Nucleotide Sequence of Potato Leafroll Luteovirus RNA

By M. A. MAYO,* D. J. ROBINSON, C. A. JOLLY AND L. HYMAN
Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, U.K.

(Accepted 23 January 1989)

SUMMARY

A sequence of 5987 nucleotides is reported for the RNA of potato leafroll luteovirus (PLRV). The sequence contains six large open reading frames, and non-coding regions of 174 nucleotides at the 5' end, 141 nucleotides at the 3' end and 197 nucleotides between two large blocks of coding sequences. The 5' coding region encodes two polypeptides of 28000 (28K) and 70K which overlap in different reading frames and circumstantial evidence suggests that the third open reading frame in the 5' block is translated by frameshift readthrough near the end of the 70K polypeptide to give a 118K polypeptide. The C-terminal part of the 118K protein contains the consensus sequence for RNA-dependent RNA polymerases. In vitro translation of PLRV RNA resulted in the synthesis mainly of 28K and 70K polypeptides and the largest product made was about 125K; these sizes are similar to those predicted for the translation products of the 5' block of coding sequence. The 3' block of coding sequence codes for three polypeptides: a 23K coat protein, a 17K polypeptide which is encoded in a different frame, and a 53K polypeptide which immediately follows the coat protein coding sequence, and is in the same reading frame. Circumstantial evidence suggests that the 53K polypeptide is translated by readthrough of the amber termination codon of the coat protein gene. The amino acid sequences encoded by the 3' block of coding sequence show many similarities with analogous polypeptides translated from the nucleotide sequences of RNA of barley yellow dwarf virus, PAV strain (BYDV) and, in particular, beet western yellows virus (BWYV). The 118K polypeptide has some similarities with the putative polymerase of southern bean mosaic virus and much more extensive similarities with the corresponding BWYV polypeptide but almost none with that of BYDV. In contrast, the amino acid sequence of the 28K polypeptide is not like that of proteins of the other luteoviruses or of viruses in other groups. The nucleotide sequences reported will appear in the EMBL, GenBank and DDBJ databases under the accession number X14600.

INTRODUCTION

Potato leafroll luteovirus (PLRV) causes an important disease of potato crops world-wide and is responsible for large crop losses. Like other luteoviruses, PLRV is transmitted in a persistent manner by aphids and its isometric particles are largely confined to the phloem of infected plants (Harrison, 1984). PLRV particles contain an $M_r \approx 10^6$ RNA in a protein coat consisting of $M_r \approx 26$K subunits (Takanami & Kubo, 1979; Rowhani & Stace-Smith, 1979). PLRV RNA has messenger activity in vitro, is not polyadenylated and is bound to a protein with an $M_r$ of approx. 7K (VPg) which may be analogous to the genome-linked proteins of other viruses (Mayo et al., 1982). Subgenomic RNA has been found in infected cells (Mayo et al., 1984) but little more is known about the molecular biology of the virus. In this paper we report the nucleotide sequence of PLRV RNA which has been determined as a first step in studies aimed at linking molecular properties of PLRV with its biological behaviour. PLRV is related serologically to several other luteoviruses including beet western yellows (BWYV) and barley yellow dwarf (BYDV) viruses (Casper, 1988; Waterhouse et al., 1988). Recently the nucleotide sequences of RNA from BYDV (PAV isolate) (Miller et al., 1988a) and from BWYV (Veidt et al., 1988) have been published. The relatedness of PLRV to these viruses, as assessed by
comparisons of the amino acid sequences deduced for the predicted translation products (including the coat proteins) of each virus, is discussed.

**METHODS**

**Virus propagation and purification.** The PLRV isolate used was derived from strain 1, a field isolate (Tamada et al., 1984) obtained from infected tubers of the potato variety Cara grown in Scotland. The isolate was maintained in potato tubers and virus particles were extracted from infected potato shoots and leaves and purified as described by Tamada & Harrison (1980) except that 5% Celluclast (Novo Enzyme Products) was used in place of Driselase to macerate the vascular tissues (Waterhouse & Helms, 1984).

**Extraction and analysis of RNA.** Virus particles were sedimented from sucrose gradient fractions by high-speed centrifugation and RNA was extracted from the pellets as described by Mayo et al. (1982). The Mr of the RNA was estimated by electrophoresis of glyoxal-treated RNA in agarose gels as described by Murant et al. (1981). Mr markers were RNA from particles of brome mosaic and tobacco mosaic viruses, and *Escherichia coli* ribosomal RNA.

**Molecular cloning.** In some experiments RNA was polyadenylated in reaction mixtures containing 30 mM-Tris-HCl pH 7.9, 0.25 mM-NaCl, 10 mM-MgCl₂, 2.5 mM-MnCl₂, 40 µg bovine serum albumin, 0.8 mM-ATP, 5 µg PLRV RNA and 2.5 units poly(A) polymerase in 100 µl for 15 min at 37 °C, recovered and annealed to oligo(dT) as described by Maniatis et al. (1982). In other experiments RNA was annealed to synthetic oligonucleotide primers (Applied Biosystems). These were 5' TACTTCTTTCAGGAATGTT 3'(A), 5' CCGCGTTTCTTTGTTTGGGC 3' and 5' GGGTTGTCTCTCATATA 3'(B) which are complementary to nucleotides 641 to 658, 1954 to 1973 and 4488 to 4506, respectively, in the sequence shown in Fig. 1. Double-stranded DNA was synthesized by reverse transcription followed by second-strand DNA synthesis and S1 nuclease treatment (Maniatis et al., 1982) for the first cloning or, in subsequent experiments, by the method of Gubler & Hoffmann (1983) using a commercial kit (Amersham). Recombinant plasmids were made either by the addition of 3' oligo(dC) to the cDNA followed by annealing with PstI-cut and oligo(dG)-tailed pBR322 or pUC9, by the addition of EcoRI linkers and ligation into EcoRI-cut pUC19, or by blunt-end ligation into SmaI-cut pUC19. The procedures were essentially those described by Maniatis et al. (1982) or as recommended by the suppliers of the enzymes used. Recombinant pBR322 was used to transform *E. coli* strain DH 1 made competent by treatment with CaCl₂, RbCl and MnCl₂ (C. Higgins, personal communication) and strains JM101 and DH5x (Bethesda Research Laboratories) were hosts for recombinant pUC plasmids. Candidate clones (ampicillin-sensitive pBR322 or lac-negative pUC) were tested by colony blotting (Maniatis et al., 1982) either with 32P-labelled cDNA to PLRV RNA made by random priming (Taylor et al., 1976) or with PLRV-specific sequences isolated from recombinant M13 bacteriophage DNA labelled as described by Barker et al. (1985).

**Subcloning and sequencing.** Plasmids were prepared by alkaline lysis (Birnboim & Doly, 1979; Maniatis et al., 1982) and analysed for restriction sites by digestion and electrophoresis in polyacrylamide gels. Appropriate fragments of insert DNA were recovered by electrophoretion and spermine precipitation. DNA fragments were ligated into M13mp10 or M13mp11 replicative form (RF) DNA and sequenced by dideoxynucleotide chain termination (Sanger et al., 1977) and electrophoresis in buffer gradient 6% gels (Biggin et al., 1983). Deoxy-7-deazaGTP was used in place of dGTP when Klenow polymerase was used (Mizusawa et al., 1986). In some experiments modified T4 polymerase (Sequenase; Cambridge Bioscience) was used. Sequences were assembled using the programs DBUTIL and DBAUTO (Staden, 1982a) and analysed by ANALYSEQ (Staden, 1984), DIAGON (Staden, 1982b) or GAP (Devereux et al., 1984).

**In vitro translation.** Translations were essentially as described by Mayo & Reddy (1985). RNA at 0.1 mg/ml was translated in a wheatgerm extract prepared according to Marcu & Dudock (1974). Polypeptides were labelled with [35S]methionine and analysed by electrophoresis in 10% polyacrylamide gels and fluorography. Marker proteins were β-galactosidase (Mr 116K), α-phosphorylase (97K), bovine serum albumin (66K), ovalbumin (45K) and carbonic anhydrase (29K) (Sigma).

**RESULTS**

**Molecular cloning**

Two preparations of RNA were used to make cDNA. Both preparations were polyadenylated and copied using an oligo(dT) primer but the second was also copied after priming separately with each of three synthetic oligonucleotides. Cloned cDNA from the first preparation represented about 80% of the final sequence. Most of the cloned DNA fragments had oligo(dA/dT) at one end but all but one lacked the 3' terminus of the final sequence, which showed that fragmented RNA had been polyadenylated and copied. Cloned cDNA from the second RNA preparation represented all but 65 nucleotides of the final sequence and included
Sequence of PLRV RNA

1039

---

**Sequence:**

```
---
```

---

---

---

---

---

---

---

---

---

---

---
three clones, derived using primer A, that had DNA inserts that terminated with identical 5' ends, and three clones, derived using oligo(dT) to prime polyadenylated RNA, that had DNA inserts that had identical sequences immediately 5' of the 3'-terminal oligo(dA/dT) sequence. All of the sequence was determined from DNA in both orientations and 97% was determined using DNA from more than one clone. The average was about eight determinations of each position.

Nucleotide sequence

The complete sequence is shown in Fig. 1. The 5987 nucleotide length is close to the value of 6 kb estimated by Rowhani & Stace-Smith (1979) by sedimentation of formaldehyde-denatured RNA and by us using electrophoresis of denatured RNA in agarose, and is also close to the value of 6±1 kb obtained by restriction mapping of cDNA (Prill et al., 1988). The putative 3'-terminal sequence is coterminal with that obtained by Prill et al. (1988), who used methods similar to ours, and differs from it only in that our sequence has U at position 5925 in place of C. However, because we used polyadenylated RNA, we cannot exclude the possibility that the sequence actually terminates in one or more A residues. Also, the method used to make the cDNA can yield copies that lack the 5'-terminal nucleotides of the template (D'Alessio & Gerard, 1988) and we cannot exclude this possibility, although finding three exactly coterminal clones among the 19 clones derived using primer A suggests that this was not so. Moreover, as in sequencing studies with other luteovirus RNAs (Miller et al., 1988a; Veidt et al., 1988), it is not possible to exclude the possibility that the VPg of PLRV, which is presumably at the 5' end of the RNA (Mayo et al., 1982), prevents reverse transcriptase from copying the last nucleotide(s). Thus the 5' end shown in Fig. 1 may not be the extreme end of the molecule.

The base composition of PLRV RNA (23% U, 25% C, 28% A and 24% G) is the same as that of BWYV RNA and similar to that of BYDV RNA (22% U, 24% C, 29.5% A and 24.5% G). Comparison of the sequence shown in Fig. 1 with the restriction endonuclease maps derived by Prill et al. (1988) and Smith et al. (1988) shows that most of the restriction sites are common to all three sequences, suggesting that there is little sequence difference between the Scottish, German and American isolates studied. Of the 28 sites described by Prill et al., the EcoRI site nearest the 5' end and the XhoI site nearest the 3' end are not in the sequence in Fig. 1 and eight sites in our sequence are not in the map shown by Prill et al. All the 20 sites described by Smith et al. (1988) are in the sequence shown in Fig. 1, but their positions are all about 100 nucleotides nearer the 5' end than they are in Fig. 1. The PstI site 278 nucleotides from the 5' end of the sequence in Fig. 1 is not in the map shown by Smith et al.

Open reading frames

Six substantial (> 300 nucleotides) open reading frames (ORFs) are present in PLRV RNA. Their positions are shown in Fig. 2 and the amino acid sequences of their predicted translation products are shown in Fig. 1. The frames are shown between the first methionine residue and the first termination codon, except for ORF 2b and ORF 5 which are shown as the regions between termination codons. ORF 2b is probably translated by frameshift readthrough from ORF 2a (see below). Several small ORFs are present but none corresponds in position to any in BWYV RNA and may therefore not be expressed in vivo.

The arrangement of the ORFs in PLRV RNA (Fig. 2) is very similar to that deduced for the ORFs in BWYV RNA (Veidt et al., 1988) and the BWYV ORFs are similar in size to those of PLRV and overlap to very similar extents. The ORFs in BYDV RNA (Miller et al., 1988a) also
Fig. 2. Diagram showing the arrangement of the ORFs in PLRV RNA. The horizontal line represents the 6 kb RNA and the grey blocks represent the ORFs with each reading frame on a different line. The numbers are those used in the text to describe the ORFs.

(a)

BWYV
GAUUAC-AAAUUCCUAGC-AGGCUUCG.(40).UAUCUAUUCUAUCUACC-U-AAAGA.(28).AUG

PLRV
GAUUAAU-AAAUUCCUAGC-GGGAUUUG.(40).UUUAUUUGGUGUAAACC-UAAAGA.(27).AUG

BYDV
UUACAAAU-CUUAGCGG-UUG..(44)...UACUUUGUACAAAUAGU.(26).AUG

(b)

(1682)  
CCCGCAGAGAAAACUGCUCAAACAAAC  
IJCAGCAGAGAACUGCUCCAUCAACU  
UCAGCAGAGAAAACUGCUCUAACAAAC (1762)

Fig. 3. (a) Similarities between luteovirus RNA sequences in the intergenic regions. The sequences are following nucleotides 3368 (BWYV), 3577 (PLRV) and 2745 (BYDV). Numbers in brackets are the lengths of intervening sequences in nucleotides, * indicates a match with the PLRV RNA sequence, – indicates that a gap has been inserted to enhance the amount of matching. (b) Sequence repeat in PLRV RNA. The sequence is continuous from nucleotide 1682 (top left) to nucleotide 1762 (bottom right); indicates the same nucleotide in all three repeats, ! indicates the same nucleotide in two of the three repeats.

Non-coding sequences

There are three non-coding regions in PLRV RNA. These are the 5'-terminal 174 (or more) nucleotides, the 3'-terminal 141 nucleotides and 197 nucleotides between ORF 2b and ORF 3 (intergenic region). No obvious similarities were detected between these sequences nor between the 5'-terminal sequence and the complement of the 3'-terminal sequence. The intergenic regions present in BWYV RNA and BYDV RNA (Veldt et al., 1988; Miller et al., 1988a) are similar in size to that in PLRV RNA and comparisons among the sequences of the regions revealed two stretches that apparently had a more than average degree of similarity. The similarities are shown in Fig. 3(a) and suggest that these stretches may be functionally significant. The right-hand stretch contains repeated UₚA sequences followed by AAGA, two features proposed by Marsh et al. (1988) as likely to be associated with the formation of subgenomic RNA. The abundant subgenomic fragment of PLRV RNA reported previously (Mayo et al., 1984) was detected by hybridization with two probes which represent the sequence

resemble those in PLRV RNA except that BYDV RNA has no ORF corresponding to PLRV ORF 1.
Sequence of PLRV RNA

between nucleotides 4200 and 4900 (approx.) or nucleotides 4800 and 5300 (approx.). It may therefore be a subgenomic mRNA for ORFs 3, 4 and/or 5, although the reported length of the fragment, 3-4 kb, implies that its 5' end must be well upstream of the intergenic region.

Sequence repeats

An unusual feature of PLRV RNA is an almost exact contiguous threefold repeat between nucleotides 1682 and 1762 (Fig. 3b). This feature was found in the sequences of clones from each of the RNA preparations used and is therefore unlikely to be a cloning artefact. However, no such repeat could be found in BWYV RNA and the DIAGON plot comparing the translation products of ORF 2a of PLRV and BWYV (Fig. 4a) shows a displacement of the diagonal line at a position corresponding to the repeated sequence. Possibly the repeated sequence arose as an insertion into a progenitor molecule and is not represented in the BWYV RNA sequence.

Sequence variants

Two RNA preparations were used for cloning. The first yielded cDNA clones from which 5 kb of sequence was determined. No sequence variations were detected in the 1-4 kb of sequence represented by more than one clone obtained from this source. The second RNA preparation yielded cDNA clones from which all but 65 nucleotides of the final sequence was determined. Half (3 kb) of this sequence was determined from more than one clone and four nucleotide changes were observed. These were at positions 480, 4298, 5741 and 5915. There was much more variation between the sequences determined from different RNA preparations, which resulted in the remaining 58 nucleotide exchanges shown in Fig. 1. These exchanges were not evenly distributed among the sequences that overlapped. The exchanges occurred in ORF 5 about twice as frequently (2-1%) than in the other ORFs (0-5% to 1%) although in all ORFs about half of the nucleotide exchanges did not alter the amino acid encoded by the sequence. This amount of variation between virus preparations of supposedly the same isolate of PLRV (58 changes in 5 kb) compares with 145 changes in 3484 nucleotides observed between RNA from the lettuce and sugarbeet isolates of BWYV (Veidt et al., 1988), and may reflect the difficulty in obtaining genetically homogeneous isolates of viruses that are not mechanically transmissible.

Expression of ORFs

In vitro translation of PLRV RNA using wheatgerm extracts with negligible background activity resulted in the appearance of two prominent products of 70K and 28K (Fig. 5). In previous work (Mayo et al., 1982) an endogenous product of the reticulocyte lysate used prevented the detection of a 28K product but with other batches of lysate this product was more visible, although translation was always more efficient in wheatgerm extracts than in reticulocyte lysate. Some minor intermediated sized products also appeared, which may have been the result of premature terminations, and in most experiments a small amount of a 125K product was detected (Fig. 5; Mayo et al., 1982). These principal products have the sizes predicted for the translation products of ORF 1 (28K) and ORF 2a (70K), and 125K is close to the size predicted for the product of a small amount of readthrough from ORF 2a into ORF 2b (118K).

Only trace amounts of a coat protein-sized polypeptide were detected when translation products of PLRV RNA were immunoprecipitated and this amount may have been made by translation of fragmented RNA. Partial proteolysis products of the 28K translation product did not comigrate with those of PLRV coat protein (data not shown) and thus there is no evidence of significant translation from genomic RNA of an ORF yielding coat protein.

PLRV VPg is presumably a virus-coded protein but we could find no evidence from the nucleotide sequence of PLRV RNA as to how VPg is made.

Overlapping coding regions

If it is assumed that the 70K translation product is that of ORF 2a and that it did not result from translation of fragmented RNA, then about 10% to 20% of ribosomes scanning from the
Fig. 4. DIAGON comparisons of translation products. (a) Products of ORF 2a of PLRV (horizontal) and BWYV (vertical). Arrow indicates a displacement in the line of homology between the sequences. (b) Products of ORF 2b of PLRV (horizontal) and BWYV (vertical). Vertical line in (a) and (b) shows the position assumed for the frameshift when computing the sequence of P2. (c) Products of ORF 5 of PLRV (horizontal) and BWYV (vertical). (d) Products of ORF 5 of PLRV (horizontal) and BYDV (vertical). The diagram shows proportional scores of 252 or more for a window of 21 amino acids.
Sequence of PLRV RNA

5' end of the RNA passed the initiation site of ORF 1 (nt 175) and initiated translation at the AUG codon at nucleotide 308. The initiation site for ORF 4 is only 26 nucleotides downstream of the initiation site for ORF 3 and, assuming that there is no separate subgenomic RNA for ORF 4, it must also be translated by ribosomes that have bypassed the initiation site of ORF 3. None of the nucleotide sequences around the proposed initiation sites corresponded closely to the consensus sequences for translation either by animal (Kozak, 1986) or plant (Lutcke et al., 1987) ribosomes and thus internal initiation is a possible mechanism for the translation of P2 and P4 (Kozak, 1986).

Evidence for frameshift readthrough

The first AUG codon in ORF 2b is located 900 nucleotides from the preceding 5' termination codon but it seems likely that the region 5' of this AUG is translated because its putative translation product has an amino acid sequence similar to that of the corresponding translation product of BWYV RNA (Fig. 4b). However, although this similarity extends upstream of the AUG it does not extend to the preceding termination codon. Whereas DIAGON comparison of the translation products of ORF 2a shows patches of amino acid sequence homology over the whole length of the ORF (Fig. 4a), comparison of the ORF 2b translation products shows that
Table 1. Comparison of general properties of polypeptides encoded by RNA of PLRV, BWYV and BYDV

<table>
<thead>
<tr>
<th></th>
<th>PLRV</th>
<th></th>
<th>BWYV</th>
<th></th>
<th>BYDV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M_r$</td>
<td>Charge</td>
<td>$H$</td>
<td>$M_r$</td>
<td>Charge</td>
<td>$H$</td>
</tr>
<tr>
<td>P1</td>
<td>28130</td>
<td>+3</td>
<td>25.4</td>
<td>29076</td>
<td>+3</td>
<td>31.4</td>
</tr>
<tr>
<td>P2</td>
<td>117947</td>
<td>+5</td>
<td>-72.2</td>
<td>114939</td>
<td>-1</td>
<td>-21.8</td>
</tr>
<tr>
<td>P3</td>
<td>23196</td>
<td>+24</td>
<td>-33.7</td>
<td>22284</td>
<td>+23</td>
<td>-21.8</td>
</tr>
<tr>
<td>P4</td>
<td>23196</td>
<td>+24</td>
<td>-33.7</td>
<td>22284</td>
<td>+23</td>
<td>-21.8</td>
</tr>
<tr>
<td>P5</td>
<td>53119</td>
<td>-15</td>
<td>-109.7</td>
<td>51545</td>
<td>-16</td>
<td>-99.8</td>
</tr>
</tbody>
</table>

* Charge is the difference between the total number of lysine and arginine residues and the total number of aspartic and glutamic acid residues.
† $H$ is the hydrophobicity assessed by ANALYSEQ.

homology begins only after approx. nucleotide 1960 in PLRV RNA and 1770 in BWYV RNA. This result therefore suggests that ORF 2b is translated by some of the ribosomes that translate ORF 2a undergoing frameshift near nucleotide 1960 in PLRV RNA, or 1770 in BWYV RNA, and continuing translation to the end of ORF 2b to yield a 118K product (P2).

It has been suggested that ribosome slippage resulting in frameshift may be associated with the occurrence of G residues in the second position in codons; frameshift then restores the preferred (G–non–G–N) codon structure (Trifonov, 1987). In the sequences of PLRV RNA and BWYV RNA there are successions of codons in ORF 2a with G in the second position immediately following nucleotides 1969 in PLRV RNA and 1745 in BWYV RNA. These positions are close to those suggested for frameshift from the DIAGON plots (Fig. 4b) and to a hexanucleotide (AAAGAA) which is common to the two RNA sequences.

Expression of ORF 5

Although there is an AUG codon 96 nucleotides downstream from the termination codon of ORF 3 that is a candidate for the start of the translation product of ORF 5, there is circumstantial evidence that suggests this is not so. ORF 5 is in frame with ORF 3 and translation of the nucleotide sequence between ORF 3 and ORF 5 yields an amino acid sequence very like that translated from the analogous region of both BWYV RNA and BYDV RNA. Veidt et al. (1988) have shown by translation of transcripts of DNA copies of this region of BWYV RNA that readthrough can occur in vitro, and the nucleotide sequence around the ORF 3 termination codon (CCAAAUAGGU) corresponds exactly to that around the analogous termination codons in BWYV RNA and BYDV RNA. Moreover, this sequence is similar to that around UAG termination codons that are read through during translation of the RNA of several other viruses (Miller et al., 1988a).

Properties of the putative translation products

The principal translation products of the large ORFs are shown in Table 1 in comparison with those of BWYV RNA and BYDV RNA. P2 is the product of ORF 2a and ORF 2b linked by frameshift at nucleotide 165 or 1765 for PLRV RNA and BWYV RNA, respectively, or proposed by Miller et al. (1988a) for BYDV RNA. The polypeptides from each ORF of PLRV are distinct in their general properties although there are similarities among the three luteoviruses for the corresponding proteins.

P1 of PLRV is substantially more hydrophobic than the other PLRV proteins and a major part of this is determined by the N-terminal region of the sequence. Applying the criteria described by Eisenberg et al. (1984) to the sequence of P1 results in a prediction that between residues 21 and 32 the protein is membrane-associated. A search of the protein sequence database resulted in detectable similarities between P1 and a variety of membrane-associated proteins such as cytochrome b5 and cytochrome P450, and these regions of similarity were at the N termini of the proteins. The amino acid sequence that follows residue 19 resembles that of
Sequence of PLRV RNA

Signal peptides in having 14 hydrophobic residues after a lysine (Briggs & Gierasch, 1986), which further suggests that P1 is membrane-associated. In contrast, although BWYV P1 is relatively hydrophobic (Table 1), its sequence did not suggest membrane linking when assessed by the method of Eisenberg et al. (1984). Furthermore DIAGON comparisons of P1 proteins of PLRV and BWYV showed no significant similarities between the proteins.

The amino acid sequence of P2 between residues 931 and 961 resembles the consensus sequence GXXXXXXXNX~8_s0GDD thought to represent a core sequence of RNA-dependent RNA polymerases (Kamer & Argos, 1984). As in the analogous proteins of other viruses this motif is near the C terminus of a relatively large protein. No NTP-binding site consensus sequence (Higgins et al., 1986) was detected in any of the PLRV proteins.

A DIAGON comparison between PLRV P2 and BWYV P2 showed extensive similarity especially in the C-terminal half of the molecule (Fig. 4). An alignment by GAP of the P2 sequences gave an identity score of 51% but a comparison of the sequences between residues 700 and 1041 of PLRV P2 and residues 671 and 1010 of BWYV P2 gave a value of 74% identity. In contrast, no similarities were detected by DIAGON plots of PLRV P2 and BYDV P2; the sequences around the GDD consensus motif were no more alike than either was to the corresponding region of the putative polymerase sequences of several unrelated viruses (e.g. Kamer & Argos, 1984). However, as detected with BWYV P2 (Veidt et al., 1988), there was a striking resemblance between PLRV P2 and the putative polymerase of southern bean mosaic virus (SBMV) (Wu et al., 1987). In two blocks between residues 755 and 829 and residues 905 and 964 of PLRV P2, GAP gave values of 67% and 59% identity respectively to regions in SBMV P2. These are essentially the same regions as illustrated by Veidt et al. (1988) when comparing SBMV P2 with the corresponding protein of BWYV.

P3 corresponds in position to the coat protein of BYDV (Miller et al., 1988b) and to the putative coat protein of BWYV (Veidt et al., 1988). The average of 10 separate determinations of the Mr of PLRV coat protein by electrophoresis in SDS-polyacrylamide gels was 23800 which is close to the calculated Mr of 23196 for P3 (Table 1). Transfection of HeLa cells with an expression vector containing a DNA fragment corresponding to nucleotides 3580 to 4580 of the sequence shown in Fig. 1 resulted in synthesis of a polypeptide that comigrated with PLRV coat protein and reacted specifically with antiserum to PLRV particles (B. Reavy & M. Mayo, unpublished results). We conclude that P3 is the coat protein of PLRV. P3 has a marked positive charge (Table 1) mainly because of the very arginine-rich sequence near the N terminus. This feature is shared with other luteovirus coat proteins as well as those of other unrelated viruses (Argos, 1981) including SBMV (Hermodson et al., 1982). By analogy with SBMV, we suggest that in PLRV particles this terminal region reacts with negatively charged groups of the RNA particle (Hermodson et al., 1982) and is therefore located on the inner side of the protein coat of the virus particle.

Fig. 6 shows a comparison of the sequences of the coat proteins of PLRV, BWYV and BYDV. With relatively few insertions of gaps the sequences for PLRV and BWYV align with 66% of the paired amino acids being identical (66% identity) and each can be aligned to be 49% identical to that of the BYDV coat protein. The sequences are most alike in their central regions and differ most near their N termini.

P4 lies completely within the nucleotide sequence coding for P3 and therefore the amino acid sequences of these proteins are to some extent mutually constrained. An alignment by GAP gave values of 45% and 33% identity between the P4 of PLRV and BWYV and of PLRV and BYDV respectively. The pentapeptide sequence QWLWS was near the N terminus of each of the luteovirus P4 proteins.

As discussed above, P5 is probably the C-terminal part of a readthrough product comprising P3 and P5. However, for comparative purposes it is considered here as a separate protein. The amino acid sequence that precedes the first methionine in P5 resembles the corresponding sequences of BWYV and BYDV and contains a sequence of prolines alternating with mostly serine or threonine. The occurrence of this unusual sequence at the N terminus of P5 of each of the three luteoviruses suggests that this region has an important function. As pointed out by Miller et al. (1988a) this sequence is characteristic of proteins that are rapidly degraded in cells
Fig. 6. Comparisons between the amino acid sequences of the coat proteins of PLRV, BWYV and BYDV. The PLRV sequence is represented twice to allow each sequence to be compared with both of the others. The numbers to the right are the numbers of matching amino acids/the number of pairs of amino acids in that line. The symbol * represents a match between adjacent sequences.

In DIAGON comparisons, P5 of PLRV, BWYV and BYDV showed two main regions of similarity (Fig. 4c, d). The N-terminal half of PLRV P5 is similar to P5 of BWYV and BYDV; GAP gave a value of 52% identity with BWYV P5 for the first 244 amino acids and 48% identity with BYDV for the same region. In the C-terminal regions, PLRV P5 shows little similarity with BWYV P5 and none with BYDV P5. A similar distinction between the two halves of P5 was made by Veidt et al. (1988) when looking at the frequency of sequence variations between two isolates of BWYV.

DISCUSSION

PLRV is the third luteovirus for which the nucleotide sequence is known and a comparison of the three sequences allows some generalizations to be proposed concerning the strategy of expression of the genomes of luteoviruses. The coat protein-coding regions have some unusual properties. They differ from those in most other virus RNA in that they are not the gene nearest the 3' end of the RNA, an arrangement resembling that in the genome of cymbidium ringspot tombusvirus (Russo et al., 1988) but few other virus genomes, and the luteovirus coat protein gene contains, wholly within it and in a different frame, a second coding region. Moreover it seems likely that readthrough of the termination codon of the luteovirus coat protein produces an over-size coat protein which has a C-terminal extension. It is possible that incorporation of such a modified coat protein into virus particles might contribute a particular particle property
important in the behaviour of luteoviruses such as in the complex relationship between virus particles and aphid transmission (Harrison, 1984).

Frameshift readthrough has been proposed to explain the expression of the RNA polymerase gene of each of the luteoviruses sequenced so far and therefore may also be a general characteristic of luteovirus genomes. However, although PLRV RNA resembles BWYV RNA in the arrangement of the coding sequences in the 5' half of the genome, BYDV RNA is different in that it lacks an ORF at the 5' end coding for a polypeptide of about 28K and the overlap between the ORFs thought to be involved in frameshift is only 13 nucleotides whereas that in PLRV RNA and BWYV RNA is about 475 nucleotides. Moreover the sequence of the putative polymerase of BYDV is quite different from those of the polymerases of PLRV and BWYV.

Serological studies have shown that PLRV is related to BWYV (Tamada et al., 1984), but distinct from it (Massalski & Harrison, 1987), and more distantly related to BYDV (Roberts et al., 1980). Comparisons among the amino acid sequences of the coat proteins of the three viruses confirm this pattern of relatedness. Percentage identity is not a proportionate indication of how much change has occurred during the evolutionary divergence of two sequences because of back mutations and multiple 'hits' (Doolittle, 1986). A computer simulation of the effects of random nucleotide changes in the sequence of ORF 3 on the percentage identity of the translation product with that of the original coding sequence showed that 66% identity implies an average of 103 'hits' on the coding sequence and 49% identity implies an average of 214 'hits'. The calculation assumes equal constraint on all nucleotides in the sequence and that all changes survive, neither of which is likely to be true, but the relative relatedness of the PLRV coat protein with the coat proteins of BWYV and BYDV is independent of these assumptions. For all three luteoviruses, PLRV was more similar to BWYV than either was to BYDV.

PLRV is most distinct from BWYV at the extremities of the coding sequence. Neither P1 nor the C-terminal half of P5 were like the corresponding sequences of BWYV. However, it would seem likely that corresponding sequences have similar functions in the two viruses; indeed P1 of both PLRV and BWYV are unusually hydrophobic. The three luteoviruses compared differ most notably in their host ranges; it may be that P1 reflects this difference, conceivably being dispensable for the infection of monocotyledonous hosts.

The most unexpected feature of luteovirus genomes has been commented on by Miller et al. (1988a) and Veidt et al. (1988). This is the similarities in the polymerase sequences of luteoviruses with viruses in other groups: e.g. BYDV with carnation mottle virus and BWYV with SBMV. PLRV resembles BWYV in having a polymerase like that of SBMV, an observation that further supports the idea suggested by Miller et al. (1988a) that luteovirus genomes have arisen by recombination between blocks of RNA. The essential characteristic of luteovirus genomes may be only in the 3' coding portion with the 5' portion being a piece of RNA acquired from another virus which is not necessarily similar for all the luteoviruses. Thus other luteoviruses may have 5' regions unlike either PLRV/BWYV or BYDV. Clearly taxonomic relatedness among luteoviruses would appear very different if assessed by characters determined by the 5' region or by the 3' region. Moreover this unusual diversity makes generalization about luteovirus genomes impossible and the determination of the sequences of the genomes of other luteoviruses more interesting and more important than is usual for the sequencing of further genomes from within one virus group.

Note added in proof. The nucleotide sequences reported will appear in the EMBL, GenBank and DDBJ databases under the accession number X14600.

We thank Miss Ann Jenkins for technical assistance, Drs W. Hamilton and R. Hay for helpful advice and Drs R. Goldbach and R. Hay for gifts of oligonucleotide primers. We also thank Dr D. Baulcombe for the gift of two PLRV-specific clones and Dr K. Richards and his colleagues for allowing us access to their results before publication.

REFERENCES


Sequence of PLRV RNA


(Received 16 November 1988)