Potato Leaf Roll Virus 3′ Genome Organization: Sequence of the Coat Protein Gene and Identification of a Viral Subgenomic RNA

O. P. Smith and K. F. Harris

Research plant pathologist, U. S. Department of Agriculture, Agricultural Research Service, Foreign Disease-Weed Science Research, Fort Detrick, Bldg. 1301, Frederick, MD 21701, and professor, Department of Entomology, Texas A&M University, College Station 77843.

Mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by the U. S. Department of Agriculture and does not imply approval to the exclusion of other products that also may be suitable.

We gratefully acknowledge Mr. Gary Smythers of the Advanced Scientific Computing Laboratory, National Cancer Institute, Frederick Research Facility for help with computer analyses. We thank Dr. Sean Carrol, University of Wisconsin, for kindly providing pUR291 and Mr. Andrew Stone for excellent technical assistance.

Supported in part by funds provided by the Texas Agricultural Experiment Station.

Accepted for publication 28 November 1989 (submitted for electronic processing).

ABSTRACT

Smith, O. P., and Harris, K. F. 1990. Potato leaf roll virus 3′ genome organization: Sequence of the coat protein gene and identification of a viral subgenomic RNA. Phytopathology 80:609-614.

The nucleotide sequence of the 3′ portion of the potato leaf roll virus (PLRV, isolate 4) genome representing 2,361 bases has been determined from cloned cDNA. The identity of the open reading frame (ORF) encoding the coat protein was confirmed by expression of an *Escherichia coli* β-galactosidase fusion protein and detection by a dot blot immunoassay. The coat protein gene contains 624 nucleotides and encodes a protein of 208 amino acids with a calculated molecular weight of 23.2 kDa. This gene is located internal to the 3′-end of the genome, contains the complete overlap of an ORF encoding a 17.4 kDa protein, and is followed in-frame by an ORF encoding a 56.4 kDa protein. Northern blot hybridization experiments have identified a 3′-positioned viral subgenomic RNA (approximately 2.5 kb) produced in PLRV-infected plants. This subgenomic RNA is not encapsidated and, by analogy to other RNA plant viruses, is likely the mRNA for expression of the 3′ portion of the genome. A comparison of the PLRV-4 coat protein sequence with two other luteoviruses, beet western yellows virus (isolate FL1) and barley yellow dwarf virus (isolate PAV) has revealed amino acid sequence similarities of 64 and 45%, respectively. The coat protein stop codon (*TAG*) of all three luteoviruses is contained within a conserved 14-nucleotide sequence (CCAAA*TAG*GTAGAC).

Potato leaf roll virus (PLRV) is a member of the luteovirus group (11) and causes an economically important disease of potato (19,20,27). Its isometric 25-nm virion consists of a positive-stranded, nonpolyadenylated, RNA genome encapsidated by a viral-encoded coat protein. The RNA genome and coat protein of PLRV have size estimates of 6 kb and 26 kDa, respectively (22). In vitro translation of PLRV genomic RNA yields a major polypeptide species of 71 kDa; however, a product corresponding in size to coat protein has not been observed (12).

Because PLRV is obligatorily aphid-transmitted and restricted to the phloem tissue of its host (19,20), virus propagation is resource intensive and yields of virus and viral RNA are very low. As a first step to circumvent these limitations and to facilitate molecular biology studies, we recently reported the construction of a PLRV cDNA library and identified three partial-length cDNAs that form an overlapping physical map of a majority of the viral genome (26).

Recent reports on the molecular biology of PLRV have focused on the technique of nucleotide sequencing (7,13). In this paper, we describe the 3′ genome organization and expression strategy of PLRV based on a combination of experimental approaches, including nucleotide sequencing, bacterial expression of a β-galactosidase fusion protein, and northern blot hybridization experiments. These approaches have resulted in identification of the PLRV coat protein gene and detection of a 3′, viral, subgenomic RNA produced in virus-infected plants.

MATERIALS AND METHODS

Virus and cDNA clones. PLRV isolate 4 was maintained in *Physalis floridana* Rybd. plants by serial transmissions using the green peach aphid, *Myzus persicae* (Sulzer). The cDNA clones pPLRV4-173, -228, and -323 representing the 5′, 3′, and middle genomic portions, respectively, of this viral isolate have been described (26).

DNA sequencing. Plasmid DNA was sequenced by the dideoxy method (24) with the Sequenase kit (United States Biochemical Corp., Cleveland, OH). Reactions were conducted with custom synthesized primers (Synthecell Corp., Rockville, MD) or M13 primers. The resulting data were analyzed using the sequence analysis software package of the Genetics Computer Group (5).

Construction and immunoassay of fusion proteins. The 5′ terminal *Bam*HI fragment of pPLRV4-228 was isolated by electroelution (25) and subcloned into pUR291, a β-galactosidase translational fusion vector (23), using *Escherichia coli* JM109 host cells (31).

Bacterial lysates were processed for dot blot immunoassay by using an alkaline phosphatase detection kit (Promega Corp., Madison, WI). Polyclonal antiserum (rabbit) to purified PLRV, kindly provided by Dr. R. Stace-Smith of Agriculture Canada, was used at a dilution of 1:1,000. The diluted antiserum was preincubated for 30 min at room temperature with 0.5 mg/ml total *E. coli* protein to remove cross-reacting antibodies of *E. coli*. Lysates (100 μl) were dot blotted under vacuum to nitrocellulose using a Minifold I apparatus (Schleicher and Schuell, Keene, NH). Purified PLRV (28) was used as a positive control.

Preparation of *E. coli* lysates for the dot blot immunoassay. Bacterial lysates were prepared based on the methods of Azad et al (1). Cultures (2.5 ml) were grown overnight in Luria broth containing 70 μg/ml of ampicillin and supplemented with 0.5 mM isopropyl thiogalactosidase to induce β-galactosidase expression. After centrifugation (3,000 *g*) at 4 C for 10 min, each bacterial pellet was suspended in 300 μl of TBS (50 mM Tris-Cl, pH 8.0, 150 mM NaCl) and transferred to a 1.5-ml microfuge tube. After addition of 5 μl of 10 mg/ml of lysozyme (Sigma, St. Louis, MO) and incubation on ice for 15 min, 35 μl of 10%

SDS was added and the incubation continued for 30 min at room temperature. After microcentrifugation (12,000 *g*) at 4 C for 10 min to remove bacterial debris, the lysate was stored at −80 C.

RNA isolation and northern blot hybridization. PLRV genomic RNA was isolated as described previously (26) using virus that was purified without using tissue macerases (3). Total plant RNA was isolated from *P. floridana* using a combination of methods (2,8,17). Four grams of liquid nitrogen-pulverized leaves was homogenized in 10 ml of ice-cold extraction buffer (10 mM Tris-Cl, pH 7.5, 5.0 M guanidine-HCl, 5 mM EGTA, 100 mM β-mercaptoethanol, and 0.1% sodium laurylsarcosine) for 45 sec using a tissuemizer (Tekmar Co., Cincinnati, OH) (17). After removal of plant debris by centrifugation (10,000 *g*) at 4 C for 20 min, the supernatant was extracted twice with 0.5 volumes phenol/chloroform/isoamyl alcohol (25:24:1) (8), followed by addition of 0.5 volume ethanol (2) and storage at −20 C overnight. After centrifugation, the pellets were washed twice with 70% ethanol (−20 C) and suspended in 10 ml of TNE (10 mM Tris-Cl, pH 7.5, 100 mM NaCl, and 0.1 mM EDTA) containing 0.1% sodium laurylsarcosine and 300 μg/ml of proteinase K (Bethesda Research Laboratories, Inc., Gaithersburg, MD). After incubation for 1 hr at 37 C, the mixture was extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1). After addition of one-tenth volume 3.0 M sodium acetate, pH 5.5, and 2.5 volumes ethanol, the RNA was stored overnight at −20 C. RNA was recovered by centrifugation, suspended in sterile distilled water, and quantitated spectrophotometrically (9). Aliquots of RNA (40 μg) were stored in 1.5-ml microfuge tubes as ethanol/sodium acetate suspensions (−20 C) and collected by microcentrifugation immediately before electrophoresis.

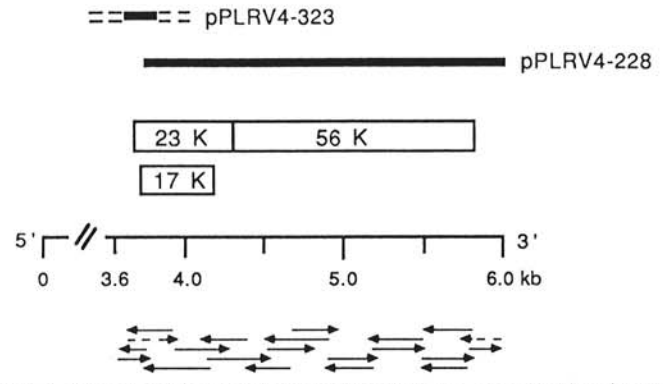

Fig. 1. Summary of the sequencing strategy and open reading frames (ORFs) for the 3′ portion of the potato leaf roll virus (PLRV) genome. Arrows show the location, direction, and extent of data obtained by complete sequence analysis of pPLRV4-228 and partial sequence analysis of pPLRV4-323 using custom primers (solid arrows) and M13 primers (dashed arrows). A dot blot immunoassay confirmed that the 23K ORF encodes the coat protein (see Fig. 3).

```
                                CATTTTCAGTAGCCGGTTTATATTTTGTTTACCTAAAGATTTCCTCCCACGTGCGATCAATTGTTAATGAGTACGGTCGTGGTTAAAGGAAATGTCAATGGTGGTGTACAACAACCAAGA  120
Coat Protein (23 K) ORF                                                                                 M  S  T  V  V  V  K  G  N  V  N  G  G  V  Q  Q  P  R
17 K ORF                                                                                                                M  S  M  V  V  Y  N  N  Q  E
                                AGGCGAAGAAGGCAATCCCTTCGCAGGCGCGCTAACAGAGTTCAGCCAGTGGTTATGGTCACGGCCCCTGGGGAACCCAGGCGCCGAAGACGCAGAAGAGGAGGCAATCGCCGCTCAAGA  240
Coat Protein (23 K) ORF  R  R  R  R  Q  S  L  R  R  R  A  N  R  V  Q  P  V  V  M  V  T  A  P  G  E  P  R  R  R  R  R  R  R  G  G  N  R  R  S  R
17 K ORF                  G  E  E  G  N  P  F  A  G  A  L  T  E  F  S  Q  W  L  W  S  R  P  L  G  N  P  G  A  E  D  A  E  E  E  A  I  A  A  Q  E
                                AGAACTGGAGTTCCCCGAGGACGAGGCTCAAGCGAGACATTCGTGTTTACAAAGGACAACCTCGTGGGCAACACCCAAGGAAGTTTCACCTTCGGGCCGAGTCTATCAGACTGTCCGGCA  360
Coat Protein (23 K) ORF  R  T  G  V  P  R  G  R  G  S  S  E  T  F  V  F  T  K  D  N  L  V  G  N  T  Q  G  S  F  T  F  G  P  S  L  S  D  C  P  A
17 K ORF                  E  L  E  F  P  E  D  E  A  Q  A  R  H  S  C  L  Q  R  T  T  S  W  A  T  P  K  E  V  S  P  S  G  R  V  Y  Q  T  V  R  H
                                TTCAAGGATGGAATACTCAAGGCCTACCATGAGTATAAGATCACAAGCATCTTACTTCAGTTCGTCAGCGAGGCCTCTTCCACCTCCTCCGGTTCCATCGCTTATGAGTTGGACCCCCAT  480
Coat Protein (23 K) ORF  F  K  D  G  I  L  K  A  Y  H  E  Y  K  I  T  S  I  L  L  Q  F  V  S  E  A  S  S  T  S  S  G  S  I  A  Y  E  L  D  P  H
17 K ORF                  S  R  M  E  Y  S  R  P  T  M  S  I  R  S  Q  A  S  Y  F  S  S  S  A  R  P  L  P  P  P  P  V  P  S  L  M  S  W  T  P  I
                                TGCAAAGTATCATCCCTCCAGTCCTACGTCAACAAGTTCCAAATTACGAAGGGCGGCGCCAAAACTTATCAAGCGCGGATGATAAATGGGGTAGAATGGCACGATTCTTCTGAGGATCAG  600
Coat Protein (23 K) ORF  C  K  V  S  S  L  Q  S  Y  V  N  K  F  Q  I  T  K  G  G  A  K  T  Y  Q  A  R  M  I  N  G  V  E  W  H  D  S  S  E  D  Q
17 K ORF                  A  K  Y  H  P  S  S  P  T  S  T  S  S  K  L  R  R  A  A  P  K  L  I  K  R  G  *
                                TGTCGGATACTGTGGAAGGGAAATGGAAAATCTTCAGATACCGCAGGATCCTTCAGAGTCACCATCAGGGTGGCTTTGCAAAACCCCAAATAGGTAGACTCCGGATCAGAGCCTGGTCCA  720
Coat Protein (23 K) ORF/56 K ORF  C  R  I  L  W  K  G  N  G  K  S  S  D  T  A  G  S  F  R  V  T  I  R  V  A  L  Q  N  P  K  *  V  D  S  G  S  E  P  G  P
                                AGCCCACAACCAACACCCACTCCAACTCCCCAGAAGCACGAGCGATTTATTGCTTACGTTGGCATACCTATGCTAACCATTCAGGCCAGGGAGAACGACGACCAAATCATATTGGGTTCC  840
56 K ORF                 S  P  Q  P  T  P  T  P  T  P  Q  K  H  E  R  F  I  A  Y  V  G  I  P  M  L  T  I  Q  A  R  E  N  D  D  Q  I  I  L  G  S
                                TTAGGGAGCCAAAGGATGAAATATATAGAAGACGAGAACCAGAACTATACAAATGTTAGTTCTGAGTATTACTCTCAATCGAGCATGCAAGCCGTCCCTATGTATTACTTCAATGTCCCG  960
56 K ORF                 L  G  S  Q  R  M  K  Y  I  E  D  E  N  Q  N  Y  T  N  V  S  S  E  Y  Y  S  Q  S  S  M  Q  A  V  P  M  Y  Y  F  N  V  P
                                AAAGGGCAATGGTCAGTCGACATCAGCTGCGAAGGGTATCAACCCACCAGCAGCACCTCGGATCCAAACCGGGGTAGGAGTGACGGAATGATCGCGTATTCAAACGCGGACTCCGATTAT 1080
56 K ORF                 K  G  Q  W  S  V  D  I  S  C  E  G  Y  Q  P  T  S  S  T  S  D  P  N  R  G  R  S  D  G  M  I  A  Y  S  N  A  D  S  D  Y
                                TGGAATGTTGGTGAAGCGGATGGTGTCAAAATTTCGAAGCTACGCAACGATAACACCTACCGCCAAGGTCACCCAGAACTTGAAATTAACTCGTGTCATTTTCGCGAGGGCCAACTCCTT 1200
56 K ORF                 W  N  V  G  E  A  D  G  V  K  I  S  K  L  R  N  D  N  T  Y  R  Q  G  H  P  E  L  E  I  N  S  C  H  F  R  E  G  Q  L  L
                                GAACGGGACGCTACAATTAGCTTCCACGTTGAAGCGCCTACTGATGGGCGATTCTTTCTCGTTGGTCCCGCTATCCAGAAAACCGCAAAGTATAACTATACTATCTCATACGGTGACTGG 1320
56 K ORF                 E  R  D  A  T  I  S  F  H  V  E  A  P  T  D  G  R  F  F  L  V  G  P  A  I  Q  K  T  A  K  Y  N  Y  T  I  S  Y  G  D  W
                                ACGGACCGAGACATGGAACTGGGGCTGATCACCGTGGTGCTTGATGAACATTTAGAAGGCACTGGTTCGGCTAACAGAGTGCGGCGGCCCCTACGGGAGGGCCACACCTATATGGCCTCG 1440
56 K ORF                 T  D  R  D  M  E  L  G  L  I  T  V  V  L  D  E  H  L  E  G  T  G  S  A  N  R  V  R  R  P  L  R  E  G  H  T  Y  M  A  S
                                CCGCGTGAACCGGAAGGAAAACCGGTTGGAAATAAACCAAGGGACGAAACCCCAATACAAACGCAGGAAAGACAACCTGATCAAACTCCGTCTGACGACGTGTCCGATGCTGGTTCGGTA 1560
56 K ORF                 P  R  E  P  E  G  K  P  V  G  N  K  P  R  D  E  T  P  I  Q  T  Q  E  R  Q  P  D  Q  T  P  S  D  D  V  S  D  A  G  S  V
                                AACAACGGCGGCTCAACTGAGTCGCTGCAATTGGAGTTCGGGGTAAACTCAGATAGTACCTACGATGCTACAGTCGATGGTACAGACTGGCCCAGAATTCCTCCACCAAGGCACCCACCT 1680
56 K ORF                 N  N  G  G  S  T  E  S  L  Q  L  E  F  G  V  N  S  D  S  T  Y  D  A  T  V  D  G  T  D  W  P  R  I  P  P  P  R  H  P  P
                                GAACCTGGAGTTTCCGGCAATTCAAGAACTGTTACTGACTTTTCTTCAAAAGCCGATCTATTGGAGAATTGGGATGCCGAACACTTCGACCCTGGTTATTCCAAAGAAGATGTCGCTGCT 1800
56 K ORF                 E  P  G  V  S  G  N  S  R  T  V  T  D  F  S  S  K  A  D  L  L  E  N  W  D  A  E  H  F  D  P  G  Y  S  K  E  D  V  A  A
                                GCTACTATTATAGCGCACGGCAGTATTCAAGATGGGCGAAGTATGCTGGAGAAGAGAGAGGAAAATGTCAAGAACAAAACCTCCTCCTGGAAGCCCCCGTTATCTAAAGCGGTGAGCCCA 1920
56 K ORF                 A  T  I  I  A  H  G  S  I  Q  D  G  R  S  M  L  E  K  R  E  E  N  V  K  N  K  T  S  S  W  K  P  P  L  S  K  A  V  S  P
                                GCCATAGCCAAATTGCGCTCGATTCGCAAATCCCAACCCCTCGAGGGAGGGACCCTTAATAAAGACGCCACTGATGGTGTCTCATCTATTGGCAGTGGTTCTCTAACAGGTGGCACGCTT 2040
56 K ORF                 A  I  A  K  L  R  S  I  R  K  S  Q  P  L  E  G  G  T  L  N  K  D  A  T  D  G  V  S  S  I  G  S  G  S  L  T  G  G  T  L
                                AAGAGGAAGGCAACTATTGAAGAGCGTTTACTGCAGACCTTAACAACTGAACAAAGGCTGTGGTACGAGAATTTGAAGAAAACTAACCCTCCAGCTGCCACCCAATGGCTGTTTGAATAT 2160
56 K ORF                 K  R  K  A  T  I  E  E  R  L  L  Q  T  L  T  T  E  Q  R  L  W  Y  E  N  L  K  K  T  N  P  P  A  A  T  Q  W  L  F  E  Y
                                CAGCCACCTCCCCAAGTGGATAGAAACATAGCTGAAAAGCCATTCCAAGGGAGGAAATGAGTCGACTCACGACTTAAAACTGAGCGTCCGCCGGACATTAAGCGGAACGAAAGCCGAAAG 2280
56 K ORF                 Q  P  P  P  Q  V  D  R  N  I  A  E  K  P  F  Q  G  R  K  *
                                GTGATTAGGCTCTCAACGCCTGCTAGAGACCGTCGAAAGACGCGACTGTGTAGCCAGGATCCTCTTACAGGGTTGTGTAGT                                          2361
```

Fig. 2. Sequence of the 3′-proximal 2,361 nucleotides of the potato leaf roll virus (PLRV) genome. Amino acids are shown in single letter code below the nucleotide sequence. Asterisks represent stop codons. The designated open reading frames (ORFs) are shown on the left. The underlined 14-nucleotide sequence, containing the coat protein stop codon (TAG), is conserved in two other luteoviruses, beet western yellows virus and barley yellow dwarf virus (see Discussion). Nucleotides 1-107 and 108-2361 are from pPLRV4-323 and pPLRV4-228, respectively.

Samples for northern blot hybridization were processed for denaturation by the methods of Maniatis et al (9) and electrophoresed in a 1.2% agarose gel containing 0.66 M formaldehyde, followed by transfer to nitrocellulose using 10× SSC (1× is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0) (4). Prehybridization, hybridization, and autoradiography were as described previously (26) using nick-translated ^{32}P-labeled plasmids (21). RNA size standards (Bethesda Research Laboratories) were labeled by hybridization to nick-translated, ^{32}P-labeled, lambda DNA (International Biotechnologies, Inc., New Haven, CT).

RESULTS

Nucleotide sequence analysis of the 3′ end of PLRV genomic RNA. The sequence of the 3′ portion of the PLRV genome was determined by analysis of 2,361 nucleotides obtained from clones pPLRV4-228 and -323. The three largest open reading frames (ORFs) are shown in Figure 1. The complete sequence is presented in Figure 2. The first ORF (designated 23K) begins at an ATG located 2,295 nucleotides from the 3′ terminus of cDNA 228. This ORF continues to a TAG stop codon located 1,671 nucleotides from the 3′ terminus to produce a reading frame of 624 nucleotides encoding a protein of 208 amino acids with a calculated molecular weight of 23.2 kDa. *E. coli* expression of a major portion of this ORF as a fusion protein, followed by immunological testing, has confirmed that this ORF encodes PLRV coat protein (see below). The 23K ORF contains the complete overlap of a second ORF in the +1 frame. This second ORF (designated 17K) initiates at an ATG located 2,270 nucleotides from the 3′ terminus of cDNA 228 and continues for 468 nucleotides to a TGA stop codon located 1,802 nucleotides from the 3′ terminus. This overlapping ORF encodes a protein of 156 amino acids with a calculated molecular weight of 17.4 kDa. Immediately after the 23K ORF is a larger in-frame ORF. This third ORF (designated 56K) continues to a TGA stop codon located 144 nucleotides from the 3′ terminus of cDNA 228 to produce a reading frame of 1,524 nucleotides encoding a protein of 508 amino acids with a calculated molecular weight of 56.4 kDa. The 56K ORF is followed by a 3′ noncoding region of 141 nucleotides.

***E. coli* expression of a β-galactosidase PLRV coat protein fusion.** Based on the availability of compatible restriction sites and the sequence of the 5′ terminus of cDNA 228, the 5′ terminal *Bam*HI fragment of cDNA 228, containing amino acids 15–195 of the 23K ORF, was subcloned into pUR291 to produce a translational fusion protein. As outlined in Figure 3, the sense (+) construct of the fusion protein and purified virus reacted positively in a dot blot immunoassay using PLRV polyclonal antisera, whereas the antisense (−) construct was not immunologically recognized. These results indicate that the 23K ORF encodes the viral coat protein.

Identification of a PLRV 3′ subgenomic RNA produced in virus-infected plants. Comparison of a 5′- versus a 3′-hybridization probe in northern blot hybridization analyses of total RNA isolated from virus-infected plants resulted in the identification and mapping of a PLRV subgenomic RNA. As shown in Figure

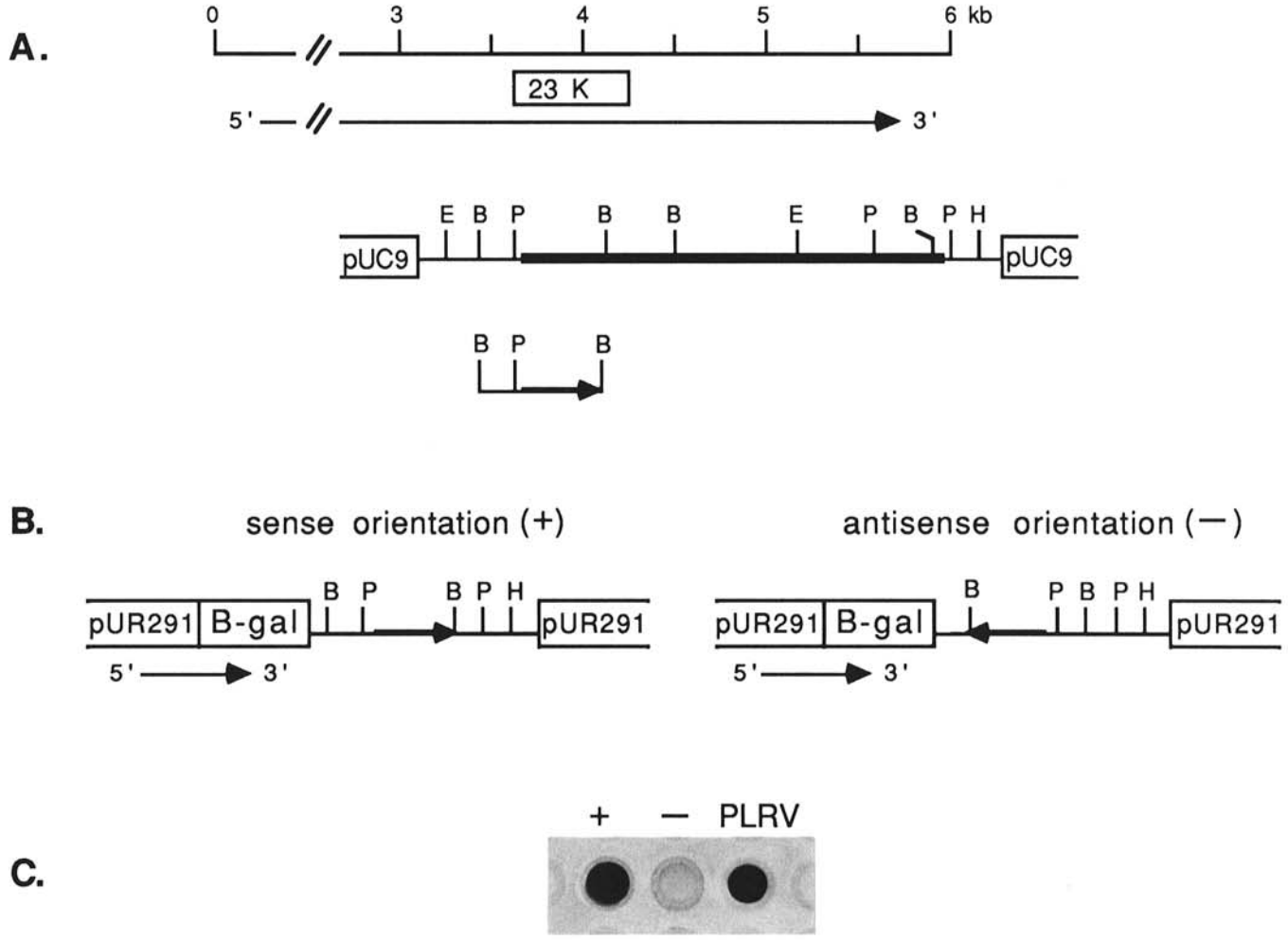

Fig. 3. Confirmation that the 23K open reading frame (ORF) encodes the potato leaf roll virus (PLRV) coat protein. The 5′-terminal *Bam*HI fragment of pPLRV4-228 containing amino acids 15-195 of the 23K ORF (**A**) was subcloned in the sense (+) and antisense (−) orientation into pUR291, a β-galactosidase translational fusion vector (**B**). After expression by *E. coli*, crude lysates were processed by a dot blot immunoassay using purified PLRV as a positive control. Immunological recognition of the sense (+) fusion by antisera to virions of PLRV (**C**) confirms that the 23K ORF encodes the coat protein. Restriction enzyme abbreviations are B (*Bam*HI), E (*Eco*RI), P (*Pst*I), and H (*Hin*dIII).

4, the 3′-hybridization probe detected two major species of viral RNA in extracts of virus-infected plants, representing genomic and subgenomic sizes of approximately 6.2 and 2.5 kb (lane 3), respectively, whereas the 5′-hybridization probe detected only the genomic-sized RNA (lane 1). The 3′-probe did not detect the subgenomic species in RNA isolated from purified virus (lane 5), indicating that this species is not encapsidated to detectable levels. These results indicate that PLRV replication is characterized by the *de novo* synthesis of a 3′-positioned subgenomic RNA.

Amino acid sequence similarity of PLRV 3′ ORFs to other luteoviruses. Table 1 summarizes the results of comparing the deduced amino acid sequence of each 3′ ORF of PLRV-4 in pairwise combination with the corresponding 3′ ORF of the FL-1 isolate of beet western yellows virus (BWYV) and the PAV-isolate of barley yellow dwarf virus (BYDV). These results show that the amino acid sequences of the three PLRV ORFs are more similar to BWYV than to BYDV and that the coat protein shares the highest amino acid similarity of the three ORFs.

DISCUSSION

Sequence analysis of 2,361 nucleotides of cloned cDNA indicates that the 3′ portion of the PLRV genome is organized into three ORFs, two of which overlap in different frames, followed by a terminal 141-nucleotide noncoding region. The calculated molecular weight of the protein potentially encoded by the 23K ORF is in good agreement with the reported molecular weight for PLRV coat protein (22), suggesting that the 23K ORF encodes this viral gene product. This hypothesis was confirmed by the construction and immunological testing of a β-galactosidase fusion protein. *E. coli* expression of a major portion of this ORF and confirmation of the expressed protein as PLRV coat protein-specific by a dot blot immunoassay (Fig. 3) confirms that this ORF encodes PLRV coat protein.

As shown in Figure 1, the physical arrangement of PLRV 3′ ORFs is characterized by an internal 23K coat protein ORF that contains the complete overlap of a smaller 17K ORF and is followed in-frame by a larger 56K ORF. Similar results have been reported recently for two other PLRV isolates (7,13). Our northern blot hybridization experiments have identified a 2.5-kb subgenomic RNA. This PLRV subgenomic RNA is produced upon infection, maps to the 3′ portion of the genome, and is not encapsidated (Fig. 4). Other minor hybridizable RNAs were detected by both probes in our northern blot hybridization experiments (Fig. 4); however, we conclude that these RNAs are likely artifacts of gel electrophoresis. By analogy to tobacco mosaic virus (18), these minor hybridizable RNAs probably result from partial degradation products of the two major hybridizable RNAs (6.2 kb and 2.5 kb) 'stacking' with plant ribosomal RNA during electrophoresis. By analogy to the 3′ subgenomic RNAs of other plant viruses (6), the 2.5-kb subgenomic RNA is the likely mRNA for expression of the 3′ portion of the viral genome. The 3′ organization of the PLRV genome is similar but not identical to two other luteoviruses, BWYV-PAV (14) and BWYV-FL1 (29). The notable difference is that a fourth 3′-terminal 6.7K ORF and a second subgenomic RNA have been reported for BYDV-PAV (14). We have not identified a similar ORF or subgenomic RNA for PLRV-4 nor has this ORF been reported for BWYV-FL1 (29) or a Scottish isolate of PLRV (13). These results indicate that there is some variation in the 3′ genome organization of luteoviruses and that PLRV is more closely related to BWYV than to BYDV. The latter is supported by the greater amino acid similarity of PLRV ORFs to BWYV ORFs (Table 1).

The roles of the potential proteins encoded by the 17K overlapping ORF and 56K in-frame ORF remain to be investigated. Similarly sized and arranged ORFs have been reported for the 3′ regions of BYDV-PAV (14) and BWYV-FL1 (29), suggesting that these ORFs encode functional proteins. There is evidence to suggest that the 17K ORF encodes a genome-linked protein for BYDV-PAV (14,15); however, the reported molecular weight for the PLRV genome-linked protein is 7 kDa (12). As discussed for BYDV-PAV (14) and by analogy to other RNA plant viruses (6), the 56K ORF is likely expressed by read-through translation of the coat protein (23K) ORF. The predicted size of this product would be 79 kDa. Future work should emphasize the identification of these putative viral proteins. Waterhouse et al (30) have recently reported that the read-through protein of

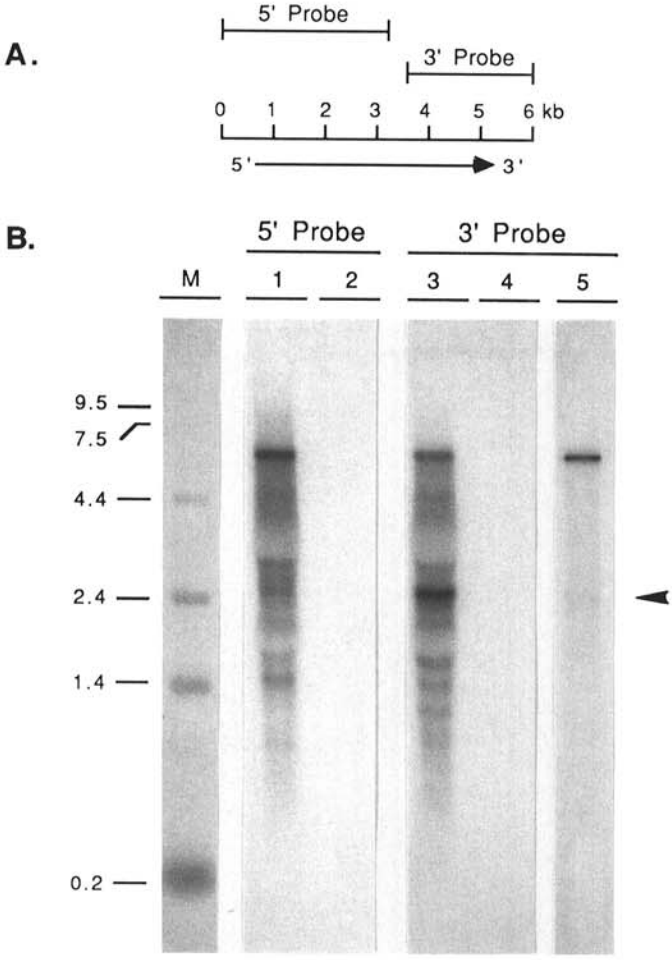

Fig. 4. Identification of a 3′, 2.5-kb, viral subgenomic RNA in extracts of potato leaf roll virus (PLRV)-infected plants. **A,** Location of the 5′ cDNA probe (pPLRV4-173) and 3′ cDNA probe (pPLRV4-228) relative to the PLRV RNA genome (from 26). **B,** Northern blot hybridization analyses using nick-translated probes. Lane M, 4 μg of BRL RNA ladder (in kb) hybridized to lambda DNA. Lanes 1 and 3, 40 μg of total RNA from PLRV-infected *Physalis floridana* plants 2-wk postinoculation. Lanes 2 and 4, 40 μg of total RNA from PLRV-free plants. Lane 5, 0.2 μg of RNA from purified PLRV. The arrow marks the location of the 2.5 kb subgenomic RNA in lane 3.

TABLE 1. Amino acid sequence similarity of potato leaf roll virus (PLRV, isolate 4) 3′ open reading frames (ORFs) to the corresponding ORFs of two other luteoviruses, beet western yellows virus (BWYV, isolate FL1) and barley yellow dwarf virus (BYDV, isolate PAV)

	Amino acid sequence similarity of corresponding ORF[a]	
PLRV-4 3′ ORF	BWYV-FL1[b]	BYDV-PAV[c]
Coat protein	64% (129/201)	45% (90/198)
17K	50% (73/146)	35% (46/133)
56K	42% (193/457)	27% (118/434)

[a]Determined by BESTFIT analysis (5) using gap and length weights of 5.0 and 0.3, respectively. Numerator is the number of identically matched amino acids, denominator is the number of totally matched amino acids.
[b]Data for comparison from Veidt et al (29).
[c]Data for comparison from Miller et al (14).

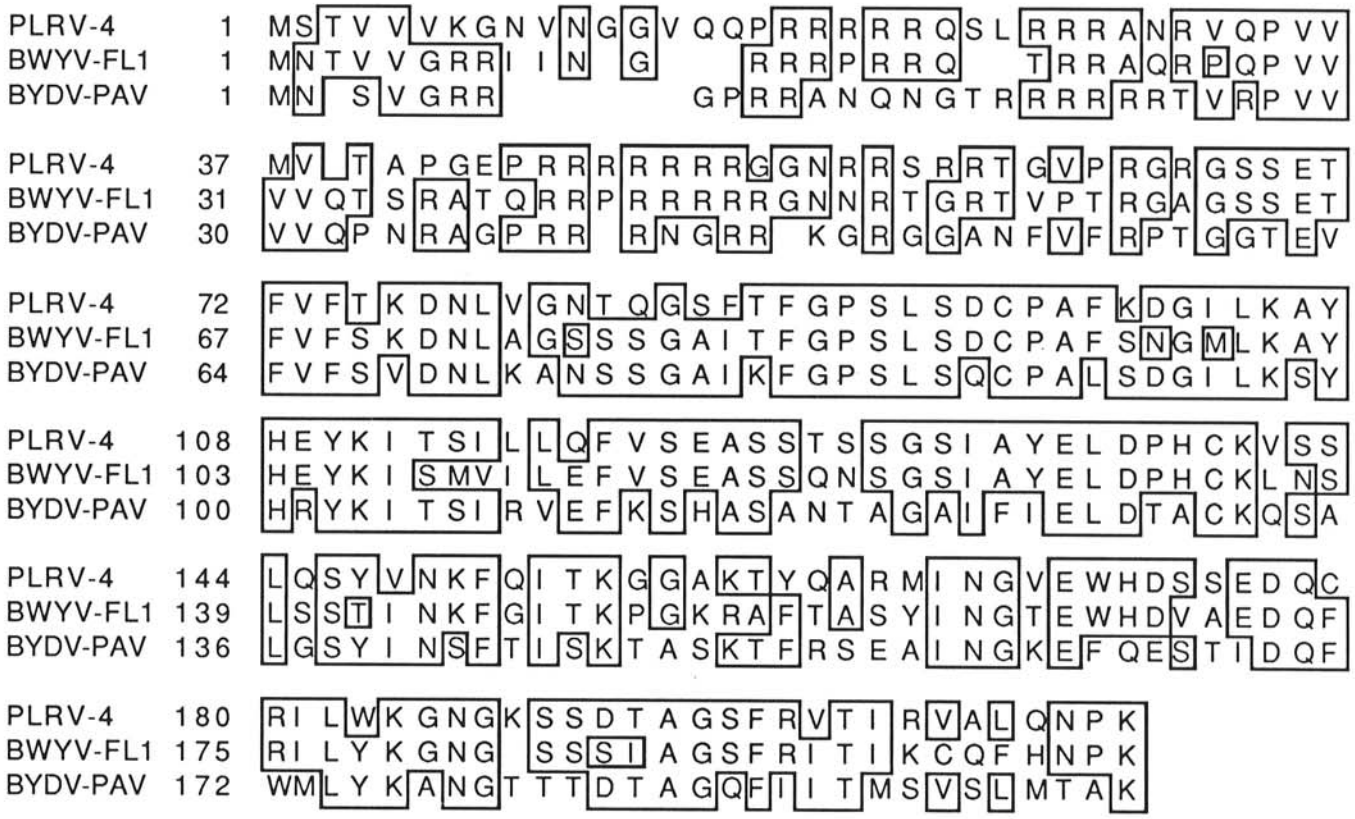

Fig. 5. Alignment of the amino acid sequences of the coat proteins of PLRV (isolate 4, 208 amino acids), BWYV (isolate FL1, 202 amino acids), and BYDV (isolate PAV, 200 amino acids). The BWYV-FL1 and BYDV-PAV sequences are from Veidt et al (29) and Miller et al (14), respectively. Sequences were computer-aligned by the algorithm of Needleman and Wunsch (16) using the methods and default parameters of Martinez (10). Identical amino acids are "boxed" to show similarities.

BYDV-PAV is present on the surface of virions and speculate that it plays a role in aphid transmission.

The coat protein amino acids of PLRV-4, BWYV-FL1, and BYDV-PAV are aligned in Figure 5. This alignment is composed of 172 three-way matched amino acids, exclusive of the initiating methionine. The middle third of the alignment, corresponding to PLRV amino acids 76–136, shows the greatest similarity. This region is characterized by 58 three-way matched amino acids containing 30 identical matches, representing an amino acid similarity of 52% for all three viruses. By similar analyses, the amino third and carboxyl third of the alignment are characterized by amino acid similarities of 33% (19/57) and 40% (23/57), respectively.

A comparison of the PLRV-4 nucleotide sequence data with BWYV-FL1 and BYDV-PAV indicates that the coat protein stop codon (*TAG*) of all three viruses is contained within a conserved 14-nucleotide sequence (CCAAA*TAG*GTAGAC), suggesting a common mechanism for the potential read-through translation of the coat protein gene. Knowledge of this conserved sequence may have practical applications to the study of luteoviruses. Based on this information, an oligonucleotide complementary to this conserved sequence could be used to clone other luteovirus coat protein genes or the restriction enzyme *Acc*I, which recognizes the hexanucleotide sequence GTAGAC (9), could be used to screen a library for putative luteovirus coat protein cDNAs.

LITERATURE CITED

1. Azad, A. A., Fahey, K. J., Barret, S. A., Erny, K. M., and Hudson, P. J. 1986. Expression in *Escherichia coli* of cDNA fragments encoding the gene for the host-protective antigen of infectious bursal disease virus. Virology 149:190-198.
2. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294-5299.
3. D'Arcy, C. J., Hewings, A. D., Burnett, P. A., and Jedlinksi, H. 1983. Comparative purification of two luteoviruses. Phytopathology 73:755-759.
4. Davis, L. G., Dibner, M. D., and Battey, J. F. 1986. Basic Methods in Molecular Biology. Elsevier, New York. 388 pp.
5. Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.
6. Goldbach, R. W. 1986. Molecular evolution of plant RNA viruses. Annu. Rev. Phytopathol. 24:289-310.
7. Kawchuk, L. M., Martin, R. R., Rochon, D. M., and McPherson, J. 1989. Identification and characterization of the potato leafroll virus putative coat protein gene. J. Gen. Virol. 70:783-788.
8. Logemann, J., Schell, J., and Willimitzer, L. 1987. Improved method for the isolation of RNA from plant tissues. Anal. Biochem. 163:16-20.
9. Maniatis, T., Fritsch, E. F., and Sambrook, J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 548 pp.
10. Martinez, H. M. 1988. A flexible multiple sequence alignment program. Nucleic Acids Res. 16:1683-1691.
11. Matthews, R. E. F. 1982. Classification and nomenclature of viruses. Fourth Report of the International Committee on the Taxonomy of Viruses. Intervirology 17:140-141.
12. Mayo, M. A., Barker, H., Robinson, D. J., Tamada, T., and Harrison, B. D. 1982. Evidence that potato leafroll virus is positive-stranded, is linked to a small protein and does not contain polyadenylate. J. Gen. Virol. 59:163-167.
13. Mayo, M. A., Robinson, D. J., Jolly, C. A., and Hyman, L. 1989. Nucleotide sequence of potato leafroll virus RNA. J. Gen. Virol. 70:1037-1051.
14. Miller, W. A., Waterhouse, P. M., and Gerlach, W. L. 1988. Sequence and organization of barley yellow dwarf virus genomic RNA. Nucleic Acids Res. 16:6097-6111.
15. Murphy, J. F., Clark, J. M., and D'Arcy, C. J. 1987. Barley yellow dwarf virus-RPV-IL RNA has a protein covalently linked to its 5′ terminus. (Abstr.) Phytopathology 77:1705.
16. Needleman, S. B., and Wunsch, D. C. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48:443-453.
17. Okita, T. W., and Greene, F. C. 1982. Wheat storage proteins: Isolation and characterization of the gliadin messenger RNAs. Plant Physiol. 69:834-839.

18. Palukaitis, P., Garcia-Arenal, F., Sulzinski, M. A., and Zaitlin, M. 1983. Replication of tobacco mosaic virus. VII. Further characterization of single- and double-stranded virus-related RNAs from TMV-infected plants. Virology 131:533-545.
19. Peters, D., and Jones, R. A. C. 1981. Potato leafroll virus. Pages 68-70 in: Compendium of Potato Diseases. W. J. Hooker, ed. American Phytopathological Society, St. Paul, MN.
20. Rich, A. E. 1983. Potato Diseases. Academic Press, New York. 238 pp.
21. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. 113:237-251.
22. Rowhani, A., and Stace-Smith, R. 1979. Purification and characterization of potato leafroll virus. Virology 98:45-54.
23. Ruther, U., and Muller-Hill, B. 1983. Easy identification of cDNA clones. EMBO J. 2:1791-1794.
24. Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
25. Smith, G. E., and Summers, M. D. 1979. Restriction maps of five *Autographa californica* MNPV variants, *Trichoplusia ni* MNPV, and *Galleria mellonella* MNPV DNAs with endonucleases *Sma* I, *Kpn* I, *Bam* HI, *Sac* I, *Xho* I, and *Eco* RI. J. Virol. 30:828-838.
26. Smith, O. P., Harris, K. F., Toler, R. W., and Summers, M. D. 1988. Molecular cloning of potato leaf roll virus complementary DNA. Phytopathology 78:1060-1066.
27. Smith, O. P., Storch, R. H., Hepler, P. R., and Manzer, F. E. 1984. Prediction of potato leafroll virus disease in Maine from thermal unit accumulation and an estimate of primary inoculum. Plant Dis. 68:863-865.
28. Takanami, Y., and Kubo, S. 1979. Enzyme-assisted purification of two phloem-limited plant viruses: Tobacco necrotic dwarf and potato leafroll. J. Gen. Virol. 44:153-159.
29. Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K., and Jonard, G. 1988. Nucleotide sequence of beet western yellows virus RNA. Nucleic Acids Res. 16:9917-9932.
30. Waterhouse, P. M., Martin, R. R., and Gerlach, W. L. 1989. BYDV-PAV virions contain readthrough protein. (Abstr.) Phytopathology 79:1215.
31. Yanisch-Peron, C., Vieira, J., and Messing, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.