The nucleotide sequence of RNA1 and RNA2 of olive latent virus 2 and its relationships in the family Bromoviridae

F. Grieco, M. Dell’Orco and G. P. Martelli

Dipartimento di Protezione delle Piante, Università degli Studi and Centro di Studio del CNR sui Virus e le Virosi delle Colture Mediterranee, Via Amendola 165/A, 70126 Bari, Italy

The complete nucleotide sequence of RNA1 and RNA2 of olive latent virus 2 (OLV-2), a virus with quasi-spherical to bacilliform particles and a non-polyadenylated tripartite ssRNA genome, was determined. RNA1 consists of 3126 nucleotides and contains a single open reading frame (ORF) coding for a polypeptide with a molecular mass of 102689 Da (p1a). RNA2 is also a monocistronic molecule, 2734 nt in length, coding for a polypeptide with a molecular mass of 90631 Da (p2a). The translation products of RNA1 and RNA2 possess the motifs proper to helicase, methyltransferase (RNA1) and RNA polymerase (RNA2), suggesting that both are involved in the replication of the viral RNA. The similarities found between OLV-2 and members of the Bromoviridae in some properties and in the sequences of all genomic products (including p1a and p2a) are strongly indicative that it belongs in this family. OLV-2, however, did not show a direct relationship with any of the current genera in the family. Rather, it revealed homologies in diverging directions with one or other of the Bromoviridae genus, thus qualifying as the possible representative of a new taxon in this family.

Introduction

Olive latent virus 2 (OLV-2), a virus isolated from symptomless olive trees in southern Italy (Savino et al., 1984), has variously shaped particles, quasi-spherical to bacilliform, a single capsid protein species and a non-polyadenylated tripartite ssRNA genome. Virions encapsidate seven RNA molecules (Grieco et al., 1992).

Two of the four major RNA species, RNA3 and RNA4, were completely sequenced (Grieco et al., 1995). RNA3 is a bicistronic molecule encoding both the movement protein and the coat protein (CP) genes. RNA4 rather than a subgenomic RNA, is a co-terminal portion of RNA3, with no messenger activity and undetermined biological significance (Grieco et al., 1995). The viral CP is apparently expressed by a non-encapsidated subgenomic RNA (Grieco et al., 1995).

As reported in the present paper, the nucleotide sequences of RNA1 and RNA2 have now been completed, providing information based on which the taxonomic status of OLV-2 is defined and its relationships in the family Bromoviridae are discussed.

Methods

■ Virus source and purification. The OLV-2 isolate was that used in previous studies (Grieco et al., 1995). The virus was propagated in Nicotiana benthamiana from which it was purified as previously described (Grieco et al., 1992). Nucleic acids were extracted from virus preparations fractionated in a sucrose gradient, analysed by electrophoresis in 1.2% low melting point agarose and visualized by staining with ethidium bromide. The bands corresponding to RNA1 and RNA2 were excised and the nucleic acid extracted (Sambrook et al., 1989). The presence of a poly(A) tail was checked by chromatography on oligo(dT) cellulose columns as described by Sambrook et al. (1989).

■ cDNA synthesis and cloning. The RNA was polyadenylated with poly(A) polymerase (BRL) following the manufacturer’s instructions. RNA was reverse-transcribed using Moloney murine leukemia virus RNase H- reverse transcriptase (Superscript, BRL) and cDNA obtained following the manufacturer’s instructions. The cDNA was ligated into SmaI-cut, dephosphorylated pUC18 and cloned in Escherichia coli DH5α.

■ DNA and RNA sequencing. Small overlapping subclones were produced from the original clones as described (Grieco et al., 1989). DNA was prepared as in Hattori & Sakaki (1986) and sequence analysis done with T7 DNA polymerase (Sequenase, United States Biochemical). All sequences except for 58 and 17 nt at the 5’ termini of RNA1 and RNA2 respectively, were determined by sequencing DNA clones and subclones.
Fig. 1. Nucleotide sequence of RNA1 of OLV-2 and amino acid sequence encoded by the open reading frame. The stop codon is marked with an asterisk.
in both orientations. The 5' ends of RNA1 and RNA2 were sequenced by
dideoxy-terminated reverse transcription (Ficht & Girard, 1990) using
two different internal primers: primer 1, consisting of nucleotides in
positions 109-128 of the RNA1 sequence; primer 2, consisting of
nucleotides in positions 48-67 of the RNA2 sequence.

Computer analysis. The sequence was assembled and analysed
using both the DNA Strider (March, 1988) and the Wisconsin Genetics
Computer Group (GCG) (Anonymous, 1994) programs. Comparisons of
proteins potentially expressed by RNA1 and RNA2 of OLV-2 with
sequences in the PIR protein database (release 47.0) were made with the
automatic electronic mail servers FASTP (Pearson & Lipman, 1988) and
BLAST (Altschul et al., 1990). The PILEUP and LINEUP programs were
used for alignment of amino acid sequences of the 103 kDa and 90 kDa
polypeptides of OLV-2 with the 1a and 2a amino acid sequences of the
following viruses: cucumber mosaic strain Q (CMV-Q), tomato
asperm (TAV) and peanut stunt (PSV) cucumoviruses; broad bean
mottle (BBMV), cowpea chlorotic mottle (CCMV) and brome mosaic
(BMV) bromoviruses; alfalfa mosaic alfamovirus (AMV); citrus rugose
leaf ilarivirus (CiRLV); the 191 kDa protein of raspberry bushy dwarf
virus (RBDV) and the 129 kDa and 87 kDa proteins of barley stripe
mosaic hordeivirus (BSMV). Sequence alignments were displayed
utilizing the PRETTYPLOT program (Anonymous, 1994). Dendrograms
were constructed with the programs SEQBOOT, PROTDIST,
NEIGHBOR and CONSENSE (Felsenstein, 1989). A multiple dataset was
generated by SEQBOOT using aligned sequences as input, distance
measure was computed utilizing PROTDIST, and trees were produced by
NEIGHBOR and CONSENSE programs. RNA secondary structures were
produced with the MFOLD and PLOTFOLD programmes from the GCG
Programs Package. The search for tRNA structure was carried out using
the 'tRNA Program' in the GLORIA package of the University of Bari.

Results

Nucleotide sequence of RNA1 and RNA2

The complete sequence of RNA1 was 3126 nt long (Fig. 1).
Except for the 58 5' terminal nucleotides the whole sequence
was determined on cDNA clones in both orientations. The 3' terminal
sequence was determined on five different clones
containing an artificial poly(A) tail. The first 100 nt at the 5' end were read directly on the RNA. The 5' non-coding region
(NCR) was 101 nt long. Most of the nucleotide sequence of RNA2 was determined from cDNA clones sequenced in both orientations, but the first 17 nt were identified by sequencing the RNA molecule. RNA2 was 2734 nt in length (Fig. 2) and had a 5' NCR 112 nt long.

The nucleotide composition of the 5' NCRs of RNA1 (36.63% A, 22.77% C, 17.83% G, 22.77% T) and RNA2 (42.86% A, 16.96% C, 13.39% G, 26.79% T) had a high A + T content, consistent with the composition of the leader region of other plant virus genomes (Gallie et al., 1987). When viral RNA1 and RNA2 were used as substrate in a sequencing reaction, terminal major 'strong stop' reverse transcription runoff products were found (not shown), as previously observed with capped viral RNAs (Ahlquist & Janda, 1984).
Comparison of the two leader sequences showed a high degree
of nucleotide identity (65%) with the 5'-terminal 60 nt almost in common (Fig. 3a), but not with the corresponding region of RNA3 (Grieco et al., 1995). A sequence partially complimentary to the 3' end of 18S ribosomal RNA was detected in both RNA1 (nt 14-31) and RNA2 (nt 15-32) sequences. Both these putative ribosome binding sites (Hagenbuchle et al., 1978) could potentially form 13 pairs (out of 18 residues) with the 18S ribosomal RNA 3' terminus. A stretch of T (9) and A (15) residues was located 10 nt upstream of the first starting
codon of RNA2 (nt 75-102). This sequence block resembled the internal poly(A) tract found in bromovirus RNA3 (Ahlquist et al., 1981).

The 3' NCR had a length of 298 and 258 nt, for RNA1 and RNA2, respectively. The 3'-proximal 250 nt of OLV-2 RNAs 1, 2 and 3 showed a high degree of similarity: RNA1/RNA2 (89%), RNA1/RNA3 (78%) and RNA2/RNA3 (82%) (Fig. 3b). These values were higher when the comparison was restricted to the last 100 nt, which could be folded to form a conserved secondary structure (Fig. 4). The free energies (at 37 °C) of the structures shown in Fig. 4 were calculated according to Zuker (1989) as -22.3 kJ (RNA1), --32.4 kJ (RNA2) and -28.4 kJ (RNA3). While not identical, the three nucleotide blocks presented a common conformation with three stem-loops. Although a similar conformation was found in ilar- and alfamoviruses (Zuidema & Jaspar, 1984), OLV-2 RNAs did not contain the AUGC motif that is thought to play an important role in binding to CP (Koper-Zwartzhoff & Bol, 1980). A computer search for 3'-terminal tRNA-like structures gave negative results.

Coding regions of RNA1 and RNA2

RNA1 contained one large open reading frame (ORF) (Fig. 1). No other ORFs encoding proteins > 10 kDa were detected on either the positive or negative strands. This ORF started at the first AUG codon located in position 102 and extended to a UGA (opal) stop codon in position 2828. The putative translation product is a polypeptide of 908 amino acids with a predicted molecular mass of 102689 Da (p1a). An analysis of the translation context of the starting codon of ORF1 showed a partial fit with the consensus sequence described by Lutcke et al. (1987), because of the presence of the G in position + 4 and the A in position -1. Another AUG codon was positioned 21 nt downstream of the former, but it was in a non-optimal translation context, lacking the G in + 4.

In OLV-2 RNA2 the first 5' terminal AUG codon, located in position 113, was followed by a 2364 nt long ORF which ended with a UGA (opal) stop codon in position 2476 (Fig. 2). The potentially encoded polypeptide was 787 amino acids in length with a molecular mass of 90631 Da (p2a). The initiation
codon of ORF2 was in a good context for plant translational
initiation (Lutcke et al., 1987), because of the A in position - 4, G in position + 4 and C in position - 5. No significant ORFs (> 10 kDa) were found on the negative strand of RNA2.
primary structure showed the presence of two distinct domains: the N-terminal, associated with RNA methyltransferase activity, and the C-terminal, containing the helicase signatures (Koonin & Dolja, 1993). Conserved sequence motifs for type I putative methyltransferase were found from aa 78–305, including motif II, which appeared to constitute part of the ‘ado met’ binding site (Mi et al., 1989). In the C-terminal half of p1a, six helicase-related motifs were found.
(from aa 626–883), including the two blocks (motifs A and B) involved in nucleotide binding (Gorbalenya et al., 1989).

Both the above signatures were compared using the GAP program with those of representative members of Bromoviridae. The pairwise alignments, shown in Table I, had high significance scores (SD > 10) calculated according to Koonin (1991). Thus, a clear-cut relationship among the C- and N-terminal domains of OLV-2 p1a and the comparable protein domains of viruses in the family Bromoviridae was established.

The amino acid sequence of OLV-2 p2a was compared with those of other viral proteins in the PIR database and significant similarities were found with the corresponding polypeptides of members of Bromoviridae. A domain formed by eight RNA polymerase-related motifs (Koonin & Dolja, 1993) was clearly revealed on the p2a primary structure and located from aa 433–665. The polymerase signature (xxGxxxxTxxxNx and xxxGDDxxx; Argos, 1988) was recognized in the V and VI motifs, suggesting p2a to be a protein with an RNA-dependent RNA polymerase activity (RdRp).

Pairwise alignments of the RNA polymerase domain of p2a with those of members of the Bromoviridae (Table I) indicated that OLV-2 was related to the other tripartite-genome viruses. A multiple alignment was made using progressive pairwise alignments (PILEUP program) between the methyltransferase and helicase signatures of OLV-2 with the signatures of some members of the Bromoviridae. The two alignments thus
Table 1. Percentage amino acid sequence similarity between methyltransferase, helicase and RNA polymerase signatures of OLV-2 and the corresponding regions of a number of tripartite-genome viruses

<table>
<thead>
<tr>
<th></th>
<th>Methyltransferase</th>
<th>Helicase</th>
<th>Polymerase</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMV</td>
<td>66%</td>
<td>59%</td>
<td>62%</td>
</tr>
<tr>
<td>CCMV</td>
<td>63%</td>
<td>58%</td>
<td>61%</td>
</tr>
<tr>
<td>CMV</td>
<td>57%</td>
<td>57%</td>
<td>59%</td>
</tr>
<tr>
<td>TAV</td>
<td>53%</td>
<td>54%</td>
<td>63%</td>
</tr>
<tr>
<td>AMV</td>
<td>53%</td>
<td>59%</td>
<td>61%</td>
</tr>
<tr>
<td>CiLRV</td>
<td>52%</td>
<td>51%</td>
<td>63%</td>
</tr>
<tr>
<td>RBDV</td>
<td>60%</td>
<td>61%</td>
<td>60%</td>
</tr>
<tr>
<td>BSMV</td>
<td>58%</td>
<td>55%</td>
<td>53%</td>
</tr>
</tbody>
</table>

obtained were used to generate tentative phylogenetic trees (Fig. 5a, b). The dendrogram plotted by comparing the methyltransferase signature (Fig. 5a) generated a tree in which three distinct clusters occurred: (i) the genus Bromovirus, (ii) the genus Cucumovirus and (iii) AMV, CiLRV and RBDV. The OLV-2 N-terminal domain did not cluster with any of these groups. The comparison of helicase signatures (Fig. 5b) confirmed the relationship with genera in the Bromoviridae, but revealed that OLV-2 is closer to RBDV.

The RdRp signatures of OLV-2 and other tripartite-genome viruses were aligned and a tentative phylogenetic dendrogram plotted from this data (Fig. 6). Here again, species of the genera Cucumovirus and Bromovirus clustered, as did AMV and CiLRV, whereas OLV-2 remained separate from the above and from RBDV.

Discussion

The determination of the nucleotide sequence of RNA1 and RNA2 completes the sequence of the OLV-2 genome, which consists of three RNA species encoding four proteins (Fig. 7). As previously discussed (Grieco et al., 1995) OLV-2 possesses properties typical of members of the Bromoviridae, i.e. RNA3 is bicistronic and codes for two polypeptides (movement protein and CP), with the same genomic position and functions of those of known members of this family. However, two major
features distinguish OLV-2 from the existing *Bromoviridae* genera: (i) the unusual properties of RNA4 (2078 nt), which lacks coding capacity and does not have a determined biological function; (ii) the presence of a non-encapsidated subgenomic RNA (1042 nt), probably responsible for CP expression (Grieco *et al.*, 1995).

The results of the present study indicate that RNA1 and RNA2 also share properties with the equivalent genomic portions of viruses in the *Bromoviridae* because: (i) they are remarkably similar in size and structural organization; (ii) both molecules may be capped at the 5' terminus; (iii) both RNAs code for proteins that have a size and proposed function similar to the replication-related proteins expressed by members of this family.

The 5' NCRs of OLV-2 RNA1 and RNA2 are closely conserved (Fig. 3a). They may be both implicated in replication and in initiation of translation, because of their high A+U content, as for BMV (Ahlquist *et al.*, 1981), and because they show a sequence block that could act as a ribosome binding site. The 5' NCR of RNA2 contains a stretch of T (9) and A (15) that is absent in the 5' terminal region of RNA1 and RNA3, suggesting that it may not be essential for replication of both these RNAs. Poly(A) stretches have been found in the NCRs of several plant viruses such as BMV (Ahlquist *et al.*, 1981), BSMV (Agranovsky *et al.*, 1982) and AMV (Langereis *et al.*, 1986). The presence of these nucleotide blocks could be considered as an example of the apparent flexibility which the NCRs of RNA plant viruses are allowed. Terminal NCRs within genera in *Bromoviridae* are different in size, structural characteristics and function (Symons, 1985), indicating that they are subjected to considerable selective pressure.

Genomic RNAs of tripartite plant viruses often have a highly conserved 3'-terminal region. In bromo-, cucumo-, ilar- and alfamoviruses, this genomic fragment retains almost identical sequence for the terminal 200–40 nt (Symons, 1985). A similar alignment between OLV-2 RNAs 1, 2 and 3 showed sequence identity in the 3'-terminal 19 nt and high similarity in the 239 nt upstream of this region (Fig. 3b). Notwithstanding this incomplete identity, the 3'-terminal 100 nt of the three OLV-2 RNA species could be folded in a 'stem–loops' secondary structure, similar to those in AMV, RBDV and ilarviruses, but not in bromo- and cucumoviruses.

However, the OLV-2 3' terminus has function and structure that differ from those of members of the family *Bromoviridae* as it lacks the tRNA-like terminal structure (as in cucumo- and bromoviruses) and it may not have a role in activating replication (as in alfamo- and ilarviruses). This last finding is supported by several lines of evidence: (i) OLV-2 does not require the addition of CP for infectivity in the absence of the subgenomic RNA (Grieco *et al.*, 1995); (ii) its sequence does not contain the AUGC motif present in AMV and TSV thought to have a role in CP genome binding (Houser-Scott *et al.*, 1994; Koper-Zwarthoff & Bol, 1980); (iii) OLV-2 CP also lacks the zinc-finger type domain, functionally important sequence in alfamo- and ilarviruses (Sanchez-Navarro & Pallas, 1994; Sehnke *et al.*, 1989).

OLV-2-coded movement protein and CP revealed homologies in two quite distinct directions: the former with the comparable products of cucumo- and bromovirus, the latter with the CP of alfamovirus (Grieco *et al.*, 1995). Computer-assisted analysis of OLV-2 p1a and p2a disclosed further relationships with members of the family *Bromoviridae*. These were in the RNA polymerase domain of p2a and in methyltransferase and helicase signatures of p1a. However, in no case did OLV-2 group with the clusters comprising bromovirus, cucumovirus and the couple AMV and CiLRV. Interestingly, the comparison of helicase domains showed a relationship with RBDV, which adds strength to the proposed
inclusion of this virus in the family Bromoviridae (Ziegler et al.,
1993; Scott & Ge, 1995).

In conclusion, the similarities found between OLV-2 and
members of the Bromoviridae in the sequence of all genomic
products, including those having major phylogenetic implic-
ations like the replication-related proteins (Koonin, 1991),
warrant inclusion of this virus in the family Bromoviridae.
However, OLV-2 does not have a complete straightforward
relationship with any current genera in the family. Rather,
this virus shows homologies in diverging directions with one genus
or another, thus substantiating the suggestion (Grieco et al.,
1995) that it may qualify as the representative of a novel taxon
in the family Bromoviridae.

References

3' extremity of barley stripe mosaic virus RNA: evidence for internal
poly(A) and a 3'-terminal tRNA-like structure. Virology 106, 51-59.

tion of the complete brome mosaic virus genome. Molecular and Cellular
Biology 4, 2876-2882.

sequence of brome mosaic virus RNA3. Journal of Molecular Biology

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J.
215, 403-410.

Anonymous (1994). Program Manual for the Wisconsin Package,
version 8, September 1994. Genetics Computer Group, Madison, Wis.,
USA.

Research 16, 9909-9916.

3.3). Cladistics 5, 164-166.

Fichot, O. & Girard, M. (1990). An improved method for sequencing of
RNA templates. Nucleic Acids Research 18, 6162.


Gorbulevsky, A. E., Blonov, V. M., Donchenko, A. P. & Koonin, E. V.


Grieco, F., Martelli, G. P. & Savino, V. (1995). Nucleotide sequence of

Hagenbuchle, O., Santer, M., Steitz, J. A. & Mants, R. I. (1978). Conservation of the primary structure at the 3' end of 18S rRNA from


Nucleotide sequence and structural determinants of specific binding of
coop protein or coat protein peptides to the 3' untranslated region of


Koonin, E. V. & Dolja, V. V. (1993). Evolution and taxonomy of
positive-strand RNA viruses: implications of comparative analysis of
amino acid sequence. Critical Reviews in Biochemistry and Molecular Biology 28, 375-430.

putative recognition site for coat protein in the RNAs of alfalfa mosaic

Langereis, K., Mugnier, M., Cornelissen, B. J. C., Pinck, L. & Bol, J. F.
(1986). Variable repeats and poly(A)-stretches in the leader sequences of


Marck, C. (1988). "DNA Strider": a "C" programme for the fast
analysis of DNA and protein sequences on the Apple Macintosh family

Association of sindbis virus RNA methyltransferase activity with the non


Sanchez-Navarro, J. A. & Pallas, V. (1994). Nucleotide sequence of

latent virus 2, a newly recognized virus with different shaped particles. Virology Journal 3, 256-268.

Scott, S. W. & Ge, X. (1995). The nucleotide sequence of citrus leaf


Schneekammer, J. & Saunder, V. (1994). Nucleotide sequence of

latent virus 2, a newly recognized virus with different shaped particles. Proceedings of the 5th Congress of the Mediterranean Phytopathological
Union, Cairo 1984, pp. 24-26.

Scott, S. W. & Ge, X. (1995). The nucleotide sequence of citrus leaf

(1989). A “zinc-finger”-type binding domain in tobacco streak virus

1. Polyhedral Virions with Tripartite Genomes, pp. 57-81. Edited by R. I. B.

tripartite-genomed viruses in the family Bromoviridae. Archives of Virology 131, 483-488.

molecule. Science 244, 48-52.

cocoon protein binding sites of the genomic RNAs of alfalfa mosaic and