Unusual sequence relationships between two isolates of citrus tristeza virus

Munir Mawassi, Elzbieta Mietkiewska, Rose Gofman, Guang Yang and Moshe Bar-Joseph

The S. Tolkowsky Laboratory, Department of Virology, Agricultural Research Organization, The Volcani Center, Bet Dagan 50250, Israel

The complete, 19226 nt sequence of the RNA genome from VT, a seedling yellows strain of citrus tristeza virus (CTV), was determined and found to have a genome organization identical with that of the previously determined CTV-T36 isolate, except that ORF 1 of CTV-VT was 70 nt shorter due to two widely separated 18 nt deletions. Sequence comparison of CTV-VT and CTV-T36 revealed approximately 89% identity throughout the ten 3' ORFs, but only 60–70% identity throughout ORF 1. The 5' nontranslated regions were only 60% identical whereas the 3' nontranslated regions were 97% identical. The transition between regions of similarity and deviation was gradual, suggesting that the sequence similarities and differences compared to CTV-T36 were unlikely to have arisen from a recent recombination event between a close T36 relative and a distantly related CTV isolate. This is the first attempt to compare in detail the variation between the genomes of two strains of a member of the closterovirus group. The observed deviation between the large RNA genomes of the two CTV strains is greater than that among different viruses of most other groups, raising the question of how to define the taxonomy of these viruses.

Introduction

Isolates of citrus tristeza virus (CTV), an aphid-transmitted closterovirus, cause some of the most important diseases of citrus. The symptoms on infected trees differ considerably depending on host sensitivity and the pathogenic nature of the virus isolate (Bar-Joseph & Lee, 1989). CTV virions consist of long, flexuous particles, 10–12 nm x 2000 nm, containing a single-stranded, positive-sense genomic RNA (gRNA) (Bar-Joseph et al., 1985). Recently, the complete sequence (19296 nt) of a decline isolate from Florida, CTV-T36, was found to contain 11 ORFs (Karasev et al., 1995; Pappu et al., 1994), potentially encoding at least 17 protein products including the replication-associated complex [the polymerase (RdRp), helicase (HEL) and methyltransferase (MT) domains], plus a homologue of the HSP70 proteins, a coat protein (CP), a divergent coat protein (dCP) and several other proteins of unknown function. Two papain-like proteases (P-PRO), the HEL and MT domains, and through a +1 translation frameshift the RdRp, are expressed as large polyproteins. The ten ORFs located in the 3' half of the CTV genome are thought to be expressed through a nested set of 3'-coterminal sub-genomic mRNAs (Hilf et al., 1995; Mawassi et al., 1995 a).

Earlier sequence characterization of the five 3' ORFs of CTV-VT suggested considerable similarity to CTV-T36 (Mawassi et al., 1995 a). However, the discovery and sequencing of defective RNAs (D-RNAs), which have most of the internal sequences deleted, revealed that their 3' sequences were CTV-T36-like, but the 5' sequences were much more distantly related, suggesting that the D-RNAs might be chimeras from different CTV isolates (Mawassi et al., 1995 b, c).

In this report, we present the complete sequence of CTV-VT and show that its 3' sequences are similar to CTV-T36, like the characterized D-RNAs, and that the 5' sequences are unusually dissimilar from CTV-T36, as were the CTV-VT D-RNAs. The transition between similarity and dissimilarity relative to CTV-T36 is gradual over several hundred nucleotides suggesting that this genome structure has not recently evolved from a recombination between two distantly related CTV isolates. These results demonstrate unusual sequence relationships between different isolates of CTV.

Methods

Propagation and purification of virus and RNA isolation. CTV-VT was originally isolated in 1970 from a declining sweet orange (Citrus sinensis) cv. Valencia tree grafted on the sour orange rootstock in
the Hibatt Zion area, Israel (Bar-Joseph & Loebenstein, 1973). The virus was maintained in Alemow (C. macrophylla) seedlings grown in a greenhouse with temperatures ranging between 25-30 °C. Virion RNA was obtained by phenol-chloroform extraction from purified particles (Bar-Joseph et al., 1985; Rosner et al., 1983) or directly from CTV particles trapped to polystyrene ELISA wells using an immuno-PCR protocol (Nolasco et al., 1993).

Double-stranded RNAs were isolated as previously reported (Dodds & Bar-Joseph, 1983; Dulieu & Bar-Joseph, 1989). The genomic-size, RF-dsRNA (ca. 20 kbp) was isolated by electroelution using Bio-Trap membranes (Schleicher and Schuell).

■ cDNA cloning and nucleotide sequencing. For cloning of cDNA, ssRNA was treated with DNase and 3' polyadenylated (Ashoulin et al., 1992). cDNA was synthesized using a chimeric deoxoyoligonucleotide primer consisting of a polylinker and oligo(dT) [5' GCCGGCGATTCCAGGG(T)15 3'] (P-DT) and AMV reverse transcriptase (RT) (Promega) following the manufacturer's instructions. The cDNA was poly(A) tailed using terminal transferase (Boehringer) (Ashoulin et al., 1992), followed by PCR amplification with P-DT primer. The PCR fragments were digested with BamHI (this site was included in the P-DT primer), eluted from agarose gel, and cloned into BamHI-digested pBluescript KS (Stratagene) (Maniatis et al., 1982). Sequencing of the cDNA clones was carried out on both strands using the T7 and the T3 primers and the USB sequencing kit according to the manufacturer's instructions. Sequences obtained from the primary clones (Fig. 1, lower panel) were used for the synthesis of CTV-VT-specific oligonucleotide primers. These specific primers were used for RT-PCR of internal regions on the VT genome, using ssRNA, purified virions or immunotrapped particles (Nolasco et al., 1993) or isolated full-genomic dsRNA. The larger than 0.6 kbp cDNA fragments were digested with HaeIII and AbiI, subcloned into the EcoRV site of pBluescript, and sequenced in both directions. Additionally, a large set of specific CTV-VT primers was used for sequencing by the step-by-step walking strategy. The sequence of the 5' end was determined using the primer extension procedure described by Karasev et al. (1995) and the VT-specific primer 5' TTTGCGGCGATTTACAC 3', complementary to nt 58-75. Sequences were analysed with the UWGCG program (Devereux et al., 1984).

Results and Discussion

Fig. 1 shows two series of cDNA clones obtained from polyadenylated viral RNA templates and from RT-PCR of ss- and dsRNA molecules using CTV-VT-specific primers (lower and upper panels respectively). The complete 19,226 nucleotide sequence of CTV-VT was assembled from these panels and deposited in GenBank (accession no. U56902). The genome of CTV-VT was 70 nt shorter than CTV-T36, mainly because of the appearance of two 18 nt gaps, in two widely separated regions (2322-2339 and 10809-10826). Other shorter gaps of 1-3 nt were scattered in several parts of the VT genome. The presence of these gaps was confirmed by at least two separate cloning and sequencing experiments. Five of the 1 or 2 nt gaps were within the putative ORF 1a, and are expected to cause shifts along the subsequent 15-25 amino acids.

Using the BESTFIT (Smith & Waterman, 1981) or GAP (Needleman & Wunsch, 1970) programs for nucleotide sequence alignment, the CTV-VT genome had an average identity of 72.7% with the Florida CTV-T36 strain. Sequence conservation differed considerably between the 5' and 3' halves (1-9100; 9101-19226) with 70.7% and 87.4% identity respectively (Fig. 2). There is a gradual increase in sequence identity throughout the centre of the viral genomes and specific regions with closer or more distant sequence identities scattered throughout the genomes (Fig. 2). Regardless of this major difference the two strains have a similar genomic organization (Fig. 2b).

The asymmetrical genomic conservation was particularly noticeable when comparing the 5' and 3' untranslated regions (UTR) (Table 1). The 105 nt 5' UTR of CTV-VT was only 60% identical to the 107 nt 5' UTR of CTV-T36. The 272 nt CTV-VT 3' UTR was 97% identical to the CTV-T36 3' UTR. ORF 1a of CTV-VT (nt 106-9427) encodes a putative polyprotein of 346.7 kDa (Table 1). Alignments of amino acid (aa) sequences of ORF 1a of CTV-VT (3107 aa) and CTV-T36 (3124 aa) show 79.6% and 68% similarity and identity, respectively. The identities varied considerably in different parts of these proteins (Fig. 3). Amino acids 1-346 had only 50.6% identity, whereas two regions located between aa 347-599 and 770-964 had identities of 77% and 71.8%. These regions contain two P-PRO domains. The putative catalytic residues, Cys-399, 883 (403, 896) and His-400, 943 (464, 956) of CTV-VT were nearly identical to those of CTV-T36 (amino acid positions in parentheses). The putative cleavage sites were located between Gly-480-Gly-481 and Gly-963-Gly-964, both of which are preceded by a bulky hydrophobic amino acid, Val and Met, respectively. Accordingly the predicted cleavage products of the VT ORF 1a are 480 aa (53.4 kDa), 483 aa (54 kDa) and 2144 aa (239.1 kDa). Interestingly, in a time-course in vitro translation of CTV-VT RNA from full-length particles a 54 kDa product, as expected from proteolytic processing, was apparent only after 20 min incubation (Mawassi et al., 1995a). The domains of the P-PRO are interrupted by a less conserved region located between aa 600-769 with only 35.7% identity. The computer analyses of this region failed to reveal any significant relationships with sequences in the available databases.

The putative MT domains of CTV-VT (aa 983-1462) showed high identity with T36 (88.3%) and lower with beet yellows virus (BYV) (47.5%) (Agranovsky et al., 1994) and lettuce infectious yellows closterovirus (LIYV) (22.8%) (Klaassen et al., 1995). The region downstream of the MT domain (aa 1463-2685) was 61.4% identical to CTV-T36 and 23.5% to BYV. Computer analyses failed to detect any functional motifs for this part.

The amino acid sequence between aa 2686 and the C-terminus (3107) showed 81.8% identity with CTV-T36 and 46.9% with BYV. All seven typical motifs of type I helicases were conserved in the VT strain. The HEL I domain motif contains the ATP/GTP-binding site (P-loop) sequence ARPGGGKT (aa 2697-2704) (Hodgman, 1988; Linder et al., 1989). This Gly-rich region might form a flexible loop between β- and z-helix strands which would be expected to interact.
Fig. 1. Schematic representation of cDNA clones of CTV strain VT. Clones shown in the upper panel were obtained by RT–PCR using VT-specific primers; arrowheads represent the P-dT primer (see text) used for cDNA synthesis and amplification of polyadenylated ssRNA or dsRNA templates. The cDNAs of the primary clones (lower panel) were derived from polyadenylated viral RNA templates followed by cDNA synthesis, polyadenylation with terminal transferase and RT–PCR with primer P-dT.

Fig. 2. (a) Nucleotide sequence identity between CiV strains Vi- and T36. The percentage identities were averaged over a 100 nt window obtained by GAP (Needleman & Wunsch, 1970) analysis. (b) Schematic map of the 12 ORFs from CTV-VT.

with the phosphate groups of the nucleotide (Saraste et al., 1990).

ORF 1b encodes a 477 aa (54.4 kDa) domain which would result in a 398 kDa polyprotein via a +1 ribosomal frameshift. Unlike BYV it does not contain the UAG terminator codon at the frameshift point and the GGGUUU sequence and the stem–loop structure. ORF 1b contains a 75 nt overlap (123 nt in T36) and a gap of six amino acids (aa 425–430). ORF 1b was the most conserved domain among the replication-associated proteins of the two CTV strains (90% identity) and contains all
Table 1. Genome analysis of the VT strain of citrus tristeza virus and its comparison with the corresponding regions of strain T36

<table>
<thead>
<tr>
<th>Region</th>
<th>Product</th>
<th>Start</th>
<th>Stop</th>
<th>Mol. mass (Da)</th>
<th>Length (amino acids)</th>
<th>RNA bases (%) Identity</th>
<th>% Similarity</th>
<th>% Identity</th>
<th>Amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' UTR</td>
<td>-</td>
<td>1</td>
<td>105</td>
<td>346706</td>
<td>3107</td>
<td>60.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ORF 1a pol</td>
<td>106</td>
<td>9427</td>
<td>-</td>
<td>54359</td>
<td>477</td>
<td>71.3</td>
<td>79.6</td>
<td>68.0</td>
<td>-</td>
</tr>
<tr>
<td>ORF 1b RdRp</td>
<td>9353</td>
<td>10784</td>
<td>9353</td>
<td>10784</td>
<td>477</td>
<td>79.3</td>
<td>95.6</td>
<td>90.0</td>
<td></td>
</tr>
<tr>
<td>ORF 2 p33</td>
<td>10835</td>
<td>11744</td>
<td>33969</td>
<td>303</td>
<td>84.8</td>
<td>92.4</td>
<td>84.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 3 p6</td>
<td>11816</td>
<td>11969</td>
<td>5593</td>
<td>51</td>
<td>88.9</td>
<td>96.1</td>
<td>92.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 4 p65</td>
<td>11977</td>
<td>13759</td>
<td>65211</td>
<td>594</td>
<td>87.6</td>
<td>94.1</td>
<td>90.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 5 p61</td>
<td>13685</td>
<td>15293</td>
<td>61305</td>
<td>536</td>
<td>87.3</td>
<td>94.0</td>
<td>90.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 6 p27</td>
<td>15270</td>
<td>15990</td>
<td>27362</td>
<td>240</td>
<td>88.1</td>
<td>97.1</td>
<td>93.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 7 p25 (CP)</td>
<td>16084</td>
<td>16753</td>
<td>24987</td>
<td>223</td>
<td>92.8</td>
<td>97.3</td>
<td>94.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 8 p18</td>
<td>16721</td>
<td>17222</td>
<td>18268</td>
<td>167</td>
<td>93.4</td>
<td>95.2</td>
<td>92.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 9 p13</td>
<td>17260</td>
<td>17617</td>
<td>13681</td>
<td>119</td>
<td>91.9</td>
<td>95.8</td>
<td>94.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 10 p20</td>
<td>17695</td>
<td>18241</td>
<td>20549</td>
<td>182</td>
<td>90.5</td>
<td>97.3</td>
<td>95.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF 11 p23.5</td>
<td>18325</td>
<td>18952</td>
<td>23798</td>
<td>209</td>
<td>90.3</td>
<td>93.8</td>
<td>87.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3' UTR</td>
<td>-</td>
<td>18953</td>
<td>19226</td>
<td>-</td>
<td>-</td>
<td>97.4</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. (a) Comparison of the predicted amino acid sequences of VT and T36 ORF 1a and ORF 1b. The percentage identities were averaged over a 50 aa window obtained by GAP (Needleman & Wunsch, 1970) analysis. (b) Predicted positions of the functional replication motifs. Note maximal identities in amino acid sequences of functional replication motifs.

The remarkable feature from the comparison of sequence similarities of CTV-VT and CTV-T36 is the gradual transition between the 3' regions, which are appropriately similar for two strains of a virus, and the 5' regions, which are sufficiently different to characterize these two virus isolates as different viruses when considering most other plant virus groups. One explanation for this unusual sequence similarity between two isolates of a virus would be that recombination had occurred recently between two different CTV strains, one of which was closely related to CTV-T36 and the other not.
Recombination of CTV resulting in different sizes of D-RNAs with different junction sites has been demonstrated both in CTV-VT and in other CTV isolates (Mawassi et al., 1995 b, c). Recombination between different CTV strains might have led to the construction of a full-genomic chimeric mixture of two CTV strains.

The second possibility is that the CTV-VT sequence reported is an artifact of cloning from a mixed population of viruses. This is unlikely for several reasons. The genomic libraries were constructed by two different cDNA synthesis strategies and represent information from a variety of templates (genomic and subgenomic dsRNA fragments). All together we sequenced over 200 overlapping sequences covering at least twice each part of the CTV-VT genome. None of these sequences deviated from the consensus VT sequence to the extent expected from mixed infections.

The presence of identical 3' and 5' ends in all of the previously characterized CTV-VT D-RNAs (Mawassi et al., 1995 b, c) further indicates that the observed variation reflects a genuine sequence characteristic of CTV-VT. Moreover, cDNA probes covering different parts of the CTV-VT genome gave similar hybridization signals with RNA blots from previously characterized CTV-VT D-RNAs (Mawassi et al., 1995 b, c). Further indications that jumped from one genome to another should result in a sharp sequence transition. Another possibility for the observed sequence differences is that the two halves of the CTV-VT and CTV-T36 genomes might result from differential pressures on sequence evolution due to host selection versus functional constraints of viral gene products.

Closteroviruses represent probably the most heterogeneous group of plant viruses (Dolja et al., 1994; Bar-Joseph et al., 1979). The type strains of BYV, CTV and LIYV have signature similarities for defining related viruses, that is a 5'-encoded polyprotein with protease, MT, HEL, and a +1 translation frameshift to produce RdRp. Additionally, they have five ORF cassette with similar genes. Otherwise, they have different genome structures. This is the first attempt to compare the variation between complete sequences of two CTV strains. It is noteworthy that deviation between the two CTV strains is greater than that among different viruses of most other groups (Ohira et al., 1995) raising the question of how to define the taxonomy of these viruses. Should CTV-T36 and CTV-VT be considered strains of the same virus or is the difference between these two isolates sufficient to define them as separate viruses?

The authors thank Dr R. Gafny and Mrs Lilach Ashoulin for useful comments. Drs V. Gaba, A. Karasev and W. O. Dawson for help with manuscript preparation and Mr Y. Ben-Shalom for excellent technical help. This work was supported in part by grants from the US-Israel Binational Agricultural Research and Development Fund (BARD), the German-Israeli Agricultural Research Agreement (GIARA) and the Citrus Marketing Board, Israel.

References


Received 18 March 1996; Accepted 20 May 1996