Complete sequence of an infectious full-length cDNA clone of citrus tatter leaf capillovirus: comparative sequence analysis of capillovirus genomes

Kazuyuki Ohira,‡ Shigetou Namba,⁎ Michael Rozanov, Takaaki Kusumi and Tsuneo Tsuchizaki

1 Laboratory of Bioresource Technology, Division of Agriculture and Agricultural Life Sciences, Graduate School, The University of Tokyo, Midori-cho, Tanashi, Tokyo 188, Japan, 2 Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA, 3 Plant Biotechnology Laboratory, Institute of Fundamental Research, Suntory Research Center, Mishima-gun, Osaka 618, Japan and 4 Laboratory of Plant Pathology, Koibuchi Gakuen, Koibuchi, Uchihara-machi, Higashiibaragi-gun, Ibaragi 319-03, Japan

The complete nucleotide sequence of citrus tatter leaf capillovirus (CTLV lily strain) was determined. It is 6496 nucleotides long, excluding the 3'-terminal poly(A) tract, and contains two putative overlapping open reading frames (ORFs). ORF1 (positions 37-6354) encodes a potential polyprotein of molecular mass 242 kDa. ORF2 (positions 4788-5750) codes for a 36 kDa protein. The 242 kDa polypeptide contains several non-structural protein domains (i.e. methyltransferase, NTP-binding helicase, papain-like protease and polymerase) and, at its C terminus, the putative coat protein. The N-terminal region of the 36 kDa protein displays sequence similarity to the cell-to-cell movement proteins of the ‘30K superfamily’. Such a genome structure is conserved between CTLV and apple stem grooving capillovirus. Capped transcripts from a plasmid containing the complete sequence of CTLV, with a T7 RNA promoter, successfully infected Chenopodium quinoa plants and caused symptoms characteristic of CTLV. Uncapped transcripts were non-infectious.

Capilloviruses and closteroviruses are diverse genera of structurally similar, flexuous, filamentous plant viruses. Their taxonomy has been unclear due to a lack of information other than on morphological features (Coffin & Coutts, 1993). The capillovirus genus includes apple stem grooving virus (ASGV; Lister, 1970), the type member of the group, potato virus T (PVT; Salazar & Harrison, 1978), citrus tatter leaf virus (CTLV; Nishio et al., 1989), lilac chlorotic leaf spot virus (LCLV; Brunt, 1978) and Nandina stem pitting virus (NSPV; Ahmed et al., 1983).

Recently, the genomes of PVT (Ochi et al., 1992) and ASGV (Yoshikawa et al., 1992) were sequenced. Surprisingly, the genome structure of PVT was found to be homologous to that of apple chlorotic leaf spot virus (ACLSV; German et al., 1990), a member of the closterovirus genus, while ASGV had a completely new genome structure. The nucleotide sequence and genome organization of the 3' terminus of CTLV-L RNA (Yoshikawa et al., 1993; Ohira et al., 1994) displayed a strong similarity to that of ASGV. In this paper we report the complete sequence and genome organization of CTLV-L, compare it with the genomes of other related viruses, and describe the production of infectious transcripts of CTLV-L from a full-length clone of the genome. A preliminary report of this study was presented at a Meeting of the Phytopathological Society of Japan, Tokyo, September 1993.

CTLV-L RNA was prepared as described previously (Ohira et al., 1994). cDNA was synthesized from CTLV-L RNA using the cDNA synthesis system Plus (Amersham) with NotI tagged oligo(dT) 5' ATGCGGCCGCGG(dT)18 3' as primer. cDNAs larger than 3 kbp were recovered by low melting point agarose gel electrophoresis, inserted into the EcoRV site of pBluescriptII SK(-) (Stratagene) and cloned in Escherichia coli XL1-Blue. Screening by restriction enzyme mapping, Northern blotting and partial sequencing allowed identification of the clone pCL1 containing a cDNA insert of 5.0 kbp covering the 3' terminus of CTLV-L RNA. After sequencing the 5'-terminal region of the insert of pCL1, a complementary
CTLV RNA

ASGV RNA

ACLSV RNA

PVT RNA

Fig. 1. (a) Amino acid sequence identity between CTLV and ASGV. Regions in ASGV ORFs with the same amino acids as CTLV are shown as open boxes; different amino acids are indicated by vertical black lines. (b) Comparison of the genomic organization of ACLSV with that of PVT. In both (a) and (b) horizontal bars indicate viral genomes and open boxes indicate ORFs.

For nucleotide sequencing, pCL1 and pCL1.5 were digested with PstI, HindIII, BamHI, EcoRI or SacI, and selected fragments were sub-cloned in the replicative form (RF) of M13mp18, M13mp19 or Bluescript II. Deletion mutants were also prepared from the clones using the Kilo-Sequence Deletion Kit (Takara Shuzo). Shotgun cloning was also performed by inserting cDNA fragments digested with Sau3AI or AluI into the BamHI or HindII site of M13mp18 RF, respectively. The resultant cDNA subclones were sequenced by the dideoxyribonucleotide chain termination method (Sanger et al., 1977) using Sequenase version 2.0 (United States Biochemical) or Klenow fragment (Toyobo). All of the nucleotide sequence was determined on both strands, except for the 5'-terminal base. The sequence of the 5'-terminal region was determined by extending a synthetic primer (5' AGTTCAAGTAAACCCCAATA 3') complementary to nucleotides 104–120 of the viral RNA with dideoxynucleotides using reverse transcriptase and terminal deoxynucleotidyl transferase (Deborde et al., 1986). All nucleotide sequence data were collected and analysed using the program DNA-SIS (version 7.01) (Hitachi Software Engineering).

The CTLV genome was found to consist of 6496 nucleotides, excluding the 3' poly(A) tail (GenBank accession no. D16681). This value agrees with previous estimates by electrophoresis of poly(A)-tailed RNA denatured with formaldehyde (K. Ohira and others, unpublished) and coincides with the length of the ASGV genome (Yoshikawa et al., 1992). The 3'-terminal sequence showed 99% identity with the CTLV-L sequence reported by Yoshikawa et al. (1993).

A search for putative open reading frames (ORFs) on both the positive and complementary strands of the CTLV-L genome showed that two overlapping ORFs of substantial size were present in the positive strand (Fig. 1a). ORF1 begins at the first AUG (nucleotide positions 37–39) and terminates at a UAG (amber) codon (positions 6350–6354) to yield a large potential poly-peptide with a calculated M r of 242 kDa (Fig 1a). The AUG codon at positions 37–39 fits with the optimal sequence context for plant mRNAs (LiJtcke et al., 1987).

ORF2 lies in a different reading frame, contained within ORF1, beginning at an AUG (opal) codon at positions 5748–5750. ORF2 encodes a polypeptide of M r 36142 (36 kDa). The initiation codon of ORF2 also fits with the optimal context for plant mRNAs (Lütke et al., 1987).

The 242 kDa protein encoded by ORF1 contains all the characteristic motifs for (putative) methyltransferase (MTR; amino acid positions 63–250, Rozanov et al., 1992; Fig. 2a), papain-like proteinase (P-PRO; positions 596–699, Rozanov et al., 1995; Fig. 2b), NTP-binding helicase (HEL; positions 781–788, Gorbalenya et al., 1988) and RNA-dependent RNA polymerase (RdRp;
Short communication

(a)

<table>
<thead>
<tr>
<th>Virus</th>
<th>ORF</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNBV</td>
<td>nsP1</td>
<td>30-260 VAAQVTNSQANARAFSLASK-35-RSPFDH-151-TL-HEHSLTQSWH</td>
</tr>
<tr>
<td>ACLSV</td>
<td>p217</td>
<td>63-232 SFFAYVHSQGCKLTHMLEL-43-VTAQEK-74-SEGQFPLEN-GFL</td>
</tr>
<tr>
<td>ASGV</td>
<td>p241</td>
<td>63-250 AGFGRSHPISKMIENHLLY-47-IDKQY642-97-SEGQFNANS-KWP</td>
</tr>
<tr>
<td>CTLV</td>
<td>p242</td>
<td>63-250 AGFGRSHPISKMIENHLLY-47-IDKQY642-97-SEGQFNANS-KWP</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Virus</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYMV</td>
<td>773 SNHLPQPTLALLSASVDQTKVSEEHLEWLSQTLRDQGSLNEETNL</td>
</tr>
<tr>
<td>ACLSV</td>
<td>843 KLRFIKGKVQLFVSAEIHKKEPEEMFLHIPMDCVSPG</td>
</tr>
<tr>
<td>ASGV</td>
<td>596 KLPRRRKNDQPKAISAHGIDSQEQLNLVWEDISDELDCIIEEDSG</td>
</tr>
<tr>
<td>CTLV</td>
<td>596 KCPRRRKNNLDQPKAISAHGIEYQDLNLVWEDISDELDCIIEEDSG</td>
</tr>
</tbody>
</table>

Fig. 2. (a) Alignment of the conserved 'signature' sequences of the putative methyltransferase domain of CTLV with those of ACLSV, ASGV and SNBV. Conserved amino acid sequences are boxed. (b) Alignment of the putative papain-like proteinase sequence of CTLV with those of TYMV, ACLSV and ASGV. The amino acid residues conserved through all (predicted) tymo-like proteinases are boxed.

The 36 kDa protein, a potential product of ORF2, bears in its N-terminal part the amino acid signature of the '30 K superfamily' of cell-to-cell transport proteins (data not shown; Mushegian & Koonin, 1993). The region of ORF1 that contains ORF2 is not conserved between CTLV and ASGV; only 54% identity exists in ORF1 amino acids 1583–1851. This suggests that these regions encode virus-specific proteins (such as proteins determining host range) rather than conserved proteins (such as proteins essential for virus replication), or that there are less severe evolutionary constraints than on the rest of the viral proteins (Ohira et al., 1994). If the latter is the case, it could be speculated that ACLSV or PVT evolved from capilloviruses by losing this dispensable region altogether. This could have led to a mechanism for coat protein expression in ACLSV and PVT different from that found in capilloviruses.

The C-terminal domain of the 242 kDa protein possesses the conserved amino acid sequence motifs (including those with invariant Ser, Arg and Asp) of filamentous viral coat proteins. The AUG at position 5641–5643 is in a good context for initiation of translation and the putative translation product from it has a calculated Mr of 27157 which agrees well with the 27 kDa estimated by SDS–PAGE (K. Ohira and others, unpublished).

Comparison of the predicted amino acid sequences with those of a 3′-terminal cDNA clone of CTLV-L (Yoshikawa et al., 1993) shows 99.0% similarity in ORF1 and 98.8% similarity in ORF2.

Fig. 1(a) shows the amino acid sequence identity between CTLV and ASGV. Similarity between the viruses in ORF1 is 88.5%, while in ORF2 it is 94.7%. Considering these high similarities, CTLV and ASGV may be strains of the same virus rather than different viruses. This is also supported by the 94% identity in the 3′ non-coding region, by serological cross-reactivity, and by the ability of ASGV to infect citrus and to induce CTLV-like symptoms in these hosts (T. Iwanami, personal communication). However, CTLV and ASGV have unique host preferences in different climates (i.e. citrus plants grow in warmer areas and apples in colder areas). These differences may be due to the ORF1 proteins. There are two hyper-variable regions in ORF1. The first is at positions 534–569, where the amino acid sequence similarity is only 22%. This region is upstream of the papain-like proteinase domain. The second region is at positions 1583–1851, with a similarity of 54%.

CTLV has a large ORF which includes conserved motifs for MTR, P-PRO, HEL, RdRp and coat protein. The protein may be co- or post-translationally cleaved by a mechanism which remains to be resolved. Turnip
yellow mosaic virus (TYMV) has a papain-like proteinase domain in the central region of ORF1 (Rozanov et al., 1995). It is probable that a similar proteinase domain exists in CTLV ORF1. Yoshikawa et al. (1992) stated that there was homology between the ORF2 protein of ASGV and a serine proteinase. This motif is conserved in CTLV ORF2, but further experiments are required to confirm its activity.

According to the taxonomy of Coffin & Coutts (1993), CTLV, ASGV and PVT belong to the capilloviruses while ACLSV is classified as a closterovirus. However, recent sequence comparisons have revealed that this taxonomy needs improvement because the genome organizations of ASGV and PVT differ (Fig. 1 b). ASGV and CTLV have the same genome organization, as described above. Moreover, PVT has a genome organization similar to ACLSV (Fig. 1 b and our unpublished data). It is more reasonable to classify PVT in the group with ACLSV. In accord with these results, the genus Trichovirus was proposed and accepted as a new genus at the last meeting of the International Committee on Taxonomy of Viruses (ICTV) (Martelli et al., 1994). The conclusion reached was that CTLV and ASGV remain classified in the genus Capillovirus, while ACLSV and PVT are classified as members of the new genus Trichovirus.

In order to begin straight-forward experiments on genome strategy and pathogenesis of capilloviruses we constructed a cDNA copy of the CTLV RNA, the plasmid pITCL, as follows (Fig. 3). Numbering refers to nucleotide positions (nt) in the complete CTLV sequence. The synthetic primers 5' AGTCGACTAATACGACTCATATAGATATTAATTTAACAGCGCTTAA 3' [A(Sa/I site)T7 RNA promoter]G(5'-terminal 18 nucleotides of the CTLV genome) and 5' GGTCAAGGAAAGCTATT 3' (complementary to CTLV nt 104-120) were used for amplification of a 120 bp fragment from the 5' terminus with a T7 RNA promoter just upstream of an additional guanine residue (-1; for more efficient initiation of transcription) followed by the 5'-terminal adenine residue (nt+1). The SaII and Sau3AI (nt 100) digested PCR product was inserted between the SaII and BamHI sites of pUC18. This resulted in plasmid pIT100. The PstI (nt 973)–HpaI (nt 1390) fragment and the HpaI (nt 1390)–BamHI (nt 2997) fragment of CTLV were inserted between the PstI–BamHI sites of pUC18 by three-fragment ligation. This resulted in plasmid p18PB. The Aor51HI (nt 13)–PstI (nt 973) fragment from pCL1.5 and the PstI (nt 973)–KpnI (nt 2274) fragment from p18PB were inserted between the Aor51HI (nt 13) and KpnI sites of pIT100 by three-fragment ligation. This resulted in plasmid p18SK. The KpnI (nt 2274)–SmaI
fragment of pCL1 was then inserted between the KpnI and SmaI sites of p18SK. This resulted in plasmid pITCL.

In vitro transcription reactions, plus or minus the cap analogue \([\text{m7(5')Gppp(5')G}]\), were performed using NotI-linearized pITCL and bacteriophage T7 RNA polymerase as described by the manufacturer (Stratagene). RNA (1 \(\mu\)g) in 8 \(\mu\)l of sterilized water was used for mechanical infection. RNA transcribed in vitro with cap analogue successfully infected Chenopodium quinoa plants and gave CTLV-specific symptoms (not shown). Virus particles which had the characteristic features of CTLV were also detected in local lesions on inoculated leaves and from systemic leaves (data not shown). Thus, our CTLV clone and the corresponding RNA sequence are functional. On the other hand, RNA transcribed in vitro without cap analogue did not infect C. quinoa and gave no symptoms.

This is the first report on the construction of infectious transcripts of a capillovirus. Using such transcripts we will determine the sequence of the complete genome of CTLV and analyse the gene expression strategy of capilloviruses.

We gratefully acknowledge Dr S. Kashiwazaki, National Agriculture Research Center, Tsukuba, Japan for technical advice and T. Ito, Fruits Experimental Station, Okitu, Shizuoka, Japan for technical assistance.

References


(Received 10 February 1995; Accepted 3 May 1995)