6%, so that Fny-CMV and Q-CMV may be considered as faithful representatives of CMV-I and CMV-II.

When the complete RNA genome is considered, dissimilarities between TAV and CMV-I, and between TAV and CMV-II are not much higher than the dissimilarity between CMV-I and CMV-II (Table 2). When the amino acid sequences for all four virus-encoded proteins are compared, the overall percentage sequence dissimilarity between TAV and CMV-I or CMV-II is higher than between CMV-I and CMV-II, but still less than twofold higher (Table 2). The ratio between the dissimilarities between TAV and CMV-I or CMV-II and between CMV-I and CMV-II varies for the different ORFs: it is 1.7 to 1.6 for Pla and P2a and goes up to 2.0 to 2.1 for the 3a proteins, and rises to 3.1 to 3.2 for the coat proteins (CP) (Table 2). Thus, in addition to different degrees of conservation found for the different virus-encoded proteins, the divergence between TAV and CMV-I or CMV-II has proceeded at different rates for the different ORFs. This divergence is highest for the CPs, and, in accordance to the reported serological differences between TAV and CMV, these viruses would be considered different if sequence comparisons are limited to the CPs. If the whole genome, or the whole encoded sequences, are considered, the difference between TAV and CMV is not so clear, and if taxonomy were based on other encoded proteins, such as Pla, then TAV, CMV-I and CMV-II should be considered as three strains of the same virus or else as three different viruses. It would not be unreasonable to consider TAV, CMV-I and CMV-II as three equivalent subgroups of a viral taxonomic entity.

To explain the different divergence rates shown for the 3a protein and CPs, recombinational events in the evolution of TAV and CMV have been invoked (O'Reilly et al., 1991). The different evolution rates shown for the four virus-encoded proteins are explained more simply by functional constraints limiting genetic divergence to different degrees for each ORF. This last hypothesis would appear to be supported by the fact that protein Pla, for which in the tricornaviruses a number of functional domains distributed over large regions of the molecule have been identified (Kroner et al., 1990; Traynor & Ahlquist, 1990), is the more highly conserved one, whereas the CP which in isometric viruses may show wide sequence variation while maintaining a similar tertiary structure (Rossmann & Johnson, 1989), is the less conserved one.

Table 2. Percentage dissimilarities among cucumoviruses

<table>
<thead>
<tr>
<th>Genomic RNAs (1 + 2 + 3)*</th>
<th>TAV/CMV-I</th>
<th>TAV/CMV-II</th>
<th>CMV-I/CMV-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoded proteins*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>42.5</td>
<td>37.2</td>
<td>35.4</td>
</tr>
<tr>
<td>P1a</td>
<td>36.6</td>
<td>36.0</td>
<td>20.0</td>
</tr>
<tr>
<td>P2a</td>
<td>26.5</td>
<td>26.5</td>
<td>15.0</td>
</tr>
<tr>
<td>P3a</td>
<td>43.0</td>
<td>43.0</td>
<td>27.0</td>
</tr>
<tr>
<td>Coat protein</td>
<td>34.6</td>
<td>33.0</td>
<td>16.1</td>
</tr>
<tr>
<td>Theriological dissimilarity index†</td>
<td>6.7</td>
<td>6.5</td>
<td>1.2</td>
</tr>
</tbody>
</table>

† Data from Devergne & Cardin (1975).

This work was in part supported by grant PA 86-0353, Comisión Interministerial de Ciencia y Tecnología, Spain. J. J. Bernal and E. Moriones were the recipients of Formación de Personal Investigador fellowships, Ministerio de Educación y Ciencia, Spain.
References


(Received 2 April 1991; Accepted 14 May 1991)