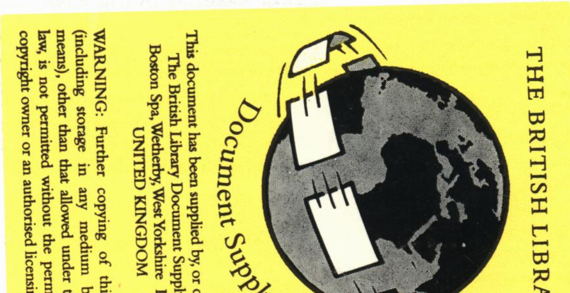
101 AGRO

101



VIROLOGY 202, 1012–1017 (1994)

Nucleotide Sequence of Cucurbit Aphid-Borne Yellows Luteovirus

H. GUILLEY,[1] CATHERINE WIPF-SCHEIBEL,* K. RICHARDS, H. LECOQ,* AND G. JONARD

*Institut de Biologie Moléculaire des Plantes du CNRS et de l'Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg Cédex, France; and *Station de Pathologie Végétale, INRA, Domaine Saint-Maurice, BP 94, 84143 Montfavet Cédex, France*

Received February 2, 1994; accepted March 22, 1994

The nucleotide sequence (5669 residues) of the genomic RNA of cucurbit aphid-borne yellows luteovirus (CABYV) is presented. Analysis of genome organization and sequence homologies indicate that CABYV is a member of luteovirus Subgroup 2 (other sequenced members: beet western yellows virus, potato leafroll virus, and barley yellow dwarf virus, RPV isolate) and appears to be most closely related to beet western yellows virus. © 1994 Academic Press, Inc.

Recent outbreaks of a severe yellowing disease of cultivated cucurbits in France have been attributed to a small isometric virus which is not mechanically transmitted, but which is efficiently transmitted in a persistent manner by the aphids *Myzus persicae* and *Aphis gossypii* (*1*). The virus, designated cucurbit aphid-borne yellows virus (CABYV), is widespread in cucurbits throughout the Meditereanean basin and has also been detected in Asia and Africa (H. Lecoq, unpublished observations) and in North America (*2*). Serological analysis and nucleic acid hybridization experiments, along with virion morphology and the aforesaid vector transmission studies, indicated that CABYV is a member of the luteovirus group (*1*).

Sequence comparisons (*3–5*) have made it clear that luteoviruses can be divided into two major subgroups, which have been referred to (*4*) as Subgroup 1 (sequenced members: barley yellow dwarf virus (BYDV), PAV, and MAV isolates (*6–7*) and Subgroup 2 (sequenced members: beet western yellows virus (BWYV (*3*)), potato leafroll virus (PLRV (*8–10*)) and BYDV, RPV isolate (*11*)). Recently, it has been suggested that soybean dwarf virus may belong to a third subgroup (*12*). Among the features distinguishing Subgroups 1 and 2 are the following: (i) Members of Subgroup 2 have a 5′-proximal open reading frame (ORF) encoding an approximately 29-kDa species, P1 (see Fig. 1A for ORF nomenclature), which is lacking in Subgroup 1, while the latter possess a short 3′-proximal ORF absent in the subgroup 2 members; (ii) the region of overlap between ORF 2 (P2) and ORF 3 in the Subgroup 2 viruses is much longer than in Subgroup 1 (Fig. 1A); (iii) the "core polymerase" domain on ORF 3 of the Subgroup 2 viruses displays clear sequence affinity with the polymerase domain of southern bean mosaic virus while the same region of the Subgroup 1 viruses is related to the polymerase domain of the carmoviruses (see Ref. *5* for a review); (iv) there is considerable sequence homology in coding regions among the members of each subgroup except for ORF 1 of Subgroup 2. Intersubgroup sequence homology also exists but is less extensive, essentially being limited to the cluster of genes nearer the 3′-terminus, i.e. the coat protein (ORF 4), ORF 5, and ORF 6. Finally, it should be noted that, despite the aforesaid differences, the members of Subgroups 1 and 2 have certain features in common with respect to gene expression, including (a) the expression of ORF 3 as a ORF 2–3 fusion protein (P2/3) due to translational frameshift in the region of overlap between ORF 2 and ORF 3 (*13–16*), (b) expression of the cluster of three 3′-terminal genes from a single subgenomic RNA (*17*), and (c) expression of ORF 6 as an ORF 4-6 fusion protein (RT protein) due to translational readthrough of the coat protein (ORF 4) termination codon (*3, 16, 17, 18*).

In this paper we describe the nucleotide sequence of 5669 residues for the genomic RNA from CABYV, obtained by characterization of cloned cDNA, and analyze its relatedness to other luteoviruses. Purified RNA from CABYV (isolate N) virions consists of a single RNA species of about 5600 nt, as judged by its mobility during electrophorsis under denaturing conditions in an agarose gel (data not shown). Sequence analysis of various luteoviruses has revealed highly conserved sequence domains in the coat protein cistron, making it possible to design luteovirus-specific primers as an aid to diagnosis and classification (*19, 20*). We have synthesized a 25-mer, 5′-ATTCATGGTAGG(C/A)CTTGAG(T/C)ATTCC (oligo 1), complementary to residues 3784–3808 of the final CABYV sequence. This region encodes the amino

Sequence data from this article have been deposited with the EMBL Data Library under Accession No. X76931.

[1] To whom reprint requests should be addressed.

0042-6822/94 $6.00
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.

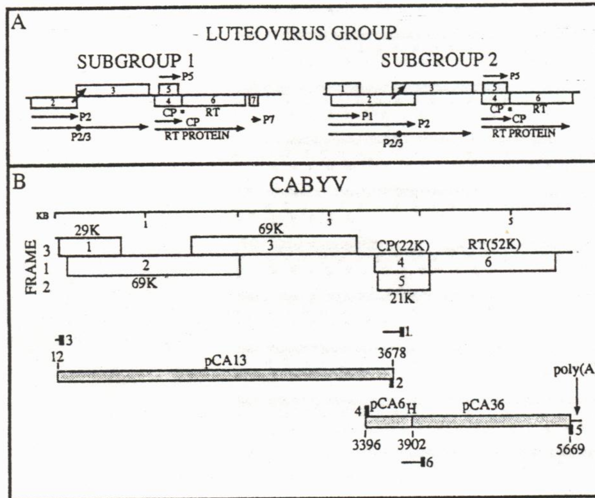


FIG. 1. Luteovirus genome organization. (A) Genetic organization of Subgroup 1 and Subgroup 2 luteoviruses. Open reading frames (ORFs) are indicated by hollow rectangles. ORFs in the map of the Subgroup 1 luteovirus (no ORF 1) have been numbered so that they correspond to the numbering of ORFs in Subgroup 2. The approximate site of frameshift between ORF 2 and ORF 3 is indicated by a slanted arrow and the approximate site of the resulting fusion between the ORF 2 and ORF 3 polypeptides (P2/3) is indicated by a black circle. The suppressible termination codon at the end of the coat protein (CP) cistron is indicated by an asterisk. RT, readthrough. (B) Position of long ORFs (hollow rectangles) in CABYV RNA (above) and the structure of cloned cDNA inserts (stipled rectangles) analyzed to determine the sequence (below). In the lower part of the figure, the numbered black squares (not to scale) represent some of the synthetic oligonucleotides used to prepare cDNA clones, to prime the polymerase chain reaction (PCR), or for direct sequence determination. In the latter case, the extent of sequence determined by primer extension is indicated by a horizontal line. The poly(A) tract enzymatically added to the 3′-end of the viral RNA and subsequently amplified by PCR using oligonucleotides 5 and 4 is represented by a thin line. H, *Hin*dIII.

acid sequence G(I/M)LK(A/S)YHE, which is perfectly or almost perfectly conserved in the coat protein of other Subgroup 2 viruses. Direct RNA sequence analysis using this primer to initiate reverse transcription (*21*) provided readable sequence of about 200 residues (Fig. 1B). For unknown reasons, attempts to produce and clone (*22*) long double-stranded cDNA fragments using oligo 1 to prime cDNA synthesis failed. However, use in the cloning protocol of another primer (oligo 2; complementary to residues 3655–3678) produced clones with suitably long cDNA inserts, one of which, pCA13 (Fig. 1B), was chosen for further analysis. The approximately 3700 residue cDNA insert of pCA13 was sequenced (*23*) on both strands principally by characterizing sets of 5′- and 3′-terminal nested deletions generated with a Pharmacia Double-Stranded Nested Deletion kit according to the supplier's instructions and whose extremities were then sequenced with the universal and reverse M13 primers. Some sequence data were also generated using synthetic primers designed to anneal to appropriate internal sequences in the insert.

To determine if the pCA13 insert extended to the 5′-terminus of the viral RNA, the sequence of the viral RNA was determined directly by reverse transcriptase-directed primer extension using oligo 3, which is complementary to residues 30–48 of the final RNA sequence (Fig. 1B). The resulting sequence was identical to that obtained by analysis of pCA13 up to a point 53 residues from the 5′-terminus of the cloned insert. At this point the two sequences diverged, with the sequence ladder obtained by primer extension ending with a strong-stop signal six residues upstream. The sequence determined by primer extension displays extensive homology with the 5′-terminal sequences reported for other luteoviruses (see below), providing additional evidence that the sequence obtained by primer extension corresponds to the authentic 5′-extremity. The presence of an "alternate" unrelated 5′-terminal sequence on pCA13 brings to mind observations with Scottish isolates of PLRV, in which the 5′-terminal 14 residues which displayed almost perfect sequence homology with an exon of tobacco chloroplast DNA ORF 196 (*24*). PCR experiments provided direct evidence for the existence of small amounts of host RNA–viral RNA chimers in the viral RNA preparation, indicating that rare recombination events can occur between host and viral RNAs (*24*). The 53-residue "alternate" 5′-terminal sequence found in pCA13 did not display significant homology with sequences in databases but residues 1–36 were perfectly complementary to residues 4045–4080 from the CABYV coat protein cistron. Thus, in the case of pCA13, the alternate 5′-terminal sequence may be a cloning artifact.

Luteovirus RNAs are not 3′-polyadenylated. Consequently, CABYV RNA was treated with *Escherichia coli* poly(A) polymerase under conditions similar to those used to 3′-polyadenylate BWYV RNA (*3*) but when the resulting 3′-polyadenylated RNA was used as template for cDNA synthesis and cloning by conventional techniques (*22, 25*), using oligo(dT) or oligo(dT)-tailed pUC9 to prime cDNA synthesis, no clones containing long cDNA inserts were obtained. Therefore, a PCR-based method was used to generate clones containing the 3′-terminal region of the RNA. cDNA was prepared from the poly(A)-tailed CABYV RNA using the oligonucleotide 5′-GCC-AAGCT$_{18}$ (oligo 5) to prime cDNA synthesis. The cDNA was then subjected to PCR using oligo 5 and oligo 4 (equivalent to nucleotides 3396–3415, Fig. 1B). The resulting 2274-bp PCR product was digested with *Hin*dIII, which produced fragments of about 1800 and 500 bp. The two fragments were purified and the larger was cloned into *Hin*dIII-cleaved pBS(−) (Stratagene) and the smaller into *Hin*dIII/*Sma*I-cleaved pBS(−) to produce pCA36 and pCA6, respectively (Fig. 1B). The insert of each clone was completely sequenced on both strands using the universal direct and reverse primers and primers specific for internal insert sequences. Sequence data were also obtained for subclones of the pCA36 insert.

```
1    ACAAAAGATACGAGCGGGTGATGCAAATTGAGTCTGTCAAGCAACAATTAATTTTCAGAC
c                 (ORF 1) M  Q  I  E  S  V  K  Q  Q  L  I  F  R  P

61   CTACCCGACGCACGAGCATAAACGATCGCAAGTTAAACGTAGCTAACTTTCTGATCAATC
c      T  R  R  T  S  I  N  D  R  K  L  N  V  A  N  F  L  I  N  H

121  ACTCTTTCTTTTTAGCGATCAATGGCACCAACCTACTTCGCCTTTTTCTTGCTCGTCTGC
a                (ORF 2) M  A  P  T  Y  F  A  F  F  L  L  V  C
c      S  F  F  L  A  I  N  G  T  N  L  L  R  L  F  L  A  R  L  P

181  CTCTGCTCATCAGTGAGCAGTTATCAGGGGACTATGTTTATACCCCTGGAGCCAGCAAAC
a    L  C  S  S  V  S  S  Y  Q  G  T  M  F  I  P  L  E  P  A  N
c      L  L  I  S  E  Q  L  S  G  D  Y  V  Y  T  P  G  A  S  K  R

241  GCATCATACTGGCTCGATTCCACCGACATTGCGGTGCCCCCCTCCCATCCTCAAGTGCAG
a    A  S  Y  W  L  D  S  T  D  I  A  V  P  P  S  H  P  Q  V  Q
c      I  I  L  A  R  F  H  R  H  C  G  A  P  L  P  S  S  S  A  V

301  TTGATCTACGACTGCCCGCCTCAAAAGATGTTGCGCGATTTTTCCTCGCGCGACATTACT
a    L  I  Y  D  C  P  P  Q  K  M  L  R  D  F  S  S  R  D  I  T
c      D  L  R  L  P  A  S  K  D  V  A  R  F  F  L  A  R  H  Y  S

361  CTAGAGCTCTGGGAGAGAGGATACAACGAAACCAGACAAGCCTTTTCAGAGGCTATGCAG
a    L  E  L  W  E  R  G  Y  N  E  T  R  Q  A  F  S  E  A  M  Q
c      R  A  L  G  E  R  I  Q  R  N  Q  T  S  L  F  R  G  Y  A  E

421  AATTTGCAAAATTTATTAATGTCTGGTGTTCGTCAATTTCACACAGGCTTGGAGAGTTTA
a    N  L  Q  N  L  L  M  S  G  V  R  Q  F  H  T  G  L  E  S  L
c      F  A  K  F  I  N  V  W  C  S  S  I  S  H  R  L  G  E  F  T

481  CTCCACGTAATTTTACAAACGGCAGCATATTTGTGGACCTCTCTAATTTGGGCCACAGCC
a    L  H  V  I  L  Q  T  A  A  Y  L  W  T  S  L  I  W  A  T  A
c      P  R  N  F  T  N  G  S  I  F  V  D  L  S  N  L  G  H  S  L

541  TGTGCGACTTGGTACTTGCTCAGCAAGTACACGATCGAGATGCTTACACTCGCCTCGCTT
a    C  A  T  W  Y  L  L  S  K  Y  T  I  E  M  L  T  L  A  S  L
c      C  D  L  V  L  A  Q  Q  V  H  D  R  D  A  Y  T  R  L  A  L

601  TACATATCCACCGTATATATGGTGAAGATGGCGGCTTGGATTTTTGGCGACTTGCCAATT
a    Y  I  S  T  V  Y  M  V  K  M  A  A  W  I  F  G  D  L  P  I
c      H  I  H  R  I  Y  G  E  D  G  G  L  D  F  W  R  L  A  N  F

661  TCCCTTCTAAAAGCTGGCCTTTCAATGGTGAGAGGTGTCTCGAGGGCTCTGTGGTACAAA
a    S  L  L  K  A  G  L  S  M  V  R  G  V  S  R  A  L  W  Y  K
c      P  S  K  S  W  P  F  N  G  E  R  C  L  E  G  S  V  V  Q  K

721  AGGAGCTACAACGCTGAGAAATCCGTAGAGGGTTATCTTTCATTCAAAATTCCACAGAAC
a    R  S  Y  N  A  E  K  S  V  E  G  Y  L  S  F  K  I  P  Q  N
c      E  L  Q  R  *

781  CCCCCCAAAAACTCGGTCTTGCAGGTCCAATATAAGGACGGATCGCATGCCGGGTATGCT
a    P  P  K  N  S  V  L  Q  V  Q  Y  K  D  G  S  H  A  G  Y  A

841  ACTTGTGTGACGCTGTACAATGGAACGAATGGGCTTCTCACGGCCTATCATGTCGCTGTC
a    T  C  V  T  L  Y  N  G  T  N  G  L  L  T  A  Y  H  V  A  V

901  CCAGGTAGCAAAGTCGTCTCCACCAGGAACGGAAACAAGGTTCCTCTCTCTGAATTCAGA
a    P  G  S  K  V  V  S  T  R  N  G  N  K  V  P  L  S  E  F  R

961  TCCATTATGGAATCTGAAAAGAGGGATCTTGTGCTGCTGGCGGGTCCCCCAAACTGGGAA
a    S  I  M  E  S  E  K  R  D  L  V  L  L  A  G  P  P  N  W  E

1021 GGCACCCTAGCCTGCAAAGCAGTTCAGTTCCAATCTGCCCAGAACCTTTGCAAATCAAAG
a    G  T  L  A  C  K  A  V  Q  F  Q  S  A  Q  N  L  C  K  S  K

1081 GCCTCCTTCTATGCTTACGACGGAGAGGGATGGATCTCATCCAATGCTGAAATTGTTGGC
a    A  S  F  Y  A  Y  D  G  E  G  W  I  S  S  N  A  E  I  V  G

1141 ATTGCGGAGGGCAAAACCCATGCATCTGTGTTGAGCAATACGGATGCTGGTCACAGTGGC
a    I  A  E  G  K  T  H  A  S  V  L  S  N  T  D  A  G  H  S  G

1201 ACCCCATATTTCAATGGCAGAACAGTTTTGGGTGTTCACGTTGGTGGAGCCAAAGACGAA
a    T  P  Y  F  N  G  R  T  V  L  G  V  H  V  G  G  A  K  D  E

1261 AACTTCAATTACATGGCCCCCATACCGCCCGTTTATGGTCTCACAAGCCCCAGCTACGAA
a    N  F  N  Y  M  A  P  I  P  P  V  Y  G  L  T  S  P  S  Y  E

1321 TTTGAGACCACGGCCCCCCAAGGACGCCTGTTTACACAGGAGGAGATAGAAGAACTCATT
a    F  E  T  T  A  P  Q  G  R  L  F  T  Q  E  E  I  E  E  L  I

1381 GAAGAATTTTCTTTCAGTGAGATCACGTCCATAATGGGGCATCGCCGATTTCATCAAATG
a    E  E  F  S  F  S  E  I  T  S  I  M  G  H  R  R  F  H  Q  M

1441 CACGACTCGCAGCGACATCAAGCTGATTATGAATACGAGTCGGGAAACGGGCAGGCGGCG
a    H  D  S  Q  R  H  Q  A  D  Y  E  Y  E  S  G  N  G  Q  A  A
c                            (ORF 3) I  R  V  G  K  R  A  G  G

1501 GCGACCGCCGAAACAACCGCACCCGATGCCACCCCCGCGGTTGGGCGCACAAGTGGAGAC
a    A  T  A  E  T  T  A  P  D  A  T  P  A  V  G  R  T  S  G  D
c      D  R  R  N  N  R  T  R  C  H  P  R  G  W  A  H  K  W  R  R

1561 GCACAGTCTCCTGCTGCTCCCTCTTCCCCGAGGGTCAACACGTACACACCTCCGAACAAA
a    A  Q  S  P  A  A  P  S  S  P  R  V  N  T  Y  T  P  P  N  K
c      T  V  S  C  C  S  L  F  P  E  G  Q  H  V  H  T  S  E  Q  T

1621 CGGGAAAACCGCACTGCTTCTGCAGTATCCCCCACCACCAGTGCTTCTACCGACACTTTG
a    R  E  N  R  T  A  S  A  V  S  P  T  T  S  A  S  T  D  T  L
c      G  K  P  H  C  F  C  S  I  P  H  H  Q  C  F  Y  R  H  F  V

1681 TCGGAAATCAAACAGGTAATAATGGACAAAATCGATGTCCATTCAATAGAGAAGCAAGTG
a    S  E  I  K  Q  V  I  M  D  K  I  D  V  H  S  I  E  K  Q  V
c      G  N  Q  T  G  N  N  G  Q  N  R  C  P  F  N  R  E  A  S  G

1741 GTCCAAGCCTTGGCGGACAAAGCCATGAAGAAACCGAGGTCGCGCAGACGCGGCAAGAGA
a    V  Q  A  L  A  D  K  A  M  K  K  P  R  S  R  R  R  G  K  R
c      P  S  L  G  G  Q  S  H  E  E  T  E  V  A  Q  T  R  Q  E  K

1801 AGTTCGAAGAACTCGCAACCGACTTCTCAAGTTTCTTCGATGCCCAGTACACCTGGGAAC
a    S  S  K  N  S  Q  P  T  S  Q  V  S  S  M  P  S  T  P  G  N
c      F  E  E  L  A  T  D  F  S  S  F  F  D  A  Q  Y  T  W  E  Q

1861 AAGGCCGGGAAGAAGCCCCAGGATTCGACAAAGTCGGATCCCTCCCCCAGTTTTACCACG
a    K  A  G  K  K  P  Q  D  S  T  K  S  D  P  S  P  S  F  T  T
c      G  G  E  E  A  P  G  F  D  K  V  G  S  L  P  Q  F  Y  H  A

1921 CTAAACAAAAGAAAAGCTCGAATTGGGGAGACAAAATCTGCAAACAGCACCCAGAAATGG
a    L  N  K  R  K  A  R  I  G  E  T  K  S  A  N  S  T  Q  K  W
c      K  Q  K  K  S  S  N  W  G  D  K  I  C  K  Q  H  P  E  M  G

1981 GTGATCTCACAAAGGGCTTCGGGTGGCCCCAGTTCGGCGCAAAAGCTGAACTAAAATCGT
a    V  I  S  Q  R  A  S  G  G  P  S  S  A  Q  K  L  N  *
c      D  L  T  K  G  F  G  W  P  Q  F  G  A  K  A  E  L  K  S  L

2041 TGCGGCTACAAGCCGCGCGTTGGCTGGAGCGTGCCCAGTCAGTTAAAATCCCCTCATCTG
c      R  L  Q  A  A  R  W  L  E  R  A  Q  S  V  K  I  P  S  S  E

2101 AGGAGCGGGAGCACGTTATAGAGAGATGCTGTCGGGCGTATCAAGCTGCTAAAACTAATG
c      E  R  E  H  V  I  E  R  C  C  R  A  Y  Q  A  A  K  T  N  G

2161 GCCCGATGGCGACGAGAGGAGATCGACTTTCCTGGGATAACTTCCTACAGGATTTCAAAC
c      P  M  A  T  R  G  D  R  L  S  W  D  N  F  L  Q  D  F  K  Q

2221 AGGCGGTCCTCTCGTTGGAAATGGACGCAGGCATAGGAGTTCCGTACATAGCTTACGGCA
c      A  V  L  S  L  E  M  D  A  G  I  G  V  P  Y  I  A  Y  G  K

2281 AGCCAACCCACCGTGGGTGGGTTGAAGACAAGAAACTCCTTCCGATATTGGCTCGTCTCA
c      P  T  H  R  G  W  V  E  D  K  K  L  L  P  I  L  A  R  L  T

2341 CCTTCAACCGGTTACAGAAGATGTTGGAGGTAAGATATGTTGACTTATCGCCTGCGGAGC
c      F  N  R  L  Q  K  M  L  E  V  R  Y  V  D  L  S  P  A  E  L

2401 TTGTGCGAAGGGGTCTCTGTGACCCTATTCGAGTATTTGTAAAAGGTGAACCGCACAAGC
c      V  R  R  G  L  C  D  P  I  R  V  F  V  K  G  E  P  H  K  Q

2461 AAGCAAAATTAGATGAAGGCCGCTACCGCCTCATCATGAGCGTCTCTCTAGTTGACCAAT
c      A  K  L  D  E  G  R  Y  R  L  I  M  S  V  S  L  V  D  Q  L

2521 TGGTAGCCCGGGTTTTATTCCAAAATCAAAACAAGCGCGAGATAGCGTTGTGGAGAGTGG
c      V  A  R  V  L  F  Q  N  Q  N  K  R  E  I  A  L  W  R  V  V

2581 TTCCCTCCAAACCCGGTTTTGGCTTGTCTACGGATGAGCAAGTGGCGGAGTTCATGCAGA
c      P  S  K  P  G  F  G  L  S  T  D  E  Q  V  A  E  F  M  Q  I

2641 TTCTCTCCGCCCAAGTCGGGCTAACGCCTTCGGATCTGATCACCGAGTGGCGAGCCCATA
c      L  S  A  Q  V  G  L  T  P  S  D  L  I  T  E  W  R  A  H  M

2701 TGATAGCAACTGACTGCTCGGGTTTTGACTGGAGCGTTTCGGACTGGCTCCTTGAAGATG
c      I  A  T  D  C  S  G  F  D  W  S  V  S  D  W  L  L  E  D  D

2761 ATATGGAGGTCCGAAACCGGCTGACGCTCGATTTAAACGAGACCACGCGCCGTTTGCGCA
c      M  E  V  R  N  R  L  T  L  D  L  N  E  T  T  R  R  L  R  S

2821 GCTGGTTATATTGCATATCGAATAGCGTCCTCTGCCTTTCAGATGGAACATTATTAGCGC
c      W  L  Y  C  I  S  N  S  V  L  C  L  S  D  G  T  L  L  A  Q

2881 AGAGAGTCCCGGGCGTGCAGAAGAGTGGCAGCTACAATACGTCATCGAGTAATTCCCGAA
c      R  V  P  G  V  Q  K  S  G  S  Y  N  T  S  S  S  N  S  R  I

2941 TCCGGGTGATGGCTGCCTACCACTGCGGCGCAGAATGGGCTATGGCGATGGGCGACGATG
c      R  V  M  A  A  Y  H  C  G  A  E  W  A  M  A  M  G  D  D  A

3001 CCCTCGAGTCAGCTTGCTCGAACCTCGAGCGTTACAAATCGCTCGGTTTTAAAGTCGAGG
c      L  E  S  A  C  S  N  L  E  R  Y  K  S  L  G  F  K  V  E  E

3061 AGTCCTCAAAACTGGAATTCTGTTCTCACATCTTTGAGAAAGAGGACCTCGCCATTCCGG
c      S  S  K  L  E  F  C  S  H  I  F  E  K  E  D  L  A  I  P  V

3121 TCAACAAGGCCAAAATGCTCTATAAGCTTATACATGGCTATGAACCGGAATGCGGCAATG
c      N  K  A  K  M  L  Y  K  L  I  H  G  Y  E  P  E  C  G  N  V

3181 TGGAAGTCCTGATCAATTATTTGGCCGCCTGCTTTTCTATTCTTAATGAATTGCGGTCTG
c      E  V  L  I  N  Y  L  A  A  C  F  S  I  L  N  E  L  R  S  D

3241 ATCCTTCCCTTGTCGAAACTCTCCATCAGTGGCTGGTTGCTCCAGTACAGCCACAAAAGA
c      P  S  L  V  E  T  L  H  Q  W  L  V  A  P  V  Q  P  Q  K  I

3301 TATAAGGGGAGTATAAAGAACACTAGCCAAGCACACACGAGTTGCAAGCATTGGAAGTCT
c      *

3361 GAGTCTCGTTACCAAGAGTCCACACAATAGATTATAAGTTTCTCGCAGGATTTTCAAGCG

3421 GTCTATTGTCTGCAATACCCATCACGGTAATCGGACTCTATTTTGTCTACCTAAAGATTT

3481 CAGCACACGTGCGTTCAATTGTTAATGAATACGGTCGCGGCTAGAAATCAAAATGCAGGG
a                    (ORF 4/6) M  N  T  V  A  A  R  N  Q  N  A  G
b                                              (ORF 5) M  Q  G
```

FIG. 2. Nucleotide sequence of CABYV RNA. The amino acid sequences of the long ORFs identified in Fig. 1 are written beneath the sequence, which is presented as DNA. The nucleotide sequence has been assigned the accession number X76931 in the EMBL Data Library.

Finally, the junction between pCA6 and pCA36 was characterized by primer extension from oligo 6 (Fig. 1B) using viral RNA as template. The complete sequence, obtained by combining the results from analysis of pCA13, pCA6, and pCA36, was 5669 residues (Fig. 2).

The position of long ORFs on the CABYV RNA plus-strand (no long ORFs were present on the minus-strand) is shown in Fig. 1B and Fig. 2. Note that the CABYV sequence possesses an ORF 1 counterpart and a long sequence overlap between ORF 2 and ORF 3, characteristics of luteovirus Subgroup 2.

The 5′-noncoding region of CABYV RNA is 20 residues with the first eight residues, 5′-ACAAAAGA, identical to the 5′-terminal sequences of BWYV and PLRV. The 5′-noncoding region of BYDV(RPV) does not contain the aforesaid octomer. The intergenic region separating the end of ORF 3 from the beginning of the coat protein cistron contains 199 residues, compared to 202, 196, and 185 nucleotides, respectively, for the intergenic regions of BWYV, PLRV, and BYDV(RPV). It is interesting to note that a second copy of the 5′-terminal consensus sequence is present near the 3′-terminus of ORF 3 in CA-

```
3541 AGGCGAAGGCGAAGAAATCAGCGCCCTGCGCGGCGCGACCGCGTGGTTGTGGTCAACCCC
a    R  R  R  R  R  N  Q  R  P  A  R  R  D  R  V  V  V  V  N  P
b     G  E  G  E  E  I  S  A  L  R  G  A  T  A  W  L  W  S  T  P

3601 ATTGGGGGACCACCGCGCGGAAGACGACAACGAAGAAACCGCCGACGCCCTAATCGAGGA
a    I  G  G  P  P  R  G  R  R  Q  R  R  N  R  R  R  P  N  R  G
b     L  G  D  H  R  A  E  D  D  N  E  E  T  A  D  A  L  I  E  E

3661 GGCAGAGCTAGAAGAGGAAGCCCAGGCGAAACATTTGTATTTTCAAAGGACAATCTCACG
a    G  R  A  R  R  G  S  P  G  E  T  F  V  F  S  K  D  N  L  T
b     A  E  L  E  E  E  A  Q  A  K  H  L  Y  F  Q  R  T  I  S  R

3721 GGCAGTTCCACAGGAAGTATCACCTTCGGGCCGTCTCTTTCAGAGAGCCCAGCATTCAGC
a    G  S  S  T  G  S  I  T  F  G  P  S  L  S  E  S  P  A  F  S
b     A  V  P  Q  E  V  S  P  S  G  R  L  F  Q  R  A  Q  H  S  A

3781 TCTGGAATACTCAAGGCCTACCATGAATATAAGATCATCATGGTCCAGCTGGAGTTCATC
a    S  G  I  L  K  A  Y  H  E  Y  K  I  I  M  V  Q  L  E  F  I
b     L  E  Y  S  R  P  T  M  N  I  R  S  S  W  S  S  W  S  S  S

3841 TCCGAGGCCTCTTCCACCTCCTCGGGTTCCATCTCTTATGAGTTGGACCCCCACTGCAAG
a    S  E  A  S  S  T  S  S  G  S  I  S  Y  E  L  D  P  H  C  K
b     P  R  P  L  P  P  P  R  V  P  S  L  M  S  W  T  P  T  A  S

3901 CTTAGCTCCCTCCAATCCACGATTAATAAATTTGGAATCACCAAGAGTGGATTGCGACGT
a    L  S  S  L  Q  S  T  I  N  K  F  G  I  T  K  S  G  L  R  R
b     L  A  P  S  N  P  R  L  I  N  L  E  S  P  R  V  D  C  D  V

3961 TGGACCGCTAAGCAGATCAACGGGATGGAATGGCATGATGCAACCGAAGACCAGTTCAAG
a    W  T  A  K  Q  I  N  G  M  E  W  H  D  A  T  E  D  Q  F  K
b     G  P  L  S  R  S  T  G  W  N  G  M  M  Q  P  K  T  S  S  R

4021 ATCCTCTATAAAGGGAATGGATCTTCCTCGGTTGCGGGCAGCTTCAGAATCACCATCAAG
a    I  L  Y  K  G  N  G  S  S  S  V  A  G  S  F  R  I  T  I  K
b     S  S  I  K  G  M  D  L  P  R  L  R  A  A  S  E  S  P  S  S

4081 TGCCAGGTCCAGAACCCGAAATAGGTAGACGGCAGTTCCCCCCCCCCCCAAGTCCTAGT
a    C  Q  V  Q  N  P  K  *  V  D  G  S  S  P  P  P  P  S  P  S
b     A  R  S  R  T  R  N  R  *

4141 CCAACCCCACCACCTCCACCACCCCCTCAGCCTCAACCCCAGCCTTGCGCCCAGCGTTTT
a    P  T  P  P  P  P  P  P  P  Q  P  Q  P  Q  P  C  A  Q  R  F

4201 TGGGGTTATGAAGGCAACCCACAAAATAAAATACTCACGGCTGAAAATTCAAGGAATATT
a    W  G  Y  E  G  N  P  Q  N  K  I  L  T  A  E  N  S  R  N  I

4261 GATTCCCGGCCGTTGAATTTCGTGCAAATGTACAAGTGGGAAGATGAGAAGTGGGACAAG
a    D  S  R  P  L  N  F  V  Q  M  Y  K  W  E  D  E  K  W  D  K

4321 GTCAACTTGCAAGCCGGATATTCTCGCAATGACCGACGTTGTATGGAGACGTACCTCACA
a    V  N  L  Q  A  G  Y  S  R  N  D  R  R  C  M  E  T  Y  L  T

4381 ATTCCAGCAGATAAAGGAAAATTTCATGTTTATCTTGAAGCTGATGGTGAGTTCGTCGTC
a    I  P  A  D  K  G  K  F  H  V  Y  L  E  A  D  G  E  F  V  V

4441 AAGCACATCGGTGACGAGCTTGACGGTAGTTGGCTCGGAAATATCGCTTATGATGTTTCC
a    K  H  I  G  D  E  L  D  G  S  W  L  G  N  I  A  Y  D  V  S

4501 CAAAGAGGTTGGAACGTAGGGAATTACAAAGGCTGTAAAATCACAAATTACCAATCTAAT
a    Q  R  G  W  N  V  G  N  Y  K  G  C  K  I  T  N  Y  Q  S  N

4561 ACAGTTTTTGTGGCAGGACATCCCGACGCAACAATGAATGGTAAAAGTTTTGACACTGCG
a    T  V  F  V  A  G  H  P  D  A  T  M  N  G  K  S  F  D  T  A

4621 CGGGCCGTTGAAGTCGATTGGTTCGCTTCCTTCGAATTAGAATGTGATGATGAGGAAGGA
a    R  A  V  E  V  D  W  F  A  S  F  E  L  E  C  D  D  E  E  G

4681 AGCTGGGCTATATACCCCCCTCCCATACAGAAGGATTCTTCGTATAATTACACCGTCTCA
a    S  W  A  I  Y  P  P  P  I  Q  K  D  S  S  Y  N  Y  T  V  S

4741 TACGGGAATTACACCGAGAAATATTGCGAGTGGGGGGCAATTTCGGTATCAATCGACGAA
a    Y  G  N  Y  T  E  K  Y  C  E  W  G  A  I  S  V  S  I  D  E

4801 GATAATAATGGGAACGAGCCAAGAAGAATACCACGCAGGGGGGTAATGGCGTGGTCCACC
a    D  N  N  G  N  E  P  R  R  I  P  R  R  G  V  M  A  W  S  T

4861 CCTGAACCATCTTTTTCAGGGGATGATAGTCAAAGACAAGACTTTAACACTCCTTCGCTT
a    P  E  P  S  F  S  G  D  D  S  Q  R  Q  D  F  N  T  P  S  L

4921 GAAGAGCGAGGTTCCGATGCCCTGGAATCGGAAGAGAAGAAGGAAGAGGACAACCTTCTT
a    E  E  R  G  S  D  A  L  E  S  E  E  K  K  E  E  D  N  L  L

4981 GATCTTGAGGAAGAGAATATACCCGACGTTGATGATGACGATTTATGGAAGGGTATTTCC
a    D  L  E  E  E  N  I  P  D  V  D  D  D  D  L  W  K  G  I  S

5041 CGTGCCTCCGAAGCCGGGACTGCCGAGGATGATCGAGCGTCTACAAGCTCTCGTCTTCGC
a    R  A  S  E  A  G  T  A  E  D  D  R  A  S  T  S  S  R  L  R

5101 GGCAACCTCAAGCCAAAGGGGCTACCCAAACCACAGCCTACTCGAACCATTACCGAGTTC
a    G  N  L  K  P  K  G  L  P  K  P  Q  P  T  R  T  I  T  E  F

5161 AATCCTGGACCTGATTTGATAGAGGTGTGGAGGCCCGATCTAGCACCTGGATACTCCAAA
a    N  P  G  P  D  L  I  E  V  W  R  P  D  L  A  P  G  Y  S  K

5221 GCGGACGTCGCCGCAGCGACAGTGCTGGCAGGAGGAAGTGTCCACGAGGGAAGAGATATG
a    A  D  V  A  A  A  T  V  L  A  G  G  S  V  H  E  G  R  D  M

5281 CTCGAGCGTAGAGAAGCAAAAGTAATGGATAGTAGGAAGAAATGGGGAATCCTCTCCTCA
a    L  E  R  R  E  A  K  V  M  D  S  R  K  K  W  G  I  L  S  S

5341 ACTTCCTCTCTTACTAGCGGAGCATTGAAGAAGCTCAGCGCACAATCGGAAAAGCTCGCC
a    T  S  S  L  T  S  G  A  L  K  K  L  S  A  Q  S  E  K  L  A

5401 ACATTGACAACTGGCGAGCGAGTGCAATATCAACGACTTAAAAACTCAATGGGTTCGACT
a    T  L  T  T  G  E  R  V  Q  Y  Q  R  L  K  N  S  M  G  S  T

5461 GTTGCAGCTGAGTATCTCGAAAAGGTGCTTGCTGATAAAACCTCCTAAAGAGTTCCTACT
a    V  A  A  E  Y  L  E  K  V  L  A  D  K  T  S  *

5521 ACTCTGAAGTAGCGCGGCACCGGAATACGCCGCGAAGCTCAATCCGGAAACTCAAGAACG

5581 AGAGTATAAACGTTAACGACTCCGAAAGGATAGGCAACGAACGTTCCCACGTAGTGGAAA

5641 CAGGGGGATTCCCCCTGGCGTTTCGGTGT 5669
```

FIG. 2—*Continued*

BYV (residues 3293–3300), as is also the case in BWYV and PLRV (*26*). In PLRV and BWYV, the 5′-terminus of the subgenomic RNA responsible for expression of the ORF 4, 5, and 6 gene cluster has been mapped to the first A of this sequence (*26;* V. Ziegler-Graff, personal communication). A virus-specific RNA species of about 2.3-kb which presumably corresponds to this subgenomic RNA is readily detected in CABYV-infected *Chenopodium quinoa* protoplasts and run-off transcription experiments (data not shown) have shown that its 5′-extremity also corresponds to the first residue of the ACAAAAGA repeat.

The 3′-noncoding region of CABYV is 161 residues. The RNAs of CABYV and BWYV display considerable sequence homology in this region with, in particular, perfect identity for the 3′-terminal eight residues. No particular homology was detected between the CABYV 3′-noncoding region and the corresponding regions of PLRV and BYDV(RPV).

Table 1 shows the degree of homology between the deduced amino acid sequences of ORFs 1–6 of CABYV and their counterparts in the other Subgroup 2 viruses. The sequences were aligned with the program GAP (*27*). The alignments reveal extensive homology between ORFs 2-6 of CABYV and the corresponding ORFs of the other Subgroup 2 members, with, generally, the closest homology being with BWYV. The CABYV and BWYV P1 sequences displayed highly statistically significant homology, as tested by Monte Carlo comparisons of the Quality (*27*) of the alignment with the Quality values produced for alignments of CABYV P1 to random sequences of the same length and amino acid composition as the BWYV P1 sequence. Similar tests with the P1s of BYDV(RPV) and PLRV revealed marginal statistical significance for the CABYV–BYDV(RPV) alignment and no significance for the CABYV–PLRV alignment. We regard the similarity between the P1s of CABYV and BWYV as of particular taxonomic importance in view of the low degree of relatedness among the P1s of the other Subgroup 2 luteoviruses.

TABLE 1

AMINO ACID SEQUENCE COMPARISONS BETWEEN THE ORFs OF CABYV AND THE CORRESPONDING ORFs OF OTHER SUBGROUP 2 LUTEOVIRUSES

	BWYV[a]	PLRV[b]	BYDV(RPV)[c]
ORF 1	25	18	16
ORF 2	38	30	34
ORF 3	64	56	58
ORF 4	67	58	65
ORF 5	45	51	45
ORF 6	40	28	30

Note. Numbers are the percentage of identical amino acids in the sequences aligned using the UWGCG program GAP (*27*).

[a] Ref. *3*.

[b] Ref. *10*.

[c] Ref. *11*.

```
            <---- PROLINE HINGE ---->              LUTEOVIRUS HOMOLOGY
CABYV      VDGSSPPPPSPSPTPPPPPPPQPQPQPCAQ.......RFWGYEGNPQNKILTAENSRNIDSRPLNFVQMYKW.E 66
BWYV       VD..EEPGPSPGPS........PSPQPTPQK.....KYRFIVYTGVPVTRIMAQSTDDAISLYDMP.SQRFRYIE 59
PLRV       VDSGSEPGPSPQ..........PTPTPTPQK.....HERFIAYVGIPMLTIQARENDDQIILGSLG.SQRMKYIE 59
BYDV(RPV)  VD..AEPGPSPGPSPDPPPPPSPSPEPAPAK.....EERFIVYSGVAHTIISAQSTDDSIIVRDIP.DQRFRYVE 67
BYDV(MAV)  VD.SSTPEPSPQPQ....PEPKPDPQPTPEPR....QKRFFEYVGTPYVVIQTRESSDSIAVKAMN.DQSFQYIE 65
BYDV(PAV)  VD.SSTPEPKPAPEPTPTPQPTPAPQPTPEPTPAPVPKRFFEYIGTPTGTISTRENTDSISVSKLG.GQSMQYIE 73
Consensus  VD    P P P           P P P          RF  Y G     I        I        Q     E

CABYV      DEKWDKVNLQAG.YSRNDRRCMETYLTIPADKGKFHVYLEADGEFVVKHIGDELDGSWLGNIAY.DVSQRGWNVG 139
BWYV       DENMNWTNLDSRWYSQNSLKAIPMII.VPVPQGEWTVEISMEGYQPTSSTTDPNKDKQDGLIAYNDDLSEGWNVG 133
PLRV       DENQNYTNISSEYYSQSSMQAVPMYY.FNVPKGQWSVDISCEGYQPTSSTSDPNRGRSDGMIAYSNADSDYWNVG 133
BYDV(RPV)  NENFYWFQIAAQWYSNTNTKAVPMFV.FPVPIGEWSVEISTEGYQATSSTTDPNKGRIDGLIAY.DNSSEGWNIG 140
BYDV(MAV)  NETSEQRTVKAWWNSNNSVQAQAAFI.FPIPAGEYSVNISCEGLQSVDHIGGNRDGYWIGLIAYQSQSGDYWGVG 139
BYDV(PAV)  NEKCETKVIDSFWSTNNNVSAQAAFV.YPVPEGSYSVNISCEGFQSVDHIGGNEDGYWIGLIAYSNSSGDNWGVG 147
Consensus  *E     *      S                 G   V    EG                G IAY       W  G

CABYV      NYKGCKITNYQSNTVFVAGHPDATMNGKSFDTARAVEVDWFASFELECDDEEGSWAIYPPPIQKDSSYNYTVSYG 214
BWYV       IYNNVEITNNKADNTLKYGHPDMELNGCHFNQGQCLERDGDLTCHIKTTGDNASFFVVGPAVQKQSKYNYAVSYG 208
PLRV       EADGVKISKLRNDNTYRQGHPELEINSCHFREGQLLERDATISFHVEAPTD.GRFFLIGPAIQKTAKYNYTISYG 207
BYDV(RPV)  AGSNVTITNNKADNSWKYGHPDLEINSCHFNQNQVLEKDGIISFHVKATEKEANFFLVAPPVQKTSKYNYAVSYG 215
BYDV(MAV)  NYVGCDITNLLGTNTWRPGHEDLELNSCKFTDGQIVERDAVISFHVKARGADPKFYLMAPKTMKADKYNYVVSYG 214
BYDV(PAV)  NYKGCSFKNFLATNTWRPGHKDLKLTDCQFTDGQIVERDAVNSFHVEATGKDASFYLMAPKTNKTDKYNYVVSYG 222
Consensus                    GH D       F      E D   S       *        P   K   YNY  SYG

CABYV      NYTEKYCEWGAISVSIDEDNNGNEP....RRIPRRG.VMAWST......PEPSFSGDDS......QRQ....... 265
BWYV       AWTDRMMEIGMIAIALDEQ.GSSGSV.KTERPKRVGHSMAVSTWETIKLPE...KGNSEGYETS.QRQ.....DS 272
PLRV       DWTDRDMELGLITVVLDEHLEGTDSAYRLRRPLREGHTYMVSPNE....PEGKPVGNKPRDETPIQTQ.....ER 273
BYDV(RPV)  AWTDRDMEFGLITVTLDEK.RGSGS..PTRKSLRAGHAGVTTTTDLVALPE...MENS.GIETS.ETPSAPVTSS 282
BYDV(MAV)  GYTDKRMEFGTISVTVDESDVEAERYSRHTSTVRRTE.....NRDYGWMNVLPPY.NPD..QVPEQEDEQPVVDK 281
BYDV(PAV)  GYTNKRMEFGTISVTCDESDVEAERITRHAETPERSKHILVSER.YA..EPLPTIVNQGLCDVKTPEQEQTLVD. 293
Consensus    T    E G I    DE                                     *

CABYV      ......DFNTPSLEERGSDALESEEKKEEDNLLDLEEE....NIPDVDDDDLWKGISRASEAGTAEDDRASTSSR 330
BWYV       K....TP.PTAS..GG.SDTLDVE...EGGLPLPVEEE.....IPDFVGDNPW...SDLSTKNSQEEEAMSSESG 328
PLRV       K.....PDQTPS..DDVSDAGSVN...SGGSTESLHLE.....F....G...........LNSDSTYDATVDGT 317
BYDV(RPV)  K....APLPTVS.........DSE...SEDDPLSAA........PD.VG.......................FGG 309
BYDV(MAV)  EMDAGSPIDTASLTSDTEAEKAFDLKEEELTRAILEYEAATVSIPDAAPDIL............PSKSEMSSKPI 344
BYDV(PAV)  EDD....RQTVSTESDI...............ALLEYEAATAEIPDAEEDVL............PSKEQLSSKPM 337
Consensus           T S     **                                           *

                                   CABYV/BWYV/PLRV HOMOLOGY
CABYV      LRGNLKPKGLPKPQPT....RTITEFNPGPDLIEVWRPD.LAPGYSKADVAAATVLAGGSVHEGRDMLERRE... 397
BWYV       LRPQLKPPGLPKPQPI....RTIRNFDPTPDLVEAWRPD.VNPGYSKADVAAATIIAGGSIKDGRSMIDKRN... 395
PLRV       DWPRIPPPRHPPEPRVSGNSRTVTDFSPKADLLENWDAEHFDPGYSKEEVAAATIIAHGSIQDGRSMLEKREENV 392
BYDV(RPV)  TR.......LLIDTDI....KTI....PDPDVADAF....VNSAHVGVDPWA..........EVRAF..KR.... 349
BYDV(MAV)  DRD...GRSLPKSQ..............TKEVLGTYQGQNITS....DDVPPV..........IAEKL..REVNR 386
BYDV(PAV)  DTS...GNIIPKPK..............EPEVLGTYQGQNIYP....EDVPPM..........ARQKL..REAAN 379
Consensus        *                  *    D*   *  *    **** D * **  * **   *  * * R

CABYV      ...................AKVMDSRKKW.............GILSS..TSSLTSGALK.KLSAQSEK..LATLT 435
BWYV       ............KAVLDG.......RKSW.............G..SSLA.SSLTGGTLKAS..AKSEK..LAKLT 431
PLRV       KNKTSSWKPPLPKAVSPAIAKLRSIRKSQPLEGGTLKKDATDGV.SSIGSGSLTGGTLK..RKATFEENLLQTLT 464
BYDV(RPV)  .............A...Q.......RPPR..............GP.SSVASSSLSGGSLRGSLRPKTED...PKDS 384
BYDV(MAV)  APSTLLYDRQ.PKQPKNPLTRFVLSNK...............TSTASPGSQSSTAG.................MT 428
BYDV(PAV)  APSTLLYERRTPKKSGNFLSRLVEANRS............PTTPTA.P.SVSTTSN.................MT 423
Consensus                                                     S T          *      * *T

CABYV      TGERVQYQRLKNSMGSTVAAEYLEKVLADKTS............. 467
BWYV       TSERARYERIKRQQGSTRASEFLESLLAGEDPDSRF......... 467
PLRV       TEQRLWYENLKKTNPPA.ATQWLFKYQPPPQVDRNLAEKPFQGRK 508
BYDV(RPV)  SKSKSR....KWSLGSLR........................... 398
BYDV(MAV)  REQASEYTRIRKSLGLTAAKQYKASLDDT................ 457
BYDV(PAV)  REQLREYTRERNSSGITAAKAYKAQFQ.................. 450
Consensus            K           *
```

Fig. 3. Amino acid sequence homologies in the luteovirus readthrough (RT) domain (ORF 6). The positions of the proline hinge, the luteovirus homology subdomain, and the CABYV/BWYV/PLRV homology region are indicated. Residues which are identical or similar (R and K, S and T, D and E, Y and F) in all six sequences are identified in the consensus sequence. Positions in which the three viruses transmitted by *M. persicae* (CABYV, BWYV and PLRV) have identical or similar residues which are different from the other luteoviruses are indicated by asterisks in the consensus sequence. The alignment was generated with PILEUP (*27*).

Detailed analysis of the amino acid sequences of CABYV ORFs 2 and 3 reveals the presence of the consensus sequence motifs found in other luteoviruses (see Ref. *5* for a review). In particular, ORF 2 contains the putative active site residues for a chymotrypsin-like protease while ORF 3 contains the well characterized "core" RNA polymerase sequence motifs present in the RNA-dependent RNA polymerases of all plus-sense RNA viruses (see Ref. *5* for a recent compilation). As noted above, ORF 3 of luteoviruses is expressed as an ORF 2–ORF 3 translational fusion protein by a −1 frameshift in the region of sequence overlap between the two ORFs. In BWYV and PLRV, the site of the frameshift has been mapped to a "slippery" heptanucleotide located a few residues upstream of a pseudoknot (*15, 16*). A slippery heptanucleotide GGGAAAC (residues 1482–1488) followed by a pseudoknot is also present near the beginning of the ORF 2–ORF 3 overlap in CABYV.

As noted above, ORF 6 of the luteoviruses is expressed as an ORF 4–ORF 6 fusion protein (RT protein) by translational readthrough of the coat protein UAG termination codon. The sequence context in the vicinity of the "leaky" UAG is identical for all the luteoviruses . . . AAAUAGGU-AGAC . . . (termination codon underlined). The RT pro-

tein or degradation products thereof can be detected in association with partially purified virus particles (*18, 28*) and it has been suggested that the RT domain itself, which is subject to partial proteolytic degradation in the course of virus purification (*18*), is probably exposed on the virion surface where it may contain determinants important in virus–vector interactions.

Multiple sequence alignment of the RT domain of all luteoviruses reveals considerable homology (Fig. 3). The region immediately downstream of the coat protein termination codon is very rich in proline. This "proline hinge" (Fig. 3) may serve as a loose tether joining the coat protein moiety (presumably anchored in the virion capsid) to the rest of the RT domain. The rest of the RT domain can be divided into two parts: the C-terminal distal region in which there are a large number of positions absolutely conserved homology among all the sequenced luteoviruses (the "Luteovirus Homology Subdomain"; Fig. 3) and the C-terminal proximal region, which displays much less overall homology.

CABYV, like BWYV and PLRV, is efficiently transmitted by *M. persicae* but this aphid is not an efficient vector for BYDV(RPV), BYDV(MAV), and BYDV(PAV) (*29*). Evidently, comparisons between the CP and RT domain sequences of these two classes of virus could identify "*M. persicae*-specific" sequence motifs. This analysis reveals that such motifs are more abundant in the RT domain than in the CP. Thus, the aligned coat protein sequences of the aforesaid luteoviruses contain only two positions with the same amino acid in CABYV, BWYV, and PLRV but absent from the other luteoviruses (data not shown). The C-terminal portion of the RT domain, on the other hand, contains a high content of such vector-specific residues with, in particular, a region about 70 residues from the C-terminus (bracketed in Fig. 3) in which 17 of 44 residues are identical exclusively in CABYV, BWYV, and PLRV. It will be interesting to determine if this region is of particular importance in conferring specificity on the aphid transmission process.

ACKNOWLEDGMENTS

The authors thank Daniele Scheidecker for technical assistance and Olivier Poch for advice concerning sequence alignments.

REFERENCES

1. Lecoq, H., Bourdin, D., Wipf-Scheibel, C., Bon, M., Lot, H., Lemaire, O., and Herrbach, E., *Plant. Pathol.* **41,** 749–761 (1992).
2. Lemaire, O. J., Gubler, W. D., Valencia, J., Lecoq, H., and Falk, B. W., *Plant Dis.* **77,** 1169 (1993).
3. Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K., and Jonard, G., *Nucleic Acids Res.* **16,** 9917–9932 (1988).
4. Habili, N., and Symons, R. H., *Nucleic Acids Res.* **17,** 9543–9555 (1989).
5. Koonin, E. V., and Dolja, V. V., *Crit. Rev. Biochem. Mol. Biol.* **28,** 375–430 (1993).
6. Miller, W. A., Waterhouse, P. W., and Gerlach, W. L., *Nucleic Acids Res.* **16,** 6097–6111 (1988).
7. Ueng, P. P., Vincent, J. R., Kawata, E. E., Lei, C.-H., Lister, R. M., and Larkins, B. A., *J. Gen. Virol.* **73,** 487–492 (1992).
8. Van Der Wilk, F., Huisman, M. J., Cornelissen, B. J. C., Huttinga, H., and Goldbach, R., *FEBS Letters* **245,** 51–56 (1989).
9. Mayo, M. A., Robinson, D. J., Jolly, C. A., and Hyman, L., *J. Gen. Virol.* **70,** 1037–1051 (1989).
10. Keese, P., Martin, R. R., Kawchuk, L. M., Waterhouse, P. M., and Gerlach, W. L., *J. Gen. Virol.* **71,** 719–724 (1990).
11. Vincent, J. R., Lister, R. M., and Larkins, B. A., *J. Gen. Virol.* **72,** 2347–2355 (1991).
12. Rathgen, J. P., Karageorgos, L. E., Habli, N., Waterhouse, P. M., and Symons, R. H., *Virology* **198,** 671–679 (1994).
13. Brault, V., and Miller, W. A., *Proc. Natl. Acad. Sci. USA* **89,** 2262–2266 (1992).
14. Prüfer, D., Tacke, E., Schmitz, J., Kull, B., Kaufmann, A., and Rhode, W., *EMBO J.* **11,** 1111–1117 (1992).
15. Garcia, A., van Duin, J., and Pleij, C. W. A., *Nucleic Acids Res.* **21,** 401–406 (1993).
16. Kujawa, A. B., Drugeon, G., Hulanicka, D., and Haenni, A.-L., *Nucleic Acids Res.* **21,** 2165–2171 (1993).
17. Tacke, E., Prüfer, D., Salamini, F., and Rohde, W., *J. Gen. Virol.* **71,** 2265–2272 (1990).
18. Bahner, I., Lamb, J., Mayo, M. A., and Hay, R. T., *J. Gen. Virol.* **71,** 2251–2256 (1990).
19. Robertson, N. L., French, R., and Gray, S. M., *J. Gen. Virol.* **72,** 1473–1477 (1991).
20. Jones, T. D., Buck, K. W., and Plumb, R. T., *J. Virol. Meth.* **35,** 287–296 (1991).
21. Fichot, O., and Girard, M., *Nucleic Acids Res.* **18,** 6162 (1990).
22. Gubler, U., and Hoffman, B. J., *Gene* **25,** 263–269 (1983).
23. Zhang, H., Scholl, R., Browse, J., and Somerville, C., *Nucleic Acids Res.* **126,** 1220 (1988).
24. Mayo, M. A., and Jolly, C. A., *J. Gen. Virol.* **72,** 2591–2595 (1991).
25. Heidecker, G., and Messing, J., *Nucleic Acids Res.* **11,** 4891–4906 (1983).
26. Miller, J. S., and Mayo, M. A., *J. Gen. Virol.* **72,** 2633–2638 (1991).
27. Devereux, J., Haeberli, P., and Smithies, O., *Nucleic Acids Res.* **12,** 387–395 (1984).
28. Waterhouse, P. M., Martin, R. R., and Gerlach, W. L., *Phytopathology* **79,** 1215 (1989).
29. Casper, R., *Plant Viruses* **3,** 235–258 (1988).