Nucleotide sequence of the coat protein genes of strawberry latent ringspot virus: lack of homology to the nepoviruses and comoviruses

K. R. Everett,1,2* K. S. Milne2 and R. L. S. Forster1

1 The Horticulture and Food Research Institute of New Zealand Ltd, Mt Albert Research Centre, 120 Mt Albert Road, Mt Albert, Private Bag 92169, Auckland, and 2 Department of Plant Science, Massey University, Palmerston North, New Zealand

The sequence of the 3'-terminal 2424 nucleotides of RNA-2 of the flowering cherry strain of strawberry latent ringspot virus (SLRV) was determined from cDNA clones. The sequence contains a reading frame in the virus-sense strand of 2070 nucleotides, a 3' untranslated region of 552 nucleotides and a 3'-terminal poly(A) tract. The positions of the two coat proteins of SLRV within the reading frame were determined from sequence data obtained by N-terminal sequencing using Edman degradation. The larger coat protein with an Mr of 43K is located 5' of the smaller coat protein of 27K, and the two proteins are apparently cleaved at a Ser-Gly bond. Although there are numerous similarities between SLRV and the nepoviruses and comoviruses, there is no significant homology between the SLRV coat proteins and the coat proteins of either group. Furthermore, the hydropathy profiles of the SLRV coat proteins are unlike those of either group. No comparisons could be made with the fabaviruses owing to lack of sequencing information. This lack of homology suggests that SLRV is more distantly related to the nepoviruses and comoviruses than has been considered previously.

Strawberry latent ringspot virus (SLRV) has a wide host range, is seedborne and is spread by nematodes of the genus *Xiphinema* (Murant, 1974). Tubule-like structures are produced in infected cells (Hicks, 1985) which may facilitate cell-to-cell translocation through plasmodesmata (Francki et al., 1985). Purified preparations of SLRV contain 30 nm hexagonal particles which sediment in sucrose density gradients as two components of approximately 55S and 130S (Murant, 1974). The 55S 'top' component is composed of RNA-free coat protein shells and the 130S 'bottom' component contains viral RNA. Some isolates of SLRV also have a 'middle' component of 95S to 99S (Gallitelli et al., 1982). The genome of SLRV consists of two RNA species of 9.0 and 5-15 kb (Everett et al., 1994) that are translated as polyproteins (Hellen et al., 1991). The RNAs are polyadenylated at their 3' termini (Mayo et al., 1979), and contain a protein covalently bound to their 5' termini (Mayo et al., 1982). Some isolates of SLRV have a 1-2 kb satellite RNA species (Mayo et al., 1974, 1982; Gallitelli et al., 1982).

The above characteristics of SLRV are shared by the nepoviruses (Harrison & Murant, 1977), and for this reason SLRV has been tentatively placed in the nepovirus group. However, unlike all other nepoviruses, which have a single species of coat protein, of approximate Mr 55K, SLRV has two coat proteins of 44K and 29K (Mayo et al., 1974). In this respect, SLRV more closely resembles two other related virus groups, the comoviruses which have two coat proteins of 41K and of 22K to 24K (MacFarlane et al., 1991; Chen & Bruening, 1992; van Wezenbeek et al., 1983; Goldbach & van Kammen, 1985), and the fabavirus broad bean wilt virus which has coat proteins of 42K and 26K (Taylor & Stubbs, 1972). This report describes the cloning and sequencing of the SLRV coat protein genes and comparison of the derived amino acid sequences with those of the coat proteins of nepoviruses and comoviruses.

The SLRV strain was obtained from flowering cherry (Everett et al., 1993) and was maintained in cucumber and purified as described (Everett et al., 1994). SLRV from flowering cherry reacted homologically with antiserum to SLRV obtained from the Scottish Crop Research Institute, Dundee, U.K. RNA was extracted from purified virus according to Everett et al. (1994).
First-strand cDNA was synthesized from purified SLRV RNA using avian myeloblastosis virus (AMV) reverse transcriptase (Life Sciences) and oligo(dT)\(_{12-18}\) as described by Sambrook et al. (1989). Second-strand synthesis of cDNA clones to the 3' region was as described by Gubler & Hoffman (1983). The cDNA was blunt-ended with T4 DNA polymerase and cloned into plasmid pUC19.

One cDNA clone, pB6, with an insert of 2278 nucleotides (nt), was selected for sequencing. This clone was shown by Northern blot analysis to be specific to RNA-2. Clone pB6 was found to contain a 3'-terminal poly(A) tract and a single continuous reading frame beginning at the 5' terminus of the insert and terminating 552 nt from the poly(A) tract. In addition, the 3' termini of five further clones were sequenced to verify the position of the termination codon. To determine whether the coat protein genes were present in this reading frame, the N-terminal amino acid sequences of the two coat proteins (underlined in Fig. 1) were determined from purified virus by Edman degradation using an Applied Biosystems 470A automatic protein sequencer (Hewick et al., 1981). An amino acid stretch identical to that obtained from the smaller coat protein was recognized within the reading frame of clone pB6, 235 amino acid residues from the termination codon. However, the N-terminal amino acid sequence of the larger coat protein was not present, indicating that clone pB6 did not contain the complete coding region for this protein.

To obtain the 5'-terminal sequence of the larger coat protein, cDNA was synthesized using AMV reverse transcriptase and the primer 5' ATAAACCAGACAC-CCTGCA 3' (complementary to nucleotides 204 to 222 of clone pB6). The cDNA was amplified by PCR using this primer and the primer 5' AGCTGGATCCGGN(C/T)- TNCA(C/T)GA(A/G)GA(C/T)(C/T)TNNGTNCC 3' which was derived from the N-terminal amino acid sequence of the larger coat protein. The cDNA–RNA hybrid was heated for 5 min at 94°C and then amplified using Taq polymerase (Amersham), and 30 cycles of 94°C for 1 min, 42°C for 1 min, 50°C for 1 min and 72°C for 2 min. The PCR product was cloned into plasmid pUC19 and sequenced using the method of Sanger et al. (1977) in both directions. Three clones chosen for sequencing contained identical inserts of 368 nt. These clones overlapped clone pB6 by 222 nt and the nucleotide sequences were identical to the sequence of clone B6 in the region of overlap.

The combined nucleotide sequences of clone pB6 and the cloned PCR product are shown in Fig. 1. Because of the degenerate nature of the oligonucleotide used for cDNA synthesis of the 5'-terminal region of the gene encoding the larger coat protein, it is possible that the nucleotide sequence of the cloned PCR product may differ from the SLRV sequence in the positions of the redundancies. Therefore the 22 nt derived from this primer are shown in lower case.

Excluding a 3'-terminal tract of 19 A residues in clone pB6, the cloned regions of pB6 and pPCR3 span 2424 nt of SLRV RNA-2. The cloned region contains a single reading frame beginning at a Gly residue and terminating at an ochre stop codon at nt 1870. Alignment of the amino acid sequences determined by Edman degradation within the reading frame shows that the smaller SLRV coat protein is encoded 3' of the larger coat protein. This is similar to the order of coat proteins of comoviruses. The larger coat protein of SLRV has 388 amino acids and an M\(_r\) of 43K, and the smaller 235 amino acid protein has an M\(_r\) of 27K. These values are similar to the reported sizes of the coat proteins of SLRV of M\(_r\) 44K and 29K (Murant, 1974) and 45K and 30K (Everett et al., 1994) determined from purified virus particles on 6 to 10% and 12.5% polyacrylamide gels respectively.

The N-terminal sequence of the 43K protein determined by Edman degradation shown underlined in Fig. 1 differs at amino acid residue 2 from that deduced from the nucleotide sequence, presumably indicating that an inexact oligonucleotide primer molecule was incorporated during PCR. The six other amino acid residues in the region of overlap with this redundant primer and a further eight residues 3' of the primer were identical to those determined by Edman degradation. The N-terminal sequence of the 27K protein determined by Edman degradation was identical to the sequence deduced from the nucleotide sequence, indicating that the 27K protein is probably cleaved from the SLRV RNA-2 polyprotein at a Ser–Gly bond. Cleavage sites for the coat protein genes of nepoviruses and comoviruses occur at Arg–Gly [grapevine fanleaf nepovirus (GFLV), Serghini et al., 1990; arabis mosaic nepovirus (ArMV), Bertioni et al., 1991], Arg–Ala [Hungarian grapevine chrome mosaic nepovirus (GCMV), Brault et al., 1989], Lys–Ala [tomato black ring nepovirus (TBRV), Demangeat et al., 1992], Cys–Ala, [raspberry ringspot nepovirus (RRV), Blok et al., 1992], Gln–Met and Gln–Gly [cowpea mosaic comovirus (CPMV), van Wezenbeek et al., 1983], Gln–Ser and Gln–Ser [cowpea severe mosaic virus (CPSMV), Chen & Bruening, 1992, Gln–Thr and Gln–Gly [red clover mottle virus (RCMV), Shanks et al., 1986] and Gln–Met and Gln–Ser [bean pod mottle virus (BPMV), MacFarlane et al., 1991]. A Ser–Gly bond has not been reported previously as a cleavage site for a plant virus-encoded polyprotein.

To date, the nucleotide sequences of the coat proteins of six nepoviruses and four comoviruses have been determined. Rott et al. (1991) have aligned the amino acid sequences of four nepoviruses [tomato ringspot nepovirus (TomRSV), TBRV, GCMV and GFLV] and
Short communication

1823

Fig. 1. Nucleotide sequence of cDNA clones to the 3'-terminal 2424 nt of SLRV RNA-2 and, in lines below the nucleotide sequence, the deduced amino acid sequence. DNA sequences were determined by dideoxynucleotide chain termination (Sanger et al., 1977) using modified T7 DNA polymerase (Sequenase). All regions were sequenced completely in both directions using overlapping subclones generated by restriction digestion. Additional sequence information was obtained using the oligonucleotides 5' CCATCTGGGACCAAGGATA 3' and 5' CAGTCCTCAAGGCCAAGA 3', complementary to nt 1912 to 1929 and nt 1083 to 1100, respectively, and the oligonucleotide 5' TAGAAGGAGCAGGAGTGGY, corresponding to nt 1817 to 1834. The underlined amino acids correspond to the N terminus of the coat protein genes as determined from Edman degradation. The symbol • indicates the presumed cleavage site of the SLRV RNA-2 43K and 27K proteins. Nucleotides in lower case are from a PCR primer containing redundancies.
shown that 38 invariant amino acids are shared by all four viruses. Alignment of the coat protein sequences of two additional nepoviruses, ArMV (Bertioli et al., 1991) and RRV (Blok et al., 1992) with the line-up of Rott et al. (1991) shows that 35 and 28, respectively, of the invariant amino acids are also shared with these two viruses. Among the conserved amino acids are two motifs, (F/L)DA(Y/F)X(R/K)(I/Y) and (F/M)XFYGX(S/T), located 435 to 486 and 11 to 62 amino acids from the C terminus, respectively. In contrast to the nepoviruses, neither of the coat proteins show little similarity to those of the coat proteins of SLRV contain these conserved amino acids and motifs. Two somewhat similar motifs, SDAYRY and LTDYGQS, are found in the 43K SLRV coat protein, located 105 and 59 amino acids from the C terminus, respectively.

Chen & Bruening (1992) have aligned the coat proteins of four comoviruses, BPMV, CPMV, CPMV and RCMV. These share 35% amino acid identity in the larger coat protein sequences, and 22% identity in the smaller coat protein sequences. Furthermore the hydropathy profiles of the larger coat proteins of the comoviruses are similar, as are the profiles of the smaller proteins, although to a lesser degree (Chen & Bruening, 1992). However there is no significant amino acid homology between coat proteins of the comoviruses and either SLRV coat protein detectable by using the Devereux et al. (1984) GCG Wisconsin software package programs GAP and BESTFIT in a VAX 3100 computer, and the hydropathy profiles of the SLRV 43K and 27K proteins show little similarity to those of the coat proteins of the comoviruses (data not shown).

Excluding the 3' poly(A) tract, the 3' untranslated region (UTR) of SLRV RNA-2 is 552 nt and contains a putative AAUAAA polyadenylation signal separated from the poly(A) tail by 62 nt. Sequences in the 3' region of the polyprotein gene and 3' UTR were confirmed using five additional SLRV RNA-2 clones (pAl, pB10, pC4, pD8 and pE5).

Cooper et al. (1992) reported that a U.K. strain of SLRV has a 3' UTR of 350 nt; however these authors presented no sequence information to allow direct comparisons between the two SLRV strains. Although most of the nepoviruses and comoviruses have 3' UTRs shorter than that of SLRV RNA-2 [nepoviruses, TBRV, 301 nt (Meyer et al., 1986), GCMV, 251 nt (Braith et al., 1989), GFLV, 212 nt (Serghini et al., 1990) and RRV, 397 nt (Blok et al., 1992); comoviruses, RCMV, 263 nt (Shanks et al., 1986), CPSMV, 471 nt (Chen & Bruening, 1992), CPMV, 179 nt (van Wezenbeek et al., 1983), BPMV, 455 nt (MacFarlane et al., 1991)], TomRSV (Rott et al., 1991) has a significantly longer 3' UTR (1550 nt).

Serghini et al. (1990) have reported four regions of nucleotide sequence homology in the 3' UTR of RNA-2 of the nepoviruses, GFLV, TBRV and GCMV, three of which are shared with CPMV. Within the first of these regions is a stretch of 10 nt, 5' UUUUGUGUGU 3', found in GFLV that is also found in the 3' UTR of the nepovirus ArMV (Bertioli et al., 1991), the comovirus RCMV (Chen & Bruening, 1992) and in the 3' UTR of SLRV RNA-2. However, this sequence is not found in the 3' UTR of other nepoviruses and comoviruses. The other regions of homology reported by Serghini et al. (1990) are not found in the 3' UTR of SLRV, and also are not found in the more recently sequenced nepoviruses and comoviruses, raising doubts about their significance.

In conclusion, we have identified the SLRV coat protein genes in the 3' region of SLRV RNA-2. SLRV shares numerous obvious similarities with the nepoviruses, comoviruses and fabaviruses. However, the conservation of amino acid motifs among the nepoviruses and comoviruses, and the lack of any significant homology between the coat proteins of SLRV and these viruses, suggest that SLRV may represent a distinct virus group. Comparison with sequence data of fabaviruses is needed to confirm this conclusion.

The authors thank Dr D. Christie, University of Auckland, for the protein sequencing.

References


(Received 27 September 1993; Accepted 4 January 1994)