

Nucleotide sequence and organization of potato leafroll virus genomic RNA

Frank van der Wilk^{*,†}, Marianne J. Huisman, Ben J.C. Cornelissen, Harm Huttinga[†] and Rob Goldbach^{*}

^{*}*Agricultural University Wageningen, Department of Virology, PO Box 8045, 6700 EM Wageningen, [†]Research Institute for Plant Protection, PO Box 9060, 6700 GW Wageningen and MOGEN International NV, Einsteinweg 97, 2333 CB Leiden, The Netherlands*

Received 18 January 1989

The nucleotide sequence of the genomic RNA of potato leafroll virus was determined and its genetic organization deduced. The RNA is 5882 nucleotides long and contains 6 open reading frames (ORFs) encoding proteins of 70, 70, 56, 28, 23 and 17 kDa. The putative genes for the coat protein (23 kDa) and the RNA-dependent RNA polymerase (70 kDa) were identified by interval amino acid sequence homologies. For expression of the different ORFs, translational frameshift and readthrough events are proposed.

cDNA cloning; Nucleotide sequence; Amino acid sequence homology; RNA virus, plant positive-strand; (Potato leafroll virus)

1. INTRODUCTION

Potato leafroll virus (PLRV), a member of the luteovirus group [1], is a virus of great economic importance which infects potato plants worldwide [2]. Luteoviruses have isometric capsids containing a single-stranded messenger-sense RNA genome [3–5]. The PLRV RNA contains a genome-linked protein (VPg) but lacks a polyadenylate sequence at the 3'-end [6]. Recently, the complete nucleotide sequences of the genomic RNAs of barley yellow dwarf virus (BYDV) and beet western yellow virus (BWYV), two other

luteoviruses, have been determined [7,8]. So far, for PLRV cDNA physical maps [9,10] and the sequence of the 3'-terminal 141 nucleotides have been published [10]. Here, we report the complete nucleotide sequence of the PLRV genomic RNA with comparisons of its deduced amino acid sequences with those of other plant virus proteins.

2. EXPERIMENTAL

2.1. Materials

Enzymes were purchased from Bethesda Research Laboratories (BRL) and Amersham. Chemicals were obtained from Sigma and radiochemicals from Amersham. Computer programs used were from the Genetics Computer Group of the University of Wisconsin (UWCGC) (Madison, WI).

2.2. Virus and RNA isolation

PLRV (strain Wageningen) was transmitted using the aphid *Myzus persicae* and propagated in *Physalis floridana* plants. Virus was purified as follows: after thorough homogenization in PCA buffer (0.18 M phosphate/citric acid buffer, pH 7.0) containing 10 mM sodium thioglycolate and 10 mM sodium dithiocarbamate, the extract was treated with chloroform/butanol (1:1, v/v). The virus was precipitated from the aqueous phase by addition of 8% (w/v) polyethylene glycol, resuspended in PCA buffer and recollected by ultracentrifugation.

Correspondence address: F. van der Wilk, Agricultural University Wageningen, Department of Virology, PO Box 8045, 6700 EM Wageningen, The Netherlands

Abbreviations: BWYV, beet western yellows virus; BYDV, barley yellow dwarf virus; CarMV, carnation mottle virus; CPMV, cowpea mosaic virus; ORF, open reading frame; PLRV, potato leafroll virus

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under accession no. Y07496

trifugation. This washing procedure was repeated once. Virus was further purified on a 5–45% linear sucrose gradient. Viral RNA was isolated from purified virus by repeated phenol extractions in the presence of 0.5% SDS. The RNA was ethanol precipitated and dissolved in sterile water.

2.3. DNA sequence analysis

Overlapping restriction fragments from cDNA clones were subcloned into M13 tg130/131 vectors [11] and sequenced using the dideoxy method [12] with [α -³⁵S]thio-dATP as radioactive nucleotide.

2.4. RNA sequence analysis

Direct dideoxy sequence determination on genomic RNA was performed as described [13].

3. RESULTS AND DISCUSSION

The nucleotide sequence of PLRV genomic RNA was unraveled by sequence determination of overlapping cDNA clones and by direct sequencing of the RNA. For first strand cDNA synthesis reverse transcription on the viral genomic RNA was initiated by random priming with calf thymus DNA fragments or by priming with specific synthetic oligonucleotides. To obtain cDNA clones representing 3'-terminal sequences of the genomic RNA, a 15-base oligonucleotide complementary to the extreme 3'-terminus [10] was used. Double-stranded cDNA was obtained using the RNase H method [14] and cloned in either pUC19 or λ gt10 vectors. The clones selected from these libraries for sequence determination spanned approx. 99% of the viral genome (fig.1A). The 5'-proximal nucleotides were elucidated by direct dideoxy sequencing on PLRV genomic RNA by extension of an oligonucleotide complementary to nucleotides 98–115. The sequence could be determined unequivocally up to the first base shown in fig.2. The VPg covalently attached to the 5'-terminal nucleotide probably obscured two additional bases from our scrutiny: they were observed as strong stops in all lanes of the sequence ladder. The total length of the sequence elucidated is 5882 nucleotides.

The coding regions comprise six large open reading frames (ORFs) (fig.1) of which the amino acid sequences are noted above the PLRV genomic RNA (fig.2). ORF1 starts at the first AUG codon (position 70), terminates with a UGA stop codon (position 811) and could encode a product of 28127 Da. ORF1 shows no homology with any of the different ORFs of BYDV [7], BWYV [8] or of

other published plant viruses. The second ORF overlaps ORF1 by a start at position 203 in a different phase from ORF1 and it stops at the UGA codon present at position 2120. The putative ORF2 product is of 69 674 Da. It shows homology with the putative ORF2 product of BWYV (54%, calculated with the UWGCG program GAP) (fig.1). PLRV ORF3 is proposed to start at position 1540, overlapping the second ORF, and to terminate at the UGA stop codon at position 3388. ORF3 encodes a protein of 69 622 Da. For BYDV ORF2 and BWYV ORF3, the evidence suggests that these ORFs are expressed by translational frameshift. On the basis of the strong amino acid sequence homology, we propose that the PLRV ORF3 is expressed via translational frameshift like the BYDV and BWYV analogous ORFs. PLRV ORF3 contains the amino acid sequence motif ---^S/_TG---T---N^S/_T 18–37 aa GDD--- which is found in all RNA-dependent RNA polymerases of RNA plant viruses sequenced to date [15,16], and therefore the ORF3 product is the most likely candidate to represent the PLRV-encoded RNA-dependent RNA polymerase. The predicted amino acid sequence of ORF3 shows considerable homology with BWYV ORF3 (70%) and the putative RNA-dependent RNA polymerase of southern bean mosaic virus (43%) [17]. Strikingly, ORF2 of BYDV showed no homology with the PLRV ORF3 but, instead, showed homology with the putative RNA polymerase of CarMV [7]. The fourth PLRV ORF is separated from ORF3 by a non-coding region of 197 nucleotides. ORF4 spans positions 3588–4212 (UAG), hence encoding a protein of 23233 Da. PLRV ORF4 showed considerable homology with BYDV ORF3 (57%) and BWYV ORF4 (72%). Since BYDV ORF3 has been shown to correspond to the BYDV coat protein [18], the conclusion seems justified that ORF4 of both PLRV and BWYV encodes the respective viral coat proteins. The strong homology in amino acid sequences (fig.3) can explain the serological cross-reactivity between the different luteoviruses. The fifth ORF (underlying ORF4 from position 3613 to the UGA codon at position 4081) encodes a protein of 17344 Da which exhibits amino acid sequence homology to the products of the analogous ORFs (fig.1) of BWYV (72%) and BYDV (57%). For BYDV and BWYV it has been proposed that the ORFs, underlying the coat pro-

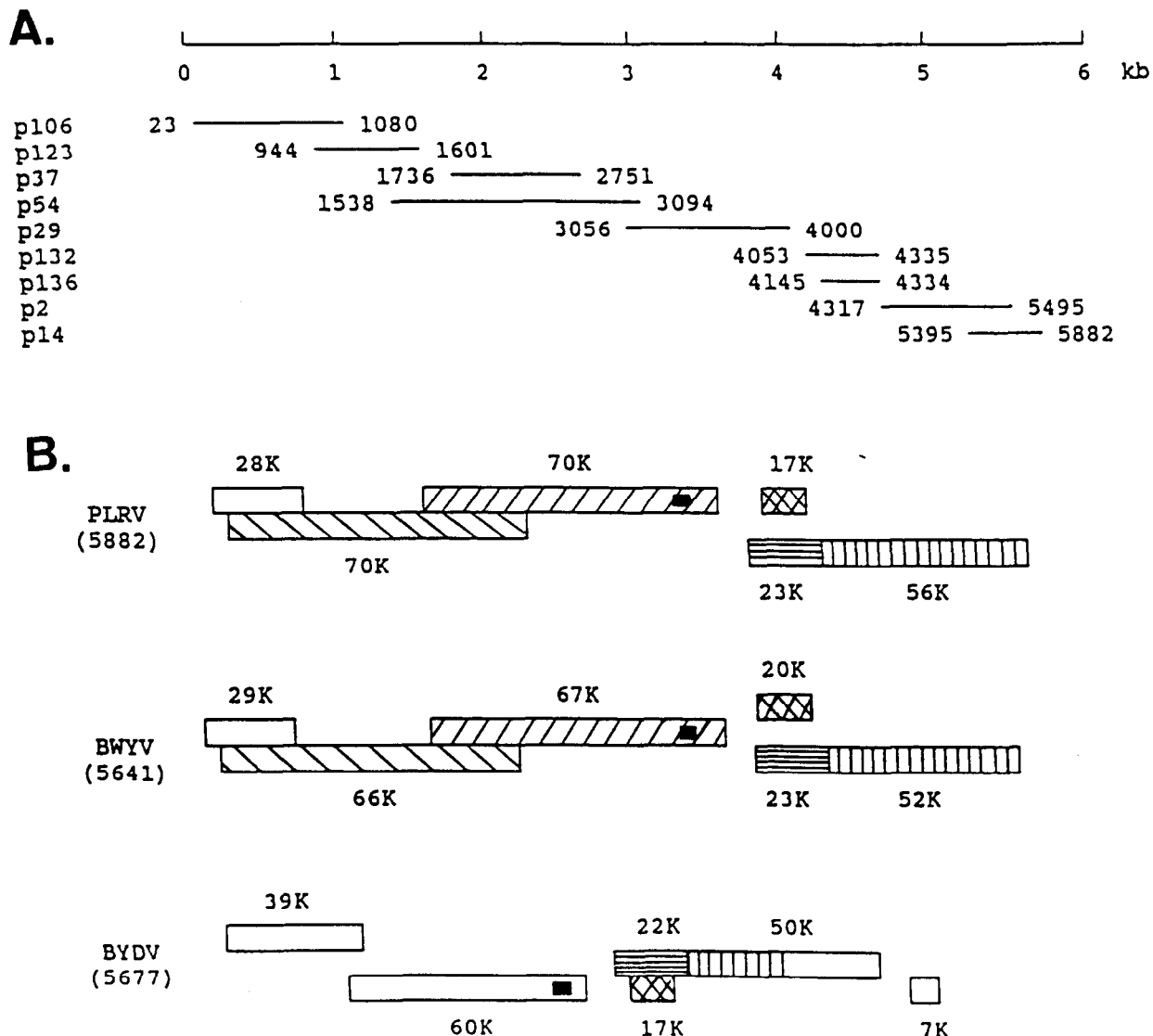


Fig.1. (A) Map of cDNA clones used to elucidate the nucleotide sequence of the PRLV RNA (p2, p29, p37, p54 represent pUC19 clones; p14, p123, p136 represent λ gt10 clones). (B) Schematic representation of the organization of the PRLV genome and comparison of the open reading frames (ORFs) in the genomic RNAs of PRLV, BYDV [7] and BWYV [8]. The size in kDa (K) of the proteins encoded by each ORF is shown. Similar shading indicates regions of high amino acid sequence homology among the ORFs of the different viruses. (■) GDD sequence motif found in plant RNA-dependent RNA polymerases (see text).

tein ORF, code for the respective VPgs. However, the VPg of PRLV has been estimated to have a molecular mass of 7 kDa only [6], whereas the coding capacity of ORF5 predicts a protein of 17 kDa. Possibly, this ORF encodes a VPg-precursor molecule from which, at the onset of RNA synthesis, the VPg molecule is released, as has been suggested for CPMV [19]. PRLV ORF6

is contiguous with the putative coat protein cistron (ORF4), separated only by the amber stop codon of ORF4. The same situation is observed for ORF5 of BYDV and ORF6 of BWYV, which show homologies to PRLV ORF6 of 45 and 59%, respectively. The sequences flanking the amber stop codons are identical in all three viruses (CCAAAUAGGUAGAC). Amber stop codons

[illegible]

4341 D Q I I L G S L G S O R M K Y I E D E N Q N Y T
 G A C C A A U C A U A U U G G U U C U A G G G A G C C A A G G A U G A A A U A U A G A G G A C G A G A A C C A G A A C U A U A
 4411 N V S S E Y Y S Q S S H Q A V P M Y Y F N V P
 C A A A U G U A U U C U G A G U A U A U C U C A A U C G A G C A U G A A G C C C U A U G U A U A U C U A A U G U C C C
 4481 K G Q W S V D I S C E G Y Q P T S S T S D P N
 G A A G G G C A A U G G U C A G U C G A C A U C A G C U G G A G G G U A A C C C A C U A G C A G C A C C U C G G A U C C A A C
 4551 R G R S D G M I A Y S N A D S D Y W N V G E A D
 C G G G U A G G A G U G A C G G G A U G C G G U A U C A A C C G G A U C C G A U A U G G A A U G G U G G A A G C G G
 4621 G V K I S K L R N D N T Y R Q G H P E L E I N
 A U G G U G A A A A A U U C G A A G C U A C G C A A C G A U A A C A C C U A C C G C C A A G G U C A C C C A G A A C U G A A A U A A
 4691 S C H F R E G G Q I L L E R D A T I S F H V E A P
 C U C G U G C A U A U U C G G A G G G C A C U C U A G A C G G A C G C U A C A U A G C U U C C A C G U G A A G C G C C U
 4761 T D G R F F L V G P A I Q K T A K Y N Y T I S Y
 A C U A G G G C G A U C U U C U C G U G G U C C G C U A U C C A G A A A A C C G A A A A U A C U A U A C U A U C U A U
 4831 G D W T D R D M E L G L I T V V L D E H L E G
 A C G G U G C U G A C G G A C G A G A C U G G A A C U G G G C U A G A C C G U G G U C U G A A A C A A C U A U A G A A G G
 4901 T G S A N R V R R P P R E G H T Y M A S P R E
 C A C U G G U C G G U A A C A G A G U G C G G C G C C C C A C G G A G G G C C A C A C U A U A U G G C U C G C C G C G G A A
 4971 P E G K P V G N K P R D E T P I Q T Q E R Q P D
 C C G G A G G A A A C C G G U G G A A A A C C A A G G A G A A A C C C G A U A C A A C G A A A C C G A A A G A A C A A C C U G
 5041 Q T P S D D V S D A G S V N S G G S T E S L Q
 A U C A A A C U C G C U C A C G A C G A U A U C C G A U C U G G U C G U A A A C G C G G C G C A C U A C U A G A G U C G C A
 5111 L E F G A N S D S T Y D A T V D G T D W P R I
 A U G G A G U U C G G G C A A A U C A G A U A G U A C U A C G A U C A C A G U C G A U G G U A C A G A C U G G C C A G A A U U

5181 P P P R H P P E P R V S G N S R T V T D F S P K
 C C U C C A C A A G G C A C C C A C C U G A A C C U A G A G U U C C G G C A A U C A A G A A C U G U A C U G A C U U U C C G A
 5251 A D L L E N W D A K H F D P G Y S K E D V A A
 A A G C C G A U C A U U G G A A U U G G A U G C C A A A C A C U C G A C C C U G G U A U U C C A A G A A G A U G C G C U G C
 5321 A T I I A H G S I Q D G R S M L E K R E E N Y
 U G C U A C U A U A U A G C G C A G G C A G U A U C A G A U G G C G A G U A U G U A G A G A G A G A G A A A A U G C
 5391 K N K T S S W K P P L P K A V S P A I A K L R S
 A A G A C A A A A C C U C C U C G A A G C C C C C U A U A C C U A A G C G G U G A G C C A C C A U A G C C A A A U G C G C U
 5461 I R K S Q P L E G G T L K K D A T D G V S S I
 C G A U U C G A A A U C C A A C C C U C G A G G A G G A C C C U A A G A A A G A G C C A C U A G G U G U C A U C A U A U
 5531 G S G S L T G G T L K R K E T I E E R L L Q T
 U G G C A G G U U C U A C A C G U G G C A C C U A A G A G A A G A A C U A U G A A G C G U U A C U G C A G A C C
 5601 L T T E Q R L W Y E N L K K T N P L A A T Q W L
 U U A C A C A C U A A A A G G C U G U G U A C G A G A A U U G A A A A A C U A A C C C U U A G C U G C U A C C A A U G G C
 5671 F E Y Q P P P Q V D R N L A E K P F Q G R K *
 U G U U G A U A U C A G C C A C C C C A A G G U G A U A A C U A G C U G A A A G C C A U C C A A G G A G G A A A U G
 5741 A G U C G A C A C A C U A A A A C U G A G U C C G C C G C A C U A U A G C G G A A C G A A A G C G A A A G G U A U A G G
 5811 C U C U A C G C C U G C U A G A G A C C G U C G A A A G A C C G C A C U G U A C C A G G A U C C U A C A G G U G U A
 5881 GU 5882

Fig.2. The nucleotide sequence of PLRV RNA. The deduced amino acid sequences of long open reading frames are depicted over the nucleotide sequence.

BYDV CP 1	M N S V - G R R G P R R A N Q N G T R R R R R R T V R - - - - - P V V V V Q P N R
PLRV CP 1	M S T V V V K G N V N G G V Q Q - P R R R R R Q S L R R R A N R V Q P V V M V K A P G
BWV CP 1	M N T V V G R R I I N G - - - - - R R R P R R Q T - - R R A Q R P Q P V V V V Q T S R
BYDV CP	A G P R R R N G R R K G R G G A N F - - - - - V F R P T G G T E V F V F S V D N L K A
PLRV CP	- - Q P R R R R R R - - R G G N R R S R R T G V P R G R G S S E T F V F T K D N L M G
BWV CP	A T Q R R P R R R R - - R G N N R T G R - T V P T R G A G S S E T F V F S K D N L A G
BYDV CP	N S S G A I K F G P S L S Q C P A L S D G I L K S Y H R Y K I T S I R V E F K A S H A
PLRV CP	N S Q G S F T F G P S L S D C P A 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
BWV CP	S S S G A I T F G P S L S D C P A F S N G M L K A Y H E Y K I S M V I L E F V S E A S
BYDV CP	S A N T A G A I F I E L D T A C K Q S A L G S Y I N S F T I S K T A S K T F R S I T M
PLRV CP	S T S S G S I A Y - E L D P H C K V S S L Q S Y V N Q F Q I P Q G G A K T Y Q A R M -
BWV CP	S Q N S G S I A Y - E L D P H C K L N S L S S T I N K F G I T K P G K R A F T A S Y -
BYDV CP	S V S L M T A E A I N G K E F Q E S T I D Q F W M L Y K A N G T T T D T A G Q F I K *
PLRV CP	- - - - - I N G V E W H D S S E D Q C R I L W K G N G K S S D T A G S F R V T
BWV CP	- - - - - I N G T E W H D V A E D Q F R I L Y K G N G S S S - I A G S F R I T
BYDV CP	
PLRV CP	I R V A L Q N P K *
BWV CP	I K C Q F H N P K *

Fig.3. Amino acid sequence comparison between the coat protein molecules of BYDV, PLRV and BWV. Direct amino acid homologies are boxed.

are renowned for their leakiness [20]. Moreover, for BWYV, evidence was presented that read-through of ORF4/6 occurs [8]. Therefore, it seems likely that ORF6 is expressed via translational readthrough, which would result in a fusion product of 79709 Da. ORF6 itself has a coding capacity for a protein of 56476 Da. The ORF6 stop codon UGA is located at position 5739.

The non-coding sequences of PLRV consist of 406 nucleotides, being 6.9% of the sequence. The minus strand does not encode ORFs of considerable length.

The PLRV genome as determined is 5882 nucleotides in length, whereas both BYDV (5677 nucleotides) and BWYV (5641 nucleotides) are smaller. The genomic organization of PLRV appears to be similar to that of BWYV and to a lesser extent to that of BYDV (fig.1B). Considering the homology between the different luteoviruses it seems that PLRV is more closely related to BWYV than to BYDV. It is suggested that BYDV and BWYV might be considered as distinct subtypes of the luteovirus group. In view of the homologies observed between the ORFs of PLRV and BWYV, PLRV appears to be part of the BWYV subgroup. However, differences between PLRV and BWYV, e.g. the nonhomologous ORF1s, make PLRV and BWYV distinctly different members.

Acknowledgements: The authors wish to thank Peter de Haan and Ves Jacobs for help in the initial experiments and Dr Mike Mayo for the exchange of data prior to publication.

REFERENCES

- [1] Matthews, R.E.F. (1982) Intervirology 17, 140–141.
- [2] Rochow, W.F. and Duffus, J.E. (1981) in: Handbook of Plant Virus Infections and Comparative Diagnosis (Kurstak, E. ed.) pp.147–170, Elsevier/North-Holland, Amsterdam.
- [3] Rowhani, A. and Stace-Smith, R. (1979) Virology 98, 45–54.
- [4] Takanami, Y. and Kubo, S. (1979) J. Gen. Virol. 44, 853–856.
- [5] Murphy, J.F., Clark, J.M. and D'Arcy, C.J. (1987) Phytopathology 33, 1705–1709.
- [6] Mayo, M.A., Barker, H., Robinson, D.J., Tameda, T. and Harrison, B.D. (1982) J. Gen. Virol. 59, 163–167.
- [7] Miller, W.A., Waterhouse, P.M. and Gerlach, W.L. (1988) Nucleic Acids Res. 13, 6097–6111.
- [8] Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guillely, H., Richards, K. and Jonard, G. (1988) Nucleic Acids Res. 16, 9917–9932.
- [9] Smith, O.P., Harris, K.F., Toler, R.W. and Summers, M.D. (1988) Phytopathology 78, 1060–1066.
- [10] Prill, B., Maiss, E., Chansilpa, N. and Casper, R. (1988) J. Gen. Virol. 69, 2397–2402.
- [11] Kieny, M.P., Lathe, R. and Lecocq, J.P. (1983) Gene 26, 91–99.
- [12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
- [13] Huisman, M.J., Linthorst, H.J.M., Bol, J.F. and Cornelissen, B.J.C. (1988) J. Gen. Virol. 69, 1789–1798.
- [14] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
- [15] Goldbach, R.W. (1986) Annu. Rev. Phytopathol. 24, 289–310.
- [16] Cornelissen, B.J.C. and Van Vloten-Doting, L. (1989) in: Cell Culture and Somatic Cell Genetics of Plants, vol.7, Molecular Biology of Plant Nuclear Genes (Schell, J. and Vasil, I.K. eds) Academic Press, New York, in press.
- [17] Wu, S., Rhinehart, C.A. and Kaesberg, P. (1987) Virology 161, 73–80.
- [18] Miller, W.A., Waterhouse, P.M., Kortt, A.A. and Gerlach, W.L. (1988) Virology 165, 306–309.
- [19] Goldbach, R. and Van Kammen, A. (1985) in: Molecular Plant Virology (Davies, J.W. ed.) vol.II, pp.83–120, CRC, Boca Raton, FL.
- [20] Van Tol, R.G.L., Van Gemenen, R. and Van Vloten-Doting, L. (1980) FEBS Lett. 118, 67–71.