Nucleotide sequence analysis and genomic organization of the NY-RPV isolate of barley yellow dwarf virus

Jeffrey R. Vincent,* Richard M. Lister and Brian A. Larkins†

Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907-1155, U.S.A.

Introduction

Luteoviruses, such as barley yellow dwarf virus (BYDV), beet western yellows virus (BWYV) and potato leafroll virus (PLRV), cause yellowing diseases in a wide range of host plants (Matthews, 1982), are not mechanically transmitted and are limited to the phloem tissue of the host plant. The physical properties of luteoviruses are similar in that they are 24 to 30 nm diameter isometric particles containing a positive-sense ssRNA genome of 5.6 to 6.0 kb (Mayo et al., 1989; Miller et al., 1988a; Veidt et al., 1988; Waterhouse et al., 1988). The genomic RNA has a small protein (VPg) covalently attached to its 5’ end (Murphy et al., 1989) but does not have a poly(A) tail (Mayo et al., 1982, 1989; Miller et al., 1988a; Veidt et al., 1988).

Unlike BWYV and PLRV, which typically infect dicotyledonous hosts, BYDV consists of a group of serologically related viruses that infect barley, oats, wheat, rice, maize and other graminaceous hosts (Rochow, 1970). Like other luteoviruses, they are transmitted only by aphid vectors in a persistent, circulative manner. Isolates of BYDV, originally distinguished and named for their predominant aphid vectors, have been separated into two groups based on their serological relationships (Aapola & Rochow, 1971; Rochow, 1970; Rochow & Carmichael, 1979; Rochow & Duffus, 1981), the cytopathological ultrastructure of infected cells (Gill & Chung, 1970) and the dsRNA profiles identified in infected tissues (Gildow et al., 1983). Based on these criteria, BYDV group 1 includes serotypes MAV, transmitted by Macrosiphon (=Sitobion) avenae Fabr., PAV, transmitted by M. avenae and Rhopalosiphum padi L., and SGV, transmitted by Schizaphis graminum Rond. (Rochow, 1970). Group 2 includes serotypes RMV, transmitted by R. maidis Fitch., and RPV, transmitted by R. padi (Rochow, 1970).

To help understand the molecular biology of luteoviruses, the nucleotide sequence and genome organization of a PAV serotype of BYDV (Miller et al., 1988a), and BWYV (Veidt et al., 1988) and PLRV (Mayo et al., 1989; Keese et al., 1990) have been determined. The coat protein genes of these luteoviruses (Miller et al., 1988b; Kawchuk et al., 1989, Prill et al., 1989; Veidt et al., 1988), and of isolates representing the two BYDV groups described have also been identified and characterized.
Here, we present the nucleotide sequence and genome organization of the RPV serotype of BYDV. This information is compared to that for a PAV serotype of BYDV, and also to that published for BWYV and PLRV. Analyses of sequence similarity and genome relationships are presented which indicate that the RPV serotype, although it infects and causes yellowing disease in graminaceous hosts, is more closely related to BWYV and PLRV, which typically infect dicotyledonous hosts, than to PAV serotypes of BYDV.

Methods

**Virus and RNA isolation.** The BYDV isolate used in this work was the New York RPV isolate of (Rochow (NY-RPV)(Rochow, 1970), named according to the suggestion of Rochow (1984) that isolates originating in New York State be prefixed by 'NY'. This isolate was maintained at Purdue University in oat plants (Avena sativa L. cv. Clintland 64) by mass transfer of viruliferous R. padi; the absence of cross-contamination was confirmed by ELISA as described previously (Barbara et al., 1987). Virus was purified and the RNA extracted as described for the MAV-PS1 isolate of BYDV (Barbara et al., 1987).

cDNA. A cDNA library representing the NY-RPV genome was constructed in λgt11 (Young & Davis, 1983), pUC18 (Vieira & Messing, 1982) and pGEM-3Z (Promega) (Barbara et al., 1987; Vincent et al., 1990). A restriction map representing the viral genome was generated by single and double restriction enzyme digests of the cloned cDNA, and by Southern blot hybridization between different restriction fragments (Maniatis et al., 1982).

c-DNA sequencing. Plasmid DNA isolated by an alkaline lysis technique (Birnboim & Doly, 1979) was sequenced by the dideoxy-nucleotide chain termination method (Sanger et al., 1977) with a modified T7 DNA polymerase (Sequenase; U.S. Biochemical). The cDNA clones representing the viral genome were sequenced by one or more of the following strategies. (i) Restriction fragments representing overlapping regions of the genome were subcloned; (ii) exonuclease III/mung bean nuclease-generated nested deletions were identified in existing cDNA clones (Putney et al., 1981); (iii) specific synthetic sequencing primers were prepared (Laboratory for Macromolecular Structure, Department of Biochemistry, Purdue University). Sequence analyses were performed using Microgenie version 4.0 (Beckman Instruments) and the Genetics Computer Group sequence analysis software, version 6.1 (Devereux et al., 1984).

Western blot analysis. For Western blot analysis of NY-RPV, purified virus particles were dissociated by the addition of an equal volume of 2 x SDS-PAGE sample buffer (Laemmli, 1970), incubated overnight at 4 °C and electrophoresed on a 10% polyacrylamide gel. Viral proteins were detected immunologically with 0.15% Tween 20 used to prevent non-specific binding (Barbara et al., 1987).

Results and Discussion

Sequencing

Overlapping cDNA clones representing NY-RPV genomic RNA were identified by restriction enzyme analysis and hybridization (Fig. 1). From them, 5600 nucleotides of the NY-RPV genome were determined (Fig. 2), of which 100% was determined from sequencing of both strands and 74% was determined from more than one independently derived cDNA clone. The cDNA inserts in clones pRP11A, pRP23, pRP29, pRP95, pRP104, pRP3-2 and pRP1-8 were sequenced in their entirety, whereas only a partial sequence for the remaining clones shown in Fig. 1 was determined. Clones pRP29 and pRP1-8 represent the extreme 5' and 3' ends of the NY-RPV sequence obtained, respectively; no other clones containing sequences coterminal with those obtained from either pRP29 or pRP1-8 were identified. Therefore, only pRP29 was used to determine the 5'-proximal 366 nucleotides and only pRP1-8 was used to determine the 3'-proximal 107 nucleotides.

Sequencing of more than one independently derived cDNA clone representing several regions of the genome identified six varying nucleotides (Fig. 2). Of these, half were C to U transitions and half were transversions, and three result in amino acid changes (Fig. 2); the other three changes, occurring in the 'wobble' position, would not result in any amino acid change. We do not have any indication whether the observed nucleotide differences are cloning artefacts or whether they represent true variability associated with the viral RNA population from which the cDNA library was constructed.

Open reading frames (ORFs)

A computer analysis of the NY-RPV nucleotide sequence identified six ORFs on the positive-sense strand capable of encoding polypeptides greater than...
The NYP-RPV isolate of BYDV

The amino acid sequences encoded by the six major positive-sense ORFs are indicated within the nucleotide sequence in Fig. 2 and are referred to as ORF1 to ORF6 in order of their proximity to the 5’ end of the sequence.

ORF1 begins at nucleotide 115, which also begins the first AUG in the genome, and potentially encodes 256 amino acids of Mr 29263. ORF2 begins with the second AUG in the genome at nucleotide 266 and potentially encodes a polypeptide of 654 amino acids (Mr 70575). Thus, ORF2 overlaps the 3′ end of ORF1 by 620 nucleotides. The first AUG of ORF3 begins at nucleotide 1813, which is 213 nucleotides downstream of the previous in-frame termination codon, and could encode a polypeptide of 633 amino acids (Mr 71785). ORF3 begins with the second AUG in the genome at nucleotide 10K, as well as two on the negative-sense strand (Fig. 3).

Located completely within ORF4, but in a different reading frame, is 9459 base ORF which is capable of encoding a protein of Mr 17211 (ORF5). Recently the Mr of the Vpg of a different RPV isolate has been determined to be 17K (Murphy et al., 1989). As only ORF5 is capable of encoding a protein of Mr 17K, it is probable that it encodes the NYP-RPV Vpg. ORF6 is a 1197 nucleotide ORF beginning at nucleotide 4302, immediately after the coat protein gene termination codon and in the same reading frame. The first AUG of ORF6 does not occur until nucleotide 4572.

Western blot analysis of purified NYP-RPV virions identified three proteins; Fig. 4 shows a typical analysis. Two smaller proteins were identified with estimated Mr s of 19K and 23K; the largest protein had an Mr of approximately 58K to 63K, and was not detectable (Fig. 4). Both the ability to detect the largest protein and its size were dependent on how often the purified virus preparation had been thawed and refrozen before analysis. When an NYP-RPV preparation had been thawed and frozen five times, the 63K protein could not be detected and a faint protein band with an apparent Mr of 58K was identified instead (Fig. 4, lane 1). When this preparation was frozen and thawed once more, the 58K band could no longer be detected.

Of the smaller proteins identified by Western blot analysis, the 23K protein was predominant and migrated in SDS-PAGE as the authentic NYP-RPV coat protein (Vincent et al., 1990); the band migrating as a 19K protein was presumably the Vpg. No separate ORF is apparent that would account for the largest protein observed by Western blot analysis of NYP-RPV virions. However, readthrough of the coat protein gene termination codon could produce a protein of Mr 65K. The implication is that ORF4 and ORF6 together can encode a 65K protein by readthrough of the ORF4 stop codon, and that this protein is associated with the viral capsid. Similarly, readthrough of the PLRV coat protein gene termination codon recently has been shown (Bahner et al., 1990).

Non-coding regions

The NYP-RPV genome contains 401 nucleotides that are not contained within the six major positive-sense coding regions. These non-coding regions are located at the extreme 5′ and 3′ ends of the genome, and 5′ to the coat protein gene (ORF4). We have no evidence that these 5′- and 3′-terminal sequences represent the actual termini of NYP-RPV. We identified 114 nucleotides in the 5′-proximal non-coding region and 102 nucleotides in the 3′-proximal non-coding region. Computer analysis of the 3′-terminal non-coding region did not identify any potential RNA-like structures. For PLRV and BWYV, luteoviruses with genomic organization similar to that of NYP-RPV, there are 141, and between 146 and 197 3′-terminal nucleotides, respectively (Keese et al., 1990; Mayo et al., 1989; Veidt et al., 1988). A comparison between the 5′ non-coding regions of these luteoviruses showed that although 114 nucleotides were identified at the 5′ end of NYP-RPV, the length of this region is variable; isolates of PLRV have between 32 and 174 (Keese et al., 1990; Mayo et al., 1989), and BWYV has 31 nucleotides at the 5′ terminus (Veidt et al., 1988).

However, for these luteoviruses, and also for Vic-PAV (Miller et al., 1988a), a consensus hexanucleotide sequence, AC/UAAAA/C, has been identified in these 5′ non-coding regions. This sequence is also found in southern bean mosaic virus (SMBV) (Keese et al., 1990; Wu et al., 1987).

Nearly half of all the non-coding nucleotides in NYP-RPV are located between ORF3 and ORF4, i.e. 5′ to the coat protein gene. This region is A–U-rich, with 58% of the 185 nucleotides being A or U residues. Uracil forms 42% of the 113 nucleotides immediately upstream of the coat protein gene initiation codon, and A and U residues together make up 62% of the nucleotides in this region. A comparison of similar regions from other luteoviruses indicates that there is a core region consisting of 113 nucleotides upstream of the coat protein gene initiation codon (Fig. 5). This core region contains the identifying features of subgenomic promoters defined for brome mosaic virus (BMV) (Marsh et al., 1988). There is extensive nucleotide identity between the core regions of NYP-RPV, BWYV.
Fig. 2. Nucleotide sequence (RNA form) of the BYDV NY-RPV isolate. The initiation codons for the six major positive-sense ORFs are indicated by an asterisk and the termination codons by +. The deduced amino acid sequence is indicated below the nucleotide sequence.
Fig. 3. ORFs determined from the positive- and negative-sense orientation of the BYDV NY-RPV nucleotide sequence. Vertical ticks above a line indicate the location of initiation codons; those below a line indicate the location of termination codons. The open boxes represent ORFs and the six major positive-sense ORFs are numbered according to their proximity to the 5' end of the genome. All ORFs are represented as beginning with an AUG initiation codon.

Fig. 4. Immunological detection of the virion-associated proteins from the BYDV NY-RPV isolate. Purified NY-RPV virions were dissociated and electrophoresed on a 10% polyacrylamide gel. After electroblotting onto nitrocellulose, virus-associated proteins were detected immunologically using a polyclonal anti-NY-RPV primary antiserum and an alkaline phosphatase-conjugated secondary antiserum. Sizes of proteins identified from purified NY-RPV preparations are shown to the left, those of protein size standards to the right. Lane 1, purified NY-RPV thawed and frozen five times; lane 2, purified NY-RPV thawed only once.

Fig. 5. Comparison of luteovirus non-coding regions located 5' to the coat protein coding region. Asterisks represent nucleotides identical in the sequences of BYDV NY-RPV isolate, BWYV and PLRV. A vertical bar represents nucleotides identical in all the luteoviruses examined. Numbering is in a direction 5' to the putative coat protein gene initiation codon. Boxed regions represent regions conserved in luteoviruses. RPV, BYDV NY-RPV isolate; BWYV, BWYV FL1 (Veidt et al., 1988); PLRV (Mayo et al., 1989); BYDV MAV (P. P. Ueng & R. M. Lister, unpublished results); BYDV PAV (Millet et al., 1988a; P. P. Ueng & R. M. Lister, unpublished results).

Comparison with other luteoviruses

The NY-RPV sequence is the second BYDV genome sequence to be reported. The genome of a BYDV PAV serotype from Victoria, Australia, Vic-PAV, has been determined to consist of 5677 nucleotides (Miller et al., 1988a) and encodes six positive-sense ORFs. The genome organizations of NY-RPV and Vic-PAV show significant differences (Fig. 6). Although the RNAs of both contain approximately 5600 nucleotides, there is no similarity in genome organization and little sequence similarity in regions of the genomes 5' to the coat protein genes. The coat protein gene itself is located 3'-proximal in NY-RPV, whereas it is near the centre of the genome of Vic-PAV.
isolates. The location of major ORFs of NY-RPV and Vic-PAV (Miller et al., 1988a) and the reading frames in which they occur are represented by boxes. The sizes of the polypeptides potentially encoded by each ORF are indicated.

However, similarities between Vic-PAV and NY-RPV are found in the 22K ORFs, in the internal 17K ORFs and in the immediately 3'-proximal ORFs (i.e. NY-RPV ORF6 and Vic-PAV ORF5). The 22K ORFs are the coat protein genes and contain the presumptive 17K VPg gene as an internal ORF. This organization has been found in all luteoviruses examined to date (Vincent et al., 1990). Nucleotide and deduced amino acid sequence identities between NY-RPV and Vic-PAV ORFs determined using Microgenie sequence alignment software were 55.2% and 48.8% for the coat protein coding region, and 44.6% and 32.2% for the full ORFs located immediately 3' to the coat protein coding region. The ORF 3'-proximal to the NY-RPV coat protein ORF is proposed to be expressed by readthrough of the coat protein gene termination codon, as with a similar ORF of Vic-PAV and, indeed, with luteoviruses in general. A conserved 14 nucleotide sequence surrounding the coat protein gene termination codon in BYDV (Fig. 6) has been identified previously, and a role for this sequence in the expression of the readthrough ORF postulated (Smith & Harris, 1990; Veidt et al., 1988). An examination of this region from NY-RPV indicates that this conserved region is actually the 12 nucleotide sequence AAAUAGGUAGAC, with the coat protein gene termination codon indicated by the underlined nucleotides.

The genome organization of NY-RPV, although it differs from that of Vic-PAV, is very similar to that reported for both BWYV (Veidt et al., 1988) and PLRV (Keese et al., 1990; Mayo et al., 1989). Furthermore, the sizes of the putative protein products encoded by each ORF of the latter viruses are similar to those encoded by the ORFs of NY-RPV.

There is considerable interest in evolutionary relationships between viruses based on amino acid sequence motifs associated with the RNA-dependent RNA polymerases (viral replicases) (Habili & Symons, 1989; Kamer & Argos, 1984; Poch et al., 1989) and nucleic acid helicases (Gorbalenya et al., 1988; Gorbalenya & Koonin, 1989; Habili & Symons, 1989; Hodgman, 1988a, b). These relationships are based both on sequence similarity within motifs and on the distance separating motifs. Using such a classification scheme, positive-sense RNA plant viruses have been placed into three supergroups, with the luteoviruses in supergroup B (Habili & Symons, 1989).

The Vic-PAV replicase shows substantial similarity to the replicase of carnation mottle virus (Miller et al., 1988a), and the replicases of BWYV and PLRV show homology to that of SBMV (Mayo et al., 1989; Veidt et al., 1988), all members of supergroup B. A BESTFIT comparative analysis (Devereux et al., 1984) of the amino acid sequence deduced from NY-RPV ORF3 with that of the putative replicase-encoding region of SBMV (Wu et al., 1987) has identified a region containing the GDD amino acid motif for RNA-dependent RNA polymerases (Kamer & Argos, 1984). This region is also similar to regions of both PLRV and BWYV (Fig. 7). The four amino acid sequence motifs of supergroup B positive-sense RNA viral replicases (Habili & Symons, 1989) (VPTDCSGFDWS-53-GVQKGSYNTSTNSRV-11-WAIAMGDDALEA-22-FCSRIF) were also identified within this region (Fig. 7), although the last motif was not represented in SBMV. The NY-RPV replicase motifs showed greater similarity in both sequence and gap distances to comparable regions in BWYV and PLRV, than to those in Vic-PAV.

Seven amino acid sequence motifs have been identified for proteins that have helicase functions (Habili & Symons, 1989; Hodgman, 1988a, b), and have been used to evaluate the taxonomic relationships between luteoviruses and other positive-sense RNA plant viruses (Habili & Symons, 1989). The NY-RPV sequence contains three of the seven helicase motifs. These regions resemble the ‘A’ and ‘B’ sites of the catalytic site of nucleoside triphosphate-utilizing enzymes (Gorbalenya et al., 1988; Gorbalenya & Koonin, 1989) and supergroup B motif VI (Habili & Symons, 1989). The potential ‘A’ and ‘B’ sites are located in ORF2 (356-AGYFY-GKT-41-LVFPEE-240), and motif VI has been identified in ORF3 (351-QSKLEDGRRYRLIMSVS-267). The NY-RPV helicase motifs were compared with those from other luteoviruses (Habili & Symons, 1989). Motif VI has been found in all luteoviruses, but although all other luteoviruses contain motif IV, no such region was identified in NY-RPV. Regions comparable to the ‘A’ and ‘B’ motifs were not identified in PLRV or Vic-PAV,
but similar sequence and gap lengths were found in BWYV (356–SPYFNGKT–38–VFAQED–194) (Veidt et al., 1988).

It has been proposed that the luteovirus viral replicase is expressed as a result of a translational frameshift allowing two overlapping ORFs to produce one protein (Keese et al., 1990; Mayo et al., 1989; Miller et al., 1988a; Veidt et al., 1988). As both ORF2 and ORF3 within the NY-RPV genome are associated with helicase function, it seems probable that both the helicase and the viral replicase are expressed by a translational frameshift mechanism. For the luteovirus Vic-PAV, a region for site-specific ribosomal slippage has been identified within the 16 nucleotide overlap between these ORFs. For other luteoviruses, including NY-RPV, the region of overlap is considerably greater, including up to several hundred nucleotides, but contains no obvious site for ribosomal slippage.

**Taxonomic considerations**

The genome of the NY-RPV isolate of BYDV shares significant properties with other luteoviruses. These include (i) a replicase-encoding region presumably expressed by translational frameshift, (ii) an ORF, presumably encoding the VPg, completely contained within the 22K coat protein gene and (iii) a highly conserved nucleotide sequence surrounding the 22K coat protein gene termination codon, which potentially allows the formation of a larger fusion protein. There is also considerable similarity between all luteovirus 22K coat proteins, although the NY-RPV coat protein is more like those from BWYV and PLRV than those from the MAV-PS1 and P-PAV isolates of BYDV (Vincent et al., 1990). The greater similarity between the NY-RPV coat protein gene sequence and those of BWYV and PLRV, the similarity between the genome organization of NY-RPV, BWYV and PLRV, and the relative dissimilarity between the genome organization of NY-RPV and Vic-PAV are all consistent with the hypothesis that NY-RPV is a strain of BWYV (Casper, 1988). Hence, NY-RPV should be classified as distinct from BWYV and PLRV, the similarity between the genome organization of NY-RPV, BWYV and PLRV, and the relative dissimilarity between the genome organization of NY-RPV and Vic-PAV are all consistent with the hypothesis that NY-RPV is a strain of BWYV (Casper, 1988). Hence, NY-RPV should be classified as distinct from BWYV and PLRV. However, differences in aphid vectors (i.e. *R. padi* and *Myzus persicae*) and host range (i.e. monocotyledonous and predominantly dicotyledonous hosts), suggest that NY-RPV should also be classified as distinct from BWYV and PLRV.

We thank C. Logan for her photographic services and M. Karanjkar for his assistance in the Western blot analysis. This work was funded by Grant 88-37263-3855 from the Competitive Research Grants Office of the U.S. Department of Agriculture and Grant 1484670 from The Quaker Oats Company. The Genetics Computer Group Sequence Analysis Software Package for the VAX was supported by NIH grant no. AI27713. This is Journal Series Article 12706 of the Purdue University Agricultural Experiment Station.
References


(Received 27 February 1991; Accepted 10 June 1991)