The nucleotide sequence of citrus leaf rugose ilarvirus RNA-2

Xin Ge and S. W. Scott*

Department of Plant Pathology and Physiology, Clemson University, Clemson, South Carolina 29634-0377, U.S.A.

The nucleotide sequence of citrus leaf rugose ilarvirus (CiLRV) RNA-2 consists of 2990 nucleotides and contains one open reading frame (ORF) which encodes a deduced translation product of 832 amino acids with a calculated M_r of 95 501 (95K). The 5' terminus of the RNA has a m7Gppp cap. Both the nucleotide sequence of CiLRV RNA-2 and its translated polypeptide share homologies with the nucleotide sequence and translated polypeptide, respectively, of RNA-2 of alfalfa mosaic virus (AlMV). The homology occurs in the central region of both the nucleic acid sequence and the polypeptide. Homologies between either CiLRV or AlMV and other Bromoviridae (cucumber mosaic virus - CMV, brome mosaic virus - BMV and cowpea chlorotic mottle virus - CCMV) were far less. Alignment of the 104 amino acid region (polymerase signature) of the 95K protein against 10 other 'alpha-like' plant viral polymerase signatures showed that CiLRV and AlMV are more closely related to each other than to CMV, BMV or CCMV. This is the first report of the full-length sequence of RNA-2 of an ilarvirus.

Citrus leaf rugose virus (CiLRV) is a member of subgroup 2 of the ilarvirus genus (Hamilton, 1991). It has a tripartite genome composed of the following single-stranded (ss) RNAs: Mr 1.1 x 10^8 (RNA-1), 1.0 x 10^6 (RNA-2) and 0.7 x 10^6 (RNA-3). A subgenomic RNA of Mr 0.3 x 10^6 (RNA-4) is also encapsidated in the virions. It is thought that ilarviruses share similar genome structure and expression strategies to those found in other members of the Bromoviridae family: bromoviruses, cucumoviruses, alfalfa mosaic virus (AlMV) (Mayo & Martelli, 1993), and the idaeovirus raspberry bushy dwarf (Ziegler et al., 1993). However, this remains to be demonstrated. To date only sequence data for the type member of the group (tobacco streak ilarvirus – TSV) has been published, although sequence data for other ilarviruses exist (Alrefai et al., 1994; Bachman et al., 1994). The full-length sequence for RNA-3 of TSV has been determined (Cornelissen et al., 1984) and data for approx. 200 nucleotides at the 3' end of RNA-2 are also known (Koper-Zwarthoff & Bol. 1980).

The RNA-2 of members of the Bromoviridae encodes a polymerase (Ahlquist et al., 1984; Cornelissen et al., 1983; Rezaian et al., 1984). As RNA dependent polymerases are probably among the most ancient of viral enzymes and appear to be more highly conserved than coat proteins, they are potentially ‘the best ruler’ for molecular taxonomy of RNA viruses (Bruenn, 1991). Although sequence data for the other members of the Bromoviridae exist, the lack of sequence data for ilarvirus genomes, especially sequences of RNA-1 and RNA-2, has made it impossible to establish a comprehensive taxonomy for the Bromoviridae.

A close relationship between ilarviruses and AlMV has been proposed (van Vloten-Doting & Jaspars, 1977; van Vloten-Doting et al., 1981). Both AlMV and TSV need coat protein, or their subgenomic RNAs, to activate the genome and initiate infection (van Vloten-Doting, 1975). However, this requirement for genomic activation is not specific and the coat proteins from these two viruses, as well as those of CiLRV and citrus variegation ilarvirus (CVV), will activate homologous as well as heterologous genomes. (Gonsalves & Garnsey, 1975). In addition, TSV and AlMV exhibit similarities in the secondary structures at the 3' ends of the RNA-3 (Koper-Zwarthoff & Bol. 1980).

Here we report the first full-length sequence of an ilarvirus RNA-2 and discuss the possible relationship between ilarviruses, AlMV, Bromoviridae, and other ‘alpha-like’ plant viruses.

Purified CiLRV was kindly provided by Dr S.M. Garnsey (USDA, Orlando, Fla., U.S.A.). Viral RNA was isolated by a guanidinium isothiocyanate/CsCl centrifugation method (Davis et al., 1986) with minor modifications. Total viral RNA was polyadenylated with Escherichia coli poly(A) polymerase (Sippel, 1973) and cDNA synthesis was completed with a cDNA synthesis kit (Amersham) using oligo(dT)_{12-18} as the primer. dsDNAs were ligated into the EcoRV site of pBluescript II SK+ (pSK) plasmid vector (Stratagene). Clones were screened by gel electrophoresis and those with inserts...
> 200 bp were retained. A clone (pCiLRV9) was obtained which, when sequenced, was found to contain a 190 nucleotide (nt) sequence at the 3' end that was almost identical to the sequence of CiLRV RNA-3 (unpublished nucleotide (nt) sequence at the 3' end that was almost obtained which, when sequenced, was found to contain a 190 nt sequence). A 20 oligonucleotide section (5' to 3') was used as the primer to synthesize cDNA that extended from the primer site toward the 5' end of the molecule. Two clones, pB11 (2.7 kb) and pB14 (1.4 kb) were isolated using this cDNA. Clone pA3 (817 bp) was obtained from a library synthesized with an 8-mer library sequence.

**Fig. 1.** Nucleotide sequence of CiLRV RNA-2 and the encoded amino acid sequence of the ORF.

---

**ORF.**
Fig. 2. Dot-matrix analysis of points of similarity between the nucleotide sequences of RNA-2 of CiLRV and AIMV. The nucleotide sequences were plotted against each other with a base window of 25 and stringency of 18. The number of points of identity is shown in the upper-left corner of the plot.

oligonucleotide (5' GCATCTCC 3') which we have previously determined to be complementary to the 3' end of CiLRV RNA-3, citrus variegation virus (CVV) RNA-3 and TSV RNA-3 (unpublished data). To determine the origin of clone pB11, CiLRV RNAs were separated in a 2.5% low melting point agarose gel. Individual bands were excised from the gel and the RNA was recovered and used as templates for PCR with synthetic primers. PCR was performed using a GeneAmp RNA PCR kit (Perkin Elmer Cetus) with primers designed from the sequence of clone pB11, 5' GCACTACGTCAGGATGG 3' (nt 1240 to 1257, negative sense) and 5' TTGGTGTTGGAGGGCGC 3' (nt 1811 to 1828, sense). PCR reactions were completed in which total CiLRV RNA, gel-isolated RNA-1, -2, or -3 of CiLRV, and total RNA from CMV were used as templates.

To confirm that the RNA had a 5' cap structure, purified CiLRV RNA was treated with tobacco acid pyrophosphatase (TAP), dephosphorylated with calf intestinal alkaline phosphatase (CIAP), labelled with [γ-32P]ATP, and the incorporation of radioisotope was determined in a scintillation counter. DNA from clones pB11, pB14 and pA3 was sequenced in both directions using an automated 373A DNA sequencer (Applied Biosystems). The 5' end sequence of the CiLRV RNA-2 was determined by reverse transcription sequencing using an internal oligonucleotide primer (5' GGATCTCAACGTCGGCC) and terminal deoxynucleotidyl transferase (DeBorde et al., 1986). Nucleic acid sequences were compared and/or translated using either programs in the GCG analysis package or by using Gene Jockey software.

Incorporation of radioactivity at the 5' end of the RNA was 6.5-fold higher when the RNA was treated with TAP and CIAP than when treated with CIAP alone. Without either treatment, no labelling was detected. Agarose gel electrophoresis of the samples revealed that the intact RNAs were labelled (data not shown). These results suggest, in toto, that the RNA possesses a cap structure at the 5' terminus.

When RNA PCR with pB11-specific primers was used, an amplified DNA fragment with the expected size (640 nt) was present with total CiLRV RNAs and gel-isolated CiLRV RNA-2, but not when gel-isolated RNA-1 or RNA-3 of CiLRV or total RNAs from CMV were used, thus confirming that pB11 was a clone to RNA-2. The complete sequence of CiLRV RNA-2 is shown in Fig. 1. The sequence contains 2990 nt, with 798 A, 617 C, 691 G and 884 U. The ORF covers the region from 74 to 2572 nt and encodes a deduced translation product of 832 amino acids with a calculated Mr of 95 501. The 5' non-coding region of the RNA contained 73 nt while the 3' noncoding region of CiLRV RNA-2 contained 418 nt. The 3' proximal 190 nt of CiLRV RNA-2 were shared with RNA-I and RNA-3 of CiLRV with only slight mismatching (four mismatches with RNA-3 and eight mismatches with RNA-1—unpublished data). This 190 nt region could be folded to form a secondary structure similar to those proposed for AIMV and TSV.

CI LRV

| ITYHKGIVM QSSPLFLSAMSRLFVYLVK S KIHP S GKNQHQLFTLDAAAF D |
| ITYHDKIVM QSSPFL PLAARILIMP L KDRIFTPS GKHQFLPSIDEA AF D |

AIMV

| I TY D K T | G V G AVDFQ RRTGD ANTYL G NTVISLCI LRV YL CT TPI VFI |
| I SDEK GVFP FNVDQ RGDAL Y TNTVI LACI LRVYL DPM VYFC V |

| ASGD D LI G SVEELPRAP EHLFTSLNFEAK FPHNQ PPC S KI L VSWOLV |
| ASGD D LI G SVEELPRAP EHLFTSLNFEAK FPHNQ PPC S KI L VSWOLV |

Fig. 3. Highly conserved area of the proteins translated by the single ORFs of CiLRV RNA-2 and AIMV RNA-2, respectively. The polymerase signatures of the two viruses are indicated in bold type.
Table 1. Amino acid sequence homologies (%) in the polymerase signature regions of the 11 viruses (strains) in ‘alpha-like’ virus group II

<table>
<thead>
<tr>
<th></th>
<th>CiLRV</th>
<th>AIMV</th>
<th>BMV</th>
<th>CCMV</th>
<th>CMVF</th>
<th>CMVQ</th>
<th>TMVv</th>
<th>TMVo</th>
<th>TRV</th>
<th>PEBV</th>
<th>BSMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CiLRV</td>
<td>64</td>
<td>51</td>
<td>47</td>
<td>48</td>
<td>47</td>
<td>33</td>
<td>32</td>
<td>35</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>AIMV</td>
<td>51</td>
<td>50</td>
<td>49</td>
<td>52</td>
<td>50</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>34</td>
<td>34</td>
<td>32</td>
</tr>
<tr>
<td>BMV</td>
<td>48</td>
<td>52</td>
<td>64</td>
<td>66</td>
<td>64</td>
<td>38</td>
<td>38</td>
<td>33</td>
<td>35</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>CCMV</td>
<td>47</td>
<td>49</td>
<td>85</td>
<td>66</td>
<td>68</td>
<td>41</td>
<td>42</td>
<td>38</td>
<td>40</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>CMVF</td>
<td>59</td>
<td>56</td>
<td>50</td>
<td>34</td>
<td>34</td>
<td>38</td>
<td>38</td>
<td>35</td>
<td>40</td>
<td>40</td>
<td>36</td>
</tr>
<tr>
<td>CMVQ</td>
<td>53</td>
<td>39</td>
<td>34</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>36</td>
<td>40</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>TMVv</td>
<td>33</td>
<td>34</td>
<td>34</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>36</td>
<td>40</td>
<td>40</td>
<td>36</td>
</tr>
<tr>
<td>TMVo</td>
<td>32</td>
<td>35</td>
<td>32</td>
<td>33</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>TRV</td>
<td>35</td>
<td>36</td>
<td>32</td>
<td>31</td>
<td>33</td>
<td>33</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>PEBV</td>
<td>34</td>
<td>34</td>
<td>35</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>BSMV</td>
<td>34</td>
<td>32</td>
<td>30</td>
<td>27</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>36</td>
<td>36</td>
<td>36</td>
</tr>
</tbody>
</table>

* According to Candresse et al. (1990).

Abbreviations used in this table and the sources of the sequence are as follows: CiLRV, citrus leaf rugose virus; AIMV, alfalfa mosaic virus RNA-2, Cornelissen et al. (1983); BMV, brome mosaic virus RNA-2, Ahlquist et al. (1984); CCMV, cowpea chlorotic mottle virus RNA-2, Allison et al. (1989); CMVF, cucumber mosaic virus strain Fny RNA-2, Rizzo & Palukaitis (1988); CMVQ, CMV strain Q RNA-2, Rezaian et al. (1984); TMVv, tobacco mosaic virus strain vulgaris, Goel et al. (1982); TMVo, TMV strain tomato, Ohno et al. (1984); TRV, tobacco rattle virus RNA-1, Hamilton et al. (1987); PEBV, pea early browning virus RNA-1, Macfarlane et al. (1989); BSMV, barley stripe mosaic virus RNA-γ, Gustafson et al. (1987).
known for other Bromoviridae. However, we do not know if the nucleotides adjacent to the cap are methylated.

The putative protein coded for by the single ORF is larger than the same protein coded for by AIMV RNA-2 but falls into the middle of the size range for translation products of the ORFs of RNA-2s of known Bromoviridae. A close relationship between ilarviruses and AIMV has been proposed. However, incorporating AIMV into the ilarvirus group has always been argued against because of differences in the mode of transmission of the two viruses. A lack of homology between coat proteins of AIMV and ilarviruses and the putative role of coat proteins in any aphid transmissibility would mitigate in favour of maintaining the separation of the two viruses. However, the RNA-2 of AIMV and CiLRV share sequence homology both at the nucleic acid and at the protein level while comparisons between CiLRV and AIMV and other Bromoviridae indicate much lower homology. In addition, the secondary RNA structure at the 3' ends of both AIMV and the two ilarviruses was similar. This ability to fold into loops does not occur in bromoviruses and cucumoviruses. The interchangeability of AIMV and ilarvirus coat proteins in initiating infection, plus the occurrence of coat protein recognition sequence (AUGC) found in AIMV and TSV and in all ilarviruses that we have examined (see Houser-Scott et al., 1994), would appear to be supporting evidence for, at the least, a common ancestor and most probably a closer relationship between AIMV and the ilarvirus group. The comparisons of the polymerase signatures of 11 of the ‘alpha-like’ viruses (group II, according to Candresse et al., 1990) confirm a closer relationship between AIMV and CiLRV (and hence the ilarviruses) than relationships to other viruses. Furthermore, only CiLRV and AIMV had 20 amino acids between the xxGxxxTxxNxx and xxxGDDxxx motifs of the signature while all the other viruses had 18. The significance of the additional two amino acids is unknown, but it nevertheless supports a relationship in which CiLRV and AIMV are more closely related to each other than to either the other viruses in group II of the ‘alpha-like viruses’ or other genera in the Bromoviridae.

This is technical contribution no. 3527 of the South Carolina Agricultural Experiment Station.

References


(Received 15 February 1994; Accepted 11 May 1994)