The complete genome sequence of the major component of a mild citrus tristeza virus isolate

M. C. Vives,1 L. Rubio,1 C. López,2 J. Navas-Castillo,1,† M. R. Albiach-Martí,3 W. O. Dawson,3 J. Guerri,1 R. Flores2 and P. Moreno1

1 Instituto Valenciano de Investigaciones Agrarias (IVIA), Apdo. Oficial, Moncada, Valencia, Spain
2 Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Universidad Politécnica de Valencia, Camino de Vera 14, 46022 Valencia, Spain
3 Citrus Research and Education Center, IFAS, University of Florida, 700 Experiment Station Road, Lake Alfred, FL 33850, USA

The genome of the Spanish mild isolate T385 of citrus tristeza virus (CTV) was completely sequenced and compared with the genomes of the severe isolates T36 (Florida), VT (Israel) and SY568 (California). The genome of T385 was 19 259 nt in length, 37 nt shorter than the genome of T36, and 33 and 10 nt longer than those of VT and SY568, respectively, but their organization was identical. T385 had mean nucleotide identities of 81.3, 89.3 and 94% with T36, VT and SY568, respectively. The 3′ UTR had over 97% identity in all isolates, whereas the 5′ UTR of T385 had 67% identity with VT, 66.3% with SY568 and only 42.5% with T36. In the coding regions, the nucleotide differences between T385 and VT were evenly distributed along the genome (around 90% identity); this was not observed between T385 and the other isolates. T385 and T36 had nucleotide identities around 90% in the eight 3′-terminal ORFs of the genome, but only 72.3% in ORF 1a, a divergence pattern similar to that reported previously for T36 and VT. T385 and SY568 had nucleotide identities close to 90% in the 5′- and 3′-terminal regions of the genome, whereas the central region had over 99% identity. Our data suggest that the central region in the SY568 genome results from RNA recombination between two CTV genomes, one of which was almost identical to T385.

Introduction

Citrus tristeza virus (CTV), a member of the genus Closterovirus, is the causal agent of one of the most economically important diseases of citrus. CTV virions are flexuous filaments about 2000 nm long (Bar-Joseph & Lee, 1989), with a single-stranded, positive-sense RNA genome and two capsid proteins of 25 and 27 kDa, coating 95 and 5% of the particle length, respectively (Febres et al., 1996). The virus is semi-persistently transmitted by several aphid species and its natural host range is restricted to the Rutaceae, mainly to species of the genera Citrus and Fortunella.

CTV isolates differ widely in their biological characteristics such as aphid transmissibility or symptoms caused in various host species and cultivars (Moreno & Guerri, 1997). These symptoms are of three general types: (i) decline and death of plants grafted on sour orange (Citrus aurantium L.); (ii) stunting and yellowing of sour orange, grapefruit (Citrus paradisi Macf.) or lemon [Citrus limon (L.) Burm. f.] seedlings (seedling yellows); and (iii) stem pitting of different citrus cultivars on their own roots or grafted on any rootstock. The genome regions involved in the expression of any of these symptoms are presently unknown. This information would be necessary to implement control measures, i.e. exclusion of the most severe isolates from budwood sources in citrus nurseries, eradication of trees infected with those isolates, monitoring cross-protected budwood sources or transformation of citrus plants to obtain protection mediated by viral sequences.
Comparison of sequences associated with a particular symptom phenotype is complicated by the fact that most CTV isolates are complex populations of different genomic and defective RNAs (Mawassi et al., 1995; Yang et al., 1997). However, most virus isolates consist of a ‘master sequence’ that usually makes up the majority of the population (Domingo et al., 1995). The complete genome sequences of the master sequence of three CTV isolates have been recently deposited in GenBank with accession numbers U16304 (isolate T36 from Florida; Karasev et al., 1995; Pappu et al., 1994), U56902 (isolate VT from Israel; Mawassi et al., 1996) and AF001623 (isolate SY568 from California). The genomes of these isolates contain about 19200–19300 nt and include 12 open reading frames (ORFs), potentially encoding at least 17 protein products, and untranslated regions (UTR) at the 5’ and 3’ termini.

The three isolates so far sequenced cause decline of plants grafted on sour orange, moderate to intense symptoms in Mexican lime (Citrus aurantifolia (Christm.) Swing.) and mild to intense seedling yellows. In addition, SY568 causes stem pitting in other citrus species like sweet orange (Citrus sinensis (L.) Osb.) or grapefruit. Here, we report the complete master sequence of the genome of a very mild isolate from Spain (T385) and compare it with those available from T36, VT and SY568 isolates. The T385 sequence differs considerably from those of T36 and VT, but it contains a genome segment conserved in T36, VT and various Spanish isolates.

**Methods**

**Virus isolates.** Isolate T385 was obtained from an old CTV-infected sweet orange/sour orange tree without decline symptoms (Moreno et al., 1991), aphid-transmitted to Mexican lime to eliminate other graft-transmissible pathogens present in the source tree, and kept in a greenhouse as part of the VIHA collection of isolates. It causes mild vein clearing in Mexican lime and Citrus macrophylla Wester, and no symptoms in other hosts.

**Cloning and sequencing.** CTV double-stranded RNA (dsRNA) was purified (Moreno et al., 1990) from infected young sweet orange bark and denatured with methylmercuric hydroxide (Sambrook et al., 1989). The approach used to clone the T385 genome included the following steps: (i) preparation of a randomly primed cDNA library and sequencing of selected clones; (ii) RT–PCR with appropriate T385-specific primers to fill the genome gaps left by this library; and (iii) preparation of clones of the 3’ and 5’ termini. All these clones were sequenced.

The randomly primed cDNA library was obtained with total dsRNA and the TimeSaver cDNA kit (Pharmacia Biotech), joining EcoRI adapters to the dsRNA fragments and ligating those to plasmid pUC18 (Albiach et al., 1989). Products of RT–PCR were cloned in linearized and thymidylated pT7Blue(R) plasmid (Novagen). To obtain clones of the 3’ and 5’ termini, denatured dsRNA was polyadenylated using yeast poly(A) polymerase (US Biochemical). After phenol-chloroform extraction and ethanol precipitation, the polyadenylated RNAs were reverse transcribed using primer PM-1 [CCGGATCCCTT-AGAGCGGCCGC(dT)12; V: in which V represents A, C or G] and then PCR-amplified using primers PM-1/RF-131 (nt 1448–1423) and PM-1/RF-116 (nt 18042–18068) to obtain the 5’ and 3’ termini of the viral RNA, respectively. The RT–PCR conditions were those described previously (López et al., 1998). Primers RF-131 and RF-116 were designed by selecting regions of the CTV RNA essentially conserved in T36 (Pappu et al., 1994; Karasev et al., 1995), VT (Mawassi et al., 1996) and in several Spanish isolates as revealed by previous RT–PCR amplifications (data not shown). PCR-amplified products were separated by agarose or polyacrylamide gel electrophoresis and ligated to the pT7Blue(R) plasmid. Sequences of the inserts were determined in both directions using chain-terminating inhibitors and T7 DNA polymerase (T7 Sequencing kit; Pharmacia Biotech) or by means of an ABI PRISM DNA Sequencer 377 (Perkin-Elmer).

The following programs of the GCG package (Devereux et al., 1984) were used: SEQED to edit sequences; ASSEMBLE to assemble cDNA clones; GAP, BESTFIT and PLOTSIMILARITY to compare sequences; PILEUP for multiple alignments; FOLDRNA and SQUIGGLES to obtain the predicted RNA secondary structure of minimum free energy; TRANSLATE to obtain amino acid sequences; and BLAST for database searching.

**Results and Discussion**

Fig. 1 outlines the relative sizes of the clones sequenced and their position in the T385 genome. The complete nucleotide sequence assembled from these clones has been deposited in the EMBL database (accession number Y18420). About 64 % of the genome was sequenced from randomly primed cDNA clones (Fig. 1, bottom), about 29.5 % was from RT–PCR products obtained with T385-specific primers, and the ends of the genome, encompassing the 5’-most 1422 nt and the 3’-most 1218 nt (Fig. 1, top), were from RT–PCR products obtained with primers containing oligo(dT) and sequenced conserved in T36, VT and various Spanish isolates.

Most of the randomly primed cDNA clones gave an intense signal in dot–blot hybridization with dsRNA extracts from T385, indicating that they probably represent the predominant sequence of this isolate. Additional confirmation of this point was provided by the fact that the overlapping regions of these clones (Fig. 1) were usually identical in nucleotide sequence, and differences were always below 0.5 %. Similarly, identical nucleotide sequences were obtained for each of the two clones analysed containing the 5’- and 3’-terminal regions of the genome. The primers based on the T385 sequence were selected so that they overlapped the randomly primed clones by at least 100 nt. The overlapping sequences were generally identical. The full genome length was sequenced at least twice.

The genome of T385 was 19259 nt in length, 37 nt shorter than the T36 genome, and 33 and 10 nt longer than the VT and SY568 genomes, respectively. T385 had a 6 nt insertion at the intergenic region between the p27 and p25 genes that was not present in the other isolates. Compared with T36, the genome of T385 had two 18 nt deletions in ORFs 1a and 1b, respectively, which were also deleted in isolates VT and SY568, and several deletions and insertions of 1–3 nt in other
Fig. 1. Schematic representation of the CTV genome and the strategy for cloning and sequencing the genome of CTV isolate T385. Boxes represent the ORFs and the corresponding putative protein products encoded by them are shown immediately above the boxes. The clones at the bottom were from a randomly primed cDNA library, and those at the top were prepared by RT–PCR using appropriate primers based on the T385 sequence obtained from the randomly primed cDNA clones. Clones of the 5′- and 3′-terminal regions (arrows) were obtained by RNA polyadenylation, reverse transcription with primer PM-1 and PCR using the primers PM-1/RF-131 and PM-1/RF-116 to amplify the 5′ and 3′ termini, respectively.

Table 1. Nucleotide (nt) identity and amino acid (aa) identity and similarity (%) between the different ORFs and UTRs of CTV isolate T385 compared to the corresponding regions of isolates T36, VT and SY568
The amino acid identity was greatest in acid identity and 81% amino acid identity was the most variable protein between isolates. Alignments of the amino acid sequence yielded predicted protein products of the two proteases, the methyl transferase and the helicase domains of this polyprotein. For example, the amino acid identities between the helicase domain of T385 and those of T36, VT and SY568 were 85–8, 93–8 and 96–7%, respectively, which were clearly higher than the amino acid identity values for the whole polyprotein.

As in other CTV isolates, ORF 1b of T385 encoded a protein with similarities to RNA-dependent RNA polymerases (RdRp) which, as suggested previously (Karasev et al., 1995), may be expressed via a +1 ribosomal frameshift mechanism. The same rare arginine codon (CGG) present in T36 and thought to be the frameshifting point (Karasev et al., 1995) is present in T385. In the T385 isolate, ORF 1b overlapped ORF 1a by 77 nt. The overlap was 123, 75 and 77 nt in T36, VT and SY568, respectively. The RdRp of T385 and those of T36, VT and SY568 had 91.2, 95.0 and 96.7% amino acid identity and 95.2, 96.7 and 96.9% amino acid similarity, respectively (Table 1). The previously defined domains I, IV, VII and VIII (Koonin & Dolja, 1993) were conserved in the RdRp of the four isolates. The amino acid sequences encoded by the remaining ORFs in T385 and in the other three isolates showed over 84–9% identity and over 89–8% similarity.

When the complete nucleotide sequence of T385 was aligned and compared with those of T36, VT and SY568, the mean nucleotide identities found were 81–3, 89–3 and 94%, respectively. However, a closer comparison showed that nucleotide differences were not evenly distributed along the genome in all isolates (Fig. 2; Table 1). The 3′ UTR region was very conserved, with nucleotide identities higher than 97% in all cases. On the other hand, the 5′ UTRs were variable and nucleotide identities between T385 and VT, SY568 and T36 were 67, 66–3 and 42.5%, respectively. Nucleotide identity between SY568 and VT was over 95%, whereas T36 had less than 60% identity with any of these isolates. Analysis of the polymorphism of the 5′ UTRs allowed the classification of CTV sequences into three discrete groups, with intragroup sequence identity higher than 88% and intergroup sequence identity as low as 44% (López et al., 1998). T36 was the isolate type for group I, VT for group II and various Spanish isolates belonged to group III. The predominant sequence of isolate T385 belonged to group III, whereas that of SY568 belonged to group II.

Nucleotide divergence in other genome regions was also variable depending on the isolates compared. For example, T385 and VT had identity values close to 90% in most ORFs (Table 1), and only in some specific regions were these values as low as 80% (Fig. 2). This divergence pattern contrasted with that reported for VT in comparison with T36. In this case, while the 3′ half of the genome was relatively conserved (nucleotide identities around 90%), nucleotide heterogeneity increased progressively toward the 5′ end reaching identity values below 60%, an unusually high difference between strains of the same virus (Mawassi et al., 1996). This asymmetric distribution of nucleotide differences was also observed when
T36 was compared with T385 or SY568. For example, T385 and T36 showed nucleotide identities around 90% in the eight ORFs close to the 3' terminus of the genome, but nucleotide identity in ORF 1a was only 72.3% (Table 1) and some regions of this ORF had identities of only 40% (Fig. 2). These comparisons indicate that T36 may be an atypical CTV isolate showing an unusually high nucleotide variability in the half of the genome close to the 5' terminus. Bar-Joseph et al. (1997) hypothesized that the asymmetric distribution of nucleotide identity between T36 and VT might indicate that one of these strains could have resulted from recombination between the largest CTV subgenomic RNA from a common ancestor and an ORF 1 of a diverse CTV strain or even from a different closterovirus. Our results suggest that T36 might be the descendant of such a recombinant virus, since the other three isolates do not show this asymmetric distribution of nucleotide variation between the two halves of the genome.

Comparison of T385 and SY568 again showed uneven distribution of nucleotide divergences. In these isolates, the two regions located at both termini of the genome had nucleotide identities close to 90%, whereas the region encompassing ORFs p33 through p61 had a nucleotide identity of over 99% (Table 1; Fig. 2). In particular, the gene encoding the small hydrophobic protein (p6) had an identical sequence in both isolates. The SY568 sequence deposited in GenBank has several positions in ORFs 1b (RdRp) and p27 at which nucleotides differed in the various clones sequenced. The nucleotide identity between T385 and SY568 for these genes was 91.6 and 92.1%, respectively (Table 1). However, considering those positions in which one of the alternative nucleotides given for SY568 coincided with the corresponding nucleotide in T385, the identity values for ORF 1b and p27 were 99.9 and 98.6%, respectively. The high degree of nucleotide identity observed in the middle region of the T385 and SY568 genomes (Fig. 2) could have resulted from a strand-switching artefact generated while preparing cDNA from this region by RT–PCR. However, two groups of observations indicate that this was probably not the case: (i) in T385, the sequence of p27 and the intergenic region downstream of it was obtained from seven independent randomly primed clones (Fig. 1) with identical sequences in overlapping regions, and the 5' terminus of the RdRp gene was located in the middle of clone h3 (Fig. 1), which seems to exclude template switching at this point; and (ii) in SY568, the two points where nucleotide divergence increased were also within cDNA clones. Moreover, RNase protection analyses indicated that the sequenced SY568 clones belonged to the predominant genome variant (E. Mirkov, personal communication). These findings strongly suggest that the abrupt change in nucleotide divergence between T385 and SY568 resulted from recombination between a common ancestor and a different CTV strain.

Various segments of the genome of a mild CTV isolate from Florida (T30) have been sequenced (M. R. Albiach-Martí, unpublished results). These segments have over 99% identity with T385, suggesting that SY568 might be the result of recombination. Furthermore, we used SY568 dsRNA and appropriate primers based on sequences conserved in T385, SY568, VT and T36 to prepare cDNA clones from three genome segments located within the potentially recombinant region. Sequencing of various clones of each segment revealed that SY568 did indeed contain two types of sequences, one of them almost identical to T385 (not shown). Recombination might have occurred between the two sequence variants present in this isolate.

Recombination between RNA genomes involves, in most cases, a replicase-driven template switching mechanism during RNA synthesis, similar to that occurring in the generation of defective RNAs (Lai, 1995). This phenomenon is more likely to occur in viruses with large genomes; thus, a recombination frequency of about 25% has been estimated for the entire genome of some coronaviruses (Lai, 1995). Closteroviruses are plant viruses with the largest known RNA genome and defective RNAs have, indeed, been observed in different CTV isolates (Mawassi et al., 1995; our unpublished results). Recently, subgenomic RNAs have been implicated in the generation of CTV defective RNAs (Yang et al., 1997) and they have also been considered as possible building blocks for modular recombination of Closterovirus genomes (Bar-Joseph et al., 1997). In view of the pattern of sequence similarity between SY568 and T385, recombination apparently occurred at the end of ORF 1a and at the end of gene p27. The intergenic region downstream of p27 has a predicted secondary structure with two stem–loops (data not shown) in the four isolates studied. Such secondary structure might cause the RNA polymerase to pause and facilitate its jumping to a new template RNA molecule (Nagy & Simon, 1997). It is noteworthy that T385 has a 6 nt insertion, which is not present in the other isolates, in one of the two potential loops in the intergenic region between p27 and p25 (the loop closest to the 5' terminus). The absence of this insertion in SY568 might indicate that recombination occurred upstream of this loop, close to the end of p27. We have not found any sequence or secondary structure that might help to explain the recombination that has apparently occurred at the end of ORF 1a.

This work was funded in part by the INIA project SC39111, the Conselleria de Agricultura, Pesca y Alimentación and the Conselleria de Cultura, Educación y Ciencia de la Generalitat Valenciana (project GV3116/95), and the INCO-DC Program of the European Commission DG-XII (contract no. ERBIC18CT960044). Luis Rubio was recipient of a fellowship from the Conselleria de Cultura, Educación y Ciencia.

References


Received 9 September 1998: Accepted 23 November 1998