The nucleotide sequence of citrus leaf rugose virus RNA 1

S. W. Scott* and Xin Ge†

Department of Plant Pathology and Physiology, Clemson University, Clemson, SC 29634-0377, USA

The nucleotide sequence of citrus leaf rugose virus (CiLRV) RNA 1 consists of 3404 nucleotides and contains one open reading frame (ORF) which encodes a putative translation product of 1051 amino acids with a calculated Mr of 118339. Both the nucleotide sequence of CiLRV RNA 1 and its translated polypeptide share similarities with those of the RNA 1 of alfalfa mosaic virus. However, the relationship is not as close as that which exists between the polymerase signatures of the two viruses, which are found on RNA 2. This is the first report of the full-length sequence for the RNA 1 of an ilarvirus and completes the first sequence for an entire ilarvirus genome. If it is typical of members of the genus then, as has long been speculated, the genomic organization of ilarviruses is identical to that of other genera in the family Bromoviridae.

Citrus leaf rugose virus (CiLRV) is a member of the genus Ilarvirus (Murphy et al., 1995). This genus, together with the genera Alfamovirus, Bromovirus and Cucumovirus, comprise the Bromoviridae, a recently established family of viruses with tripartite, ssRNA genomes (Murphy et al., 1995). The genomic segments of ilarviruses have typical Mr values of 1.1 x 10⁶ (RNA 1), 0.9 x 10⁶ (RNA 2) and 0.7 x 10⁶ (RNA 3). A subgenomic RNA of 0.3 x 10⁶ (RNA 4) is also encapsidated in the virions (Murphy et al., 1995). Sequence data indicate that the RNA 2, RNA 3 and RNA 4 of CiLRV are 2990, 2289 and 1009 nt in length, respectively (Ge & Scott, 1994; Scott & Ge, 1995). The data for RNA 3 and RNA 4, both in terms of length and organization, are in good agreement with the few published sequences available for ilarviruses and also agree with sequence data available for more extensively studied members of the Bromoviridae. However, the RNA 2 of CiLRV is at present the sole published sequence for the RNA 2 of an ilarvirus (Ge & Scott, 1994) and no sequence for the RNA 1 of an ilarvirus has yet been published.

The RNA 1 and 2 of members of the Bromoviridae encode viral replicases (Ahlquist et al., 1984; Cornelissen et al., 1983; Rezaian et al., 1984; Ahlquist, 1992). As these RNA-dependent replicases are probably among the most ancient of viral enzymes and appear to be more highly conserved than coat proteins, they are eminently suited to provide information on the evolutionary relationships among viruses (Bruenn, 1991). Although sequence data for the RNA 1 and RNA 2 of other members of the Bromoviridae exist, the lack of similar data for ilarvirus genomes has made it impossible either to establish a comprehensive taxonomy for the Bromoviridae or to compare relationships at the molecular level among all the genera in this family.

A close relationship between ilarviruses and alfalfa mosaic virus (AMV) has been proposed (van Vloten-Doting & Jaspars, 1977; van Vloten-Doting et al., 1981) on the basis of their common requirement for the presence of coat protein to activate the genome, the fact that the coat proteins of AMV and ilarviruses are interchangeable in this respect (Gonsalves & Garnsey, 1975; van Vloten-Doting, 1975) and similarities in the secondary structures at the 3' ends of RNA 3 (Koper-Zwarthoff & Bol, 1980). Comparison of the sequence for RNA 2 of CiLRV with the RNA 2 of other members of the Bromoviridae has confirmed both a close relationship between ilarviruses and AMV (the sole member of the genus Alfamovirus) and a more distant relationship to Cucumovirus and Bromovirus, the other genera in the family Bromoviridae (Ge & Scott, 1994).

Here we describe the first full-length sequence of an ilarvirus RNA 1. By comparing the N-terminal methyltransferase-like domain (N-terminal signature) and the C-terminal helicase-like domain (helicase signature) of this RNA (as defined by Candresse et al., 1990) with those of other Bromoviridae, we have confirmed the relationships among genera of the Bromo-
**Short communication**

![DNA sequence of CiLRV with protein translations](image)

**Fig. 1.** The complete sequence of RNA 1 of CiLRV. The putative translation product for the single ORF is shown under the sequence.
**viridae** that we had previously observed using the polymerase signature contained on the RNA 2 of CiLRV (Ge & Scott, 1994).

The methods used were essentially those described previously (Ge & Scott, 1994; Scott & Ge, 1995). Purified CiLRV was kindly provided by Dr S. M. Garnsey (USDA, Orlando, Fla., USA). Viral RNAs were fractionated in a 2.5% low-melting point agarose gel and a cDNA was prepared to the gel-isolated RNA 1 using the primer (5' GCATCTCC 3') which we had previously found to be complementary to the 3' terminal sequence of all the RNAs of CiLRV and to several other ilarviruses. The cDNA was cloned and a clone (pCiLRV 1-1) was isolated that when sequenced contained a 600 base insert of which 190 bp shared greater than 90% similarity with the 3' untranslated region (UTR) of both RNA 2 and RNA 3. An EcoRI site separated the highly similar region from the remainder of the clone. Digestion with this enzyme was used to produce a 400 bp fragment which was labelled with 32P and used to screen a cDNA library to CiLRV that had previously been prepared by polyadenylating the RNA prior to synthesis of the cDNA. A 1.4 kb clone (pCiLRV 1-3) was obtained and sequenced. A primer (5' TTGCGCGTCGCTACAACC 3') complementary to positions 2011-2028 in the final sequence was selected from the sequence of pCiLRV 1-3 and used to synthesize cDNA extending toward the 5' end of the molecule. Synthesis and cloning using this primer yielded a clone (pCiLRV 1-13) which was about 1.6 kb in size. The 5' end of the molecule was confirmed by 5'–3' ligation of RNA and PCR (Mandl et al., 1991) and by direct sequencing of the RNA.

The DNA was sequenced using an automated DNA 373A sequencer (Applied Biosystems) and synthetic oligonucleotide primers (KS, M13, reverse M13 and p79; De Bellis et al., 1992). All clones and their subclones were sequenced completely in both directions. Nucleic acid sequences were compared and/or translated using either programs in the University of Wisconsin Genetics Computer Group (GGC) analysis package (Devereux et al., 1984) or by using Gene Jockey. Comparisons of the levels of identity between translation products of the RNA 1 (P1 protein), the N-terminal signature and the helicase signature, and corresponding features of other Bromoviridae were made using the GAP procedure of GCG. Multiple sequence alignments were made and dendograms were drawn using the PILEUP and FIGURE procedures of GCG, respectively.

The complete sequence of CiLRV RNA 1 is shown in Fig. 1. The sequence contains 3404 nt with initiation codons at positions 74–76 and 530–532. The second initiation codon is in a good context for plant translational initiation (Lutcke et al., 1987). However, selection of this codon as the initiation point of an open reading frame (ORF) would give rise to both an unusually long 5' UTR and a small putative translation product (Mr 101000) when compared to corresponding features reported for the RNA 1 of other members of the Bromoviridae. Thus, the initiation codon at 74–76 is the most probable starting point of the translation product of this RNA, even though it is in a less than optimum context for translation, having a U at the –3 position. The single ORF covers the region from positions 74 to 3229 and encodes a putative translation product of 1051 amino acids with a calculated Mr of 118339. The 5' UTR of the RNA has 73 nt while the 3' UTR was 174 nt. The 3' proximal 190 nt of CiLRV RNA 1 were shared with RNA 2 and RNA 3 of CiLRV with only slight mismatching: eight mismatches between RNA 1 and RNA 2, eight mismatches between RNA 1 and RNA 3, and five mismatches between RNA 2 and RNA 3 (Fig. 2). This 190 nt region could be folded to form a secondary structure similar to those proposed for AMV and tobacco streak virus (TSV; Koper-Zwarthoff & Bol, 1980). While not identical, the three sequences all formed structures with three loops. If aligned from the 3' terminus, the seven occurrences of the AUGC motif in this region are completely conserved with the minor sequence differences among the three RNAs occurring between motifs. However, it should be noted that the mismatching in the region between AUGC motifs 3 and 4 gives rise to an additional unique AUGC motif in RNA 2.

Comparisons between the nucleotide sequences of RNA 1, the amino acid sequence of the putative P1 protein, the N-terminal signature and the helicase signature of CiLRV with those of other Bromoviridae showed CiLRV to be most closely related to AMV. The nucleotide sequences of the two viruses shared 47.8% similarity, the amino acid sequences of the P1 proteins exhibited 37% identity, and the N-terminal and helicase
Fig. 3. Clustering of the nucleotide sequences for the RNA 1 of members of the genera \textit{Cucumovirus}, \textit{Bromovirus} and \textit{Alfamovirus} and CiLRV produced using the PILEUP procedure of the GCG package (Devereux \textit{et al.}, 1984). Sources of sequence data: AMV, Cornelissen \textit{et al.} (1983); BBMV, broad bean mottle virus, and CCMV, cowpea chlorotic mottle virus, Dzianott \\& Bujarski (1991); BMV, brome mosaic virus, Ahlquist \textit{et al.} (1984); PSV, peanut stunt virus, Karasawa \textit{et al.} (1992); TAV, tomato aspermy virus, Bernal \textit{et al.} (1991); CMV, cucumber mosaic virus, strain Fny, Rizzo \\& Palukaitis (1989); strain Q, Rezaian \textit{et al.} (1985); strain Y, Kataoka \textit{et al.} (1990).

signatures showed 46.7 and 44.7\% identity, respectively. Figures for comparisons with other members of the \textit{Bromoviridae} were typically 6–12\% less. Multiple alignments of sequences and signatures using the PILEUP procedure always generated figures in which three distinct clusters occurred (Fig. 3): the genus \textit{Cucumovirus}, the genus \textit{Bromovirus}, and AMV and CiLRV. While the relative positions of viruses within each cluster varied slightly, the arrangement into three distinct clusters persisted even when the sequence and signatures of barley stripe mosaic virus (BSMV) were added to the alignments to act as an outgroup. Comparing the N-terminal and helicase signatures of the P1 protein of CiLRV with the signatures of the members of the larger ‘alpha-like’ virus group (Candresse \textit{et al.}, 1990) produced the three clusters observed in Fig. 3, with the other viruses arranged in additional groupings (Fig. 4\textit{a}, \textit{b}). However, comparisons among the N-terminal signatures (Fig. 4\textit{a}) showed a closer relationship between raspberry bushy dwarf virus (RBDV), BSMV, AMV and CiLRV than did comparisons made using the helicase signature (Fig. 4\textit{b}).

The sequence presented here extends the list of sequences reported for ilarviruses, is the first full-length sequence described for the RNA 1 of an ilarvirus, and completes the first sequence for the entire genome of an ilarvirus. Assuming that the genome of CiLRV is typical of those of other members of the genus \textit{Ilarvirus}, then clearly the genomic organization of ilarviruses is similar to that of other genera in the \textit{Bromoviridae}, as has been previously speculated. The arrangement of the genes on the respective genomic RNAs and the size of both the putative translation products and the UTRs are in good agreement with previously reported data for the other genera within the \textit{Bromoviridae}. Both the length of the nucleotide sequence and the size of the putative protein coded for by the single ORF of CiLRV are smaller than those reported for AMV but are larger than for members of the genera \textit{Cucumovirus} and \textit{Bromovirus}.

A close relationship between ilarviruses and AMV has been proposed and to some extent supported by our earlier findings with the RNA 2 of CiLRV (Ge \& Scott, 1994). Here we provide further evidence that the two viruses are related. However, the relationship demonstrated with the RNA 1 (levels of identity for the P1 protein, N-terminal signature and helicase signature being 37.0, 46.7 and 44.7\%, respectively) appears to be slightly more distant than that shown by the P2 protein and the polymerase signature of the RNA 2 (42.3 and 62.0\% identity, respectively). In addition, both ilarviruses and AMV have repeated AUGC motifs in the 3’ UTR of the RNAs, which are important in the protein binding-genome activation phenomenon observed in both genera (Houser-Scott \textit{et al.}, 1994; Reusken \textit{et al.}, 1994). AUGC motifs also occur in the 3’ UTR of some bromoviruses and cucumoviruses but they do so as either single or double occurrences close to the 5’ end of the region. If the highly conserved repeated AUGC motifs are considered to be the ancestral state and accumulation of point mutations has led to the loss of this state, then it can be speculated that ilarviruses and AMV are closely related and perhaps ‘older’ viruses than both bromoviruses and cucumoviruses. However, it must be noted that while the sizes of the 5’ UTR of the genomic RNAs of CiLRV and AMV are similar, the 3’ UTR of the RNA 2 of CiLRV is much longer than the corresponding area of the AMV genome and in this respect the virus more closely resembles cucumber mosaic virus.

The comparisons of the N-terminal signature and helicase signature of the ‘alpha-like’ viruses (Group II, according to Candresse \textit{et al.}, 1990) confirmed the relationships between genera in the \textit{Bromoviridae} but
Fig. 4. Clusterings of the N-terminal signature (a) and the helicase signature (b) of the 'alpha-like' viruses. Sources of sequence data: BSMV RNA a, Gustafson et al. (1989); RBDV, Ziegler et al. (1992); Sindbis virus, Rice & Strauss (1981); TMV, tobacco mosaic virus; common strain (V), Goelet et al. (1982); tomato strain (O), Ohno et al. (1984); TRV, tobacco rattle virus, Hamilton et al. (1987); PEBV, pea early browning virus, Macfarlane et al. (1989); for other viruses see legend to Fig. 3.

gave different arrangements of the non-members. RBDV is much more closely linked to AMV and the ilarviruses when comparing their N-terminal signatures than when comparing the helicase signatures. It has been proposed that RBDV, with a total genome size and genomic organization similar to that of the Bromoviridae, should be included in this family despite possessing a bipartite genome (Ziegler et al., 1993), although this has yet to be approved and RBDV is currently the sole member of the genus Idaeovirus (Murphy et al., 1995). The genome of RBDV may have arisen following recombination between the RNA 1 and RNA 2 of viruses with tripartite genomes. Thus, members of the ilarviruses may be considered potential sources of the molecules that combined to form RBDV. Evidence of viral recombination exists in the tobaviruses (Robinson et al., 1987; Angenent et al., 1990) and the equine encephalitis viruses (Hahn et al., 1988), where parts of the genome of one virus are found in the genome of a second virus. Clearly, from the comparisons in this paper and preceding papers on CiLRV (Ge & Scott, 1994; Scott & Ge, 1995), no evidence for CiLRV molecules being part of RBDV exists. Expectations of such evidence are, perhaps, simplistic and any relationship between ilarviruses and RBDV may be less defined and mimic to some degree the relationship between AMV and the ilarviruses.

The RNA 3 of a number of ilarviruses has now been sequenced (apple mosaic virus, accession number U15608; prunus necrotic ringspot virus, L38823; prune dwarf virus, L28145; CiLRV, U17390; citrus variegation virus, U17389; asparagus virus 2, X86352; lilac ring mottle virus, U17391; TSV, X00435) but no close relationship between either the putative movement proteins or the coat proteins of these viruses and RBDV has been detected. However, sufficient ilarviruses remain to be sequenced so that a member of this genus may yet be identified as the direct progenitor of RBDV. Conversely, it must be pointed out that the AUGC motif, which is found repeatedly in the 3' UTR of ilarviruses and AMV, occurs only once in the 3' UTRs of RBDV.

This is technical contribution number 4089 of the South Carolina Agricultural Experiment Station.
References


(Received 23 May 1995; Accepted 29 August 1995)