Volume 16 Number 21 1988 Nucleic Acids Research

Nucleotide sequence of beet western yellows virus RNA

Isabelle Veidt, Hervé Lot[1], Mathias Leiser[2], Danièle Scheidecker, Hubert Guilley, Kenneth Richards and Gérard Jonard

Institut de Biologie Moléculaire des Plantes, 12 rue de Général Zimmer, 67000 Strasbourg, [1]Station de Pathologie Végétale, INRA, Domaine Saint-Maurice, 84140 Montfavet, France and [2]Institut für Phytopathologie Aschersleben, Theodor-Roemer-Weg, Aschersleben 4320, GDR

Received September 30, 1988; Accepted October 10, 1988 Accession nos X13062, X13063

ABSTRACT

The nucleotide sequence of the genomic RNA (5641 nt) of beet western yellows virus (BWYV) isolated from lettuce has been determined and its genetic organization deduced. The sequence of the 3'terminal 2208 nt of RNA of a second BWYV isolate, obtained from sugarbeet, was also determined and was found to be very similar but not identical to that of the lettuce isolate. The complete sequence of BWYV RNA contains six long open reading frames (ORFs). A cluster of three of these ORFs, including the coat protein cistron, display extensive amino acid sequence homology with corresponding ORFs of a second luteovirus, the PAV isolate of barley yellow dwarf virus (BYDV) (1,2). The ORF corresponding to the putative viral RNA-dependant RNA polymerase, on the other hand, resembles that of southern bean mosaic virus. There is circumstantial evidence that expression of the BWYV RNA polymerase ORF may involve a translational frameshift mechanism. The ORF immediately following the coat protein cistron may be translated by in-frame readthrough of the coat protein cistron amber termination codon. Similar mechanisms have been proposed for expression of the corresponding ORFs of BYDV(PAV) (1).

INTRODUCTION

The luteoviruses are a large group of plant viruses which cause economically significant yellowing diseases in a wide range of hosts (3). Three important members of the group are barley yellow dwarf virus (BYDV), potato leafroll virus (PLRV) and beet western yellows virus (for reviews see 4,5). Luteoviruses are phloem-limited and are consequently present in only very low concentrations in plant tissue extracts. Furthermore, they are not transmissible by mechanical inoculation but require use of an aphid vector. Because of these properties the luteoviruses, in spite of their economic importance, have been subject to little investigation at the molecular level.

The luteoviruses for which information is available have isometric virions consisting of multiple copies of a major coat protein of ca. 24,000 M_r. The genome consists of a single-stranded plus-sense RNA of about 1.8-2X10^6 M_r with a genome linked protein (VPg) (6,7) and no polyA tail (6). Beet western yellows virus (BWYV) is probably the most commonly encountered luteovirus, infecting a wide range of dicotyledons (4). Furthermore, several luteoviruses originally classified as distinct entities such as beet mild yellowing virus and the RPV and RMV isolates of BYDV are serologically very close to BWYV and should

probably be classified as BWYV strains (5). In this paper we present the sequence of cloned overlapping DNA copies of RNA from a lettuce isolate of BWYV which, taken together, define the complete BWYV genome. Partial sequence data is also presented for an isolate of BWYV obtained from sugarbeet. The resulting information about open reading frames and encoded proteins provides a framework for further studies of luteovirus genome expression. Comparison of the BWYV sequence with data for other members of the group, notably the recently completed sequence of the PAV isolate of BYDV (1), should help to unravel luteovirus taxonomy. Finally, the sequence has revealed a hitherto unexpected relation between BWYV and southern bean mosaic virus.

MATERIALS AND METHODS

Virus. BWYV (isolate FL1) was originally collected from lettuce in the vicinity of Avignon, France and was maintained by serial transfer in *Physalis floridiana* using *Myzus persicae* as vector. The virus was propagated in *Montia perfoliata* for four weeks prior to purification as described (8) except that 1.5% Rohament P (9) or 1.5% Extractase P20X (Finnsugar Biochemicals) was used during extraction. After fractionation through a sucrose gradient, virus from the peak fractions was concentrated by centrifugation and taken up in 10 mM sodium borate, pH 7.2, 1 mM EDTA. RNA was extracted from the purified virions with phenol and the nucleic acid in the aqueous phase was concentrated by ethanol precipitation before being dissolved in sterile water. The sugarbeet isolate of BWYV, BWYV(GB1), was collected from sugarbeet in the German Democratic Republic (10) and maintained on *Montia perfoliata* using *Myzus persicae* as vector. Virus was purified as described (11). Purified virions were resuspended in 0.1 M Tris-Cl, pH 7.4, 50 mM NaCl, 10 mM EDTA and .2% SDS. Proteinase K was added at 200 µg/ml and the suspension was incubated for 1 hr at 37° after which the mixture was emulsified with an equal volume of phenol/chloroform (v/v) at 60° for 10 minutes. After centrifugation the aqueous phase was reextracted twice with chloroform, once at 60° and once at room temperature. RNA was concentrated by ethanol precipitation and dissolved in sterile water.

cDNA synthesis and cloning. BWYV RNA (5µg) was 3'polyadenylated in 50 µl reaction mixture containing 50 mM Tris-Cl, pH 8, 250 mM NaCl, 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 100 µM ATP, 40 µg/ml bovine serum albumin (BRL) and 2 units polyA polymerase (BRL). Polyadenylation was allowed to proceed for 6 min. at 37°. The reaction was stopped by phenol extraction and RNA was precipitated from the aqueous phase by addition of 2 volumes of ethanol. The RNA precipitate was dissolved in 10 µl sterile water. Recombinant cDNA clones were obtained as described (12) with minor modifications (13). In some experiments plasmids were tested for the presence of viral cDNA inserts by colony hybridization using nicked BWYV RNA which had been 5'-^{32}P-labelled with polynucleotide kinase as probe (14). In other cases DNA minipreparations from recombinant

plasmids were screened for >1kB inserts by restriction endonuclease digestion followed by agarose gel electrophoresis. The nature of each long cDNA insert was then examined by dot-blot hybridization of immobilized total RNA from BWYV-infected or healthy *Physalis floridiana* using nick-translated plasmid DNA as probe. This test was deemed essential because the viral RNA preparations contained small amounts of cellular DNA. However, all the plasmids examined proved to harbour cDNA inserts of viral origin as judged by their ability to hybridize with the RNA extracted from virus-infected but not from healthy plant tissue.

Recombinant cDNA plasmids extending further toward the 5' terminus were obtained by primer extension using the synthetic oligodeoxynucleotide 5'-CTTCAACCCACCCACGG (nt 2254-2270) as primer. After reverse transcription the RNA-cDNA hybrids were treated with RNase H (BRL) and DNA polymerase I (Boehringer), dC-tailed and cloned into Pst I-cut dG-tailed pUC 9 (15,16). Plasmids were screened for viral cDNA inserts as described above.

Sequence analysis. cDNA inserts were sequenced by partial chemical degradation (17) after digestion with appropriate restriction enzymes and 5'end-labelling with polynucleotide kinase, or by the dideoxynucleotide triphosphate chain termination method (18) after subcloning of appropriate restriction fragments into pUC 9. Miniprep plasmid DNA for sequencing experiments was prepared as described (19). The double-stranded DNA (about 8 µg in 40 µl H_2O) was denatured by addition of 10 µl 1 M NaOH followed by incubation at 0° for 10 min. An equal volume of 4 M NH_4 acetate and 4 volumes of ethanol were then added and the precipitated DNA was collected by centrifugation, washed once with 67% ethanol and resuspended in 15 µl H_2O. Sequence was determined on the denatured plasmid DNA by the dideoxynucleotide triphosphate chain termination method from direct or reverse sense synthetic oligodeoxynucleotide sequencing primers (Biolabs) and bacteriophage T7 DNA polymerase (Sequenase, USB) (20).

The sequence of the first 127 nt of the viral RNA was obtained by reverse transcription from the 5'-^{32}P-labelled oligodeoxynucleotide primer 5'-TTGCGCTCTTTGA-TTGCTCGTAC, complementary to nt 149-171 of the final sequence, and analysis of the resulting 5'-labelled cDNA runoff transcript by partial chemical degradation. In some experiments the 5'-labelled cDNA was tailed with dC or dT residues using terminal deoxynucleotidyl transferase (21) prior to sequence analysis. Sequence data were analyzed with UWGCG programs (22) on a Microvax II computer.

Synthetic Transcript. The Dra I-EcoRI fragment (nt 3313-4321) of the pBW17 cDNA insert was purified by electrophoresis through a low melting point agarose gel and fused between the Hinc II and EcoR I sites of pBS(-) (Stratagene) to yield pBW(3313-4321)BS. After linearization of the plasmid with EcoR I, a plus-strand RNA transcript corresponding to the inserted DNA was synthesized using bacteriophage T3 polymerase (Stratagene) (23). The transcript was translated in rabbit reticulocyte lysate (23) and the ^{35}S-methionine-labelled translation products were immunoprecipitatedwith BWYV(FL1) antiserum or nonimmune

serum as described (24) except that 10% Pansorbin (Calbiochem) was substituted for protein A-sepharose. Immunoprecipitated products were fractionated by SDS-polyacrylamide gel electrophoresis and visualized by autoradiography of the dried gel.

RESULTS AND DISCUSSION

Sequence analysis. No information is available concerning the 3'extremity of BWYV RNA. The genomic RNA of the luteoviruses PLRV and BYDV(PAV), however, have been shown to contain no 3'polyA sequence (1,6) and it seems reasonable to assume that this is also the case for BWYV. Consequently, the BWYV RNA for use in cDNA synthesis was first 3'polyadenylated by treatment with polyA polymerase in the presence of ATP. Recombinant clones were then synthesized from this material using Pst I-cut oligodT-tailed pUC 9 to prime first-strand cDNA synthesis (12). About 50 clones were obtained which contained viral cDNA inserts of 1 kB or longer. All the cDNA inserts were found to be 3'coterminal as judged by restriction enzyme mapping. Preliminary sequence analysis of three clones with long cDNA inserts revealed that all three had a 3'polyA tail and that the sequence immediately upstream of the tail was identical in all three cases over at least 300 nt, indicating that the polyadenylation site corresponds to an abundant 3'OH extremity, presumably that of the viral genomic RNA.

The cDNA insert of pBW17, the longest recombinant plasmid obtained in the foregoing experiment, was sequenced in its entirety. It corresponds to nt 2016-5641 of the final sequence (Fig. 1). In another cloning experiment a synthetic oligodeoxynucleotide complementary to a portion of the pBW17 sequence near its 5'extremity was used to prime a second round of cDNA synthesis. Analysis of a series of overlapping recombinant cDNA plasmids obtained in this experiment extended the sequence up to nt 127 (Fig 1).

The sequence of the 5'terminal portion of the viral RNA was established by primer extension using a 5'-^{32}P-labelled synthetic oligodeoxynucleotide complementary to nt

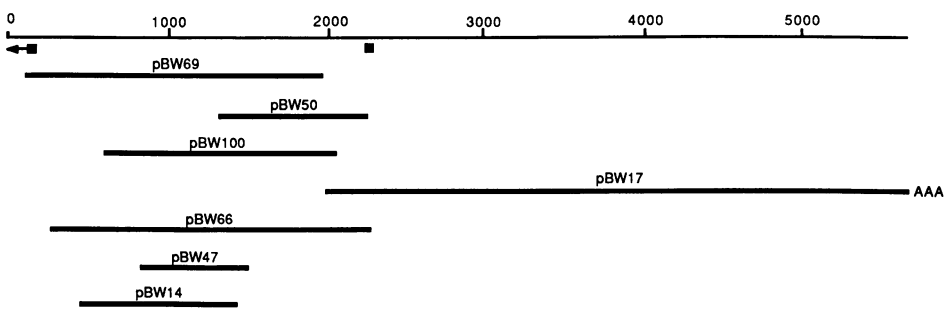


Figure 1. Map of cDNA inserts of clones used to determine the BWYV(FL1) sequence. Filled squares indicate positions of synthetic oligodeoxynucleotides used to prime specific cDNA synthesis. Arrow indicates 5'terminal region sequenced by primer extension directly from the RNA.

149-171. After hybridization of the primer to viral RNA and extension to the 5'terminus of the RNA with reverse transcriptase, the resulting 5'-labelled full-length runoff cDNA migrated as a single band upon polyacrylamide gel electrophoresis (data not shown). The band was eluted from the gel and sequenced by partial chemical degradation. The resulting sequence could be read with confidence up to nt 2 at which point the heavy radioactive band of nondegraded cDNA extends across all four lanes of the sequencing gel and obscures the pattern (Fig. 2A). Similar results were obtained whether or not the RNA was treated with proteinase K prior to cDNA synthesis (data not shown). In order to characterize the terminal residues a second aliquot of the 5'-labelled cDNA was tailed with dT or dC residues before being subjected to sequence analysis. This treatment shifts the heavy band of nondegraded material to lower mobility in the sequencing gel and so permits the last 2 residues of the cDNA sequence proper,3'-TG..., to be determined (Fig. 2B,2C). It remains to be seen if cDNA synthesis in these experiments has proceeded all the way to the RNA 5'extremity, as appears to be the case for other RNAs possessing a VPg (25) or whether synthesis has terminated prematurely due to steric hindrance by the VPg. It should be noted, however, that, if the established sequence, 5'-ACAAAAGAAA..., does indeed represent its 5'extremity, then BWYV RNA resembles the VPg-containing RNA of the potyvirus tobacco vein mottling virus in beginning with an A residue (25). Other viral RNAs known to posess a VPg begin with a pyrimidine (e.g. 26-29).

As noted above, the three plasmids containing the 3'terminal region that were analyzed proved to be 3'coterminal upstream of the polyA sequence, strongly suggesting that this site corresponds to the 3'extremity of the viral RNA. The possibility remains open,

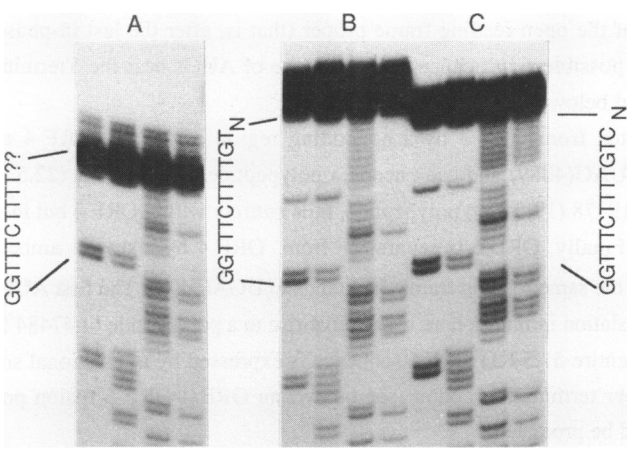


Figure 2. Partial chemical degradation of 5'-labelled runoff cDNA complementary to the 5'terminal sequence of BWYV(FL1) RNA (A). The annotated sequence corresponds to the cDNA and is complementary to that of viral RNA. (B,C) The same as in panel A except that the runoff cDNA was tailed with dT residues (B) or dC residues (C) prior to sequence analysis. The chemical degradation reactions are specific for (from left to right) G, G+A, C+T and C.

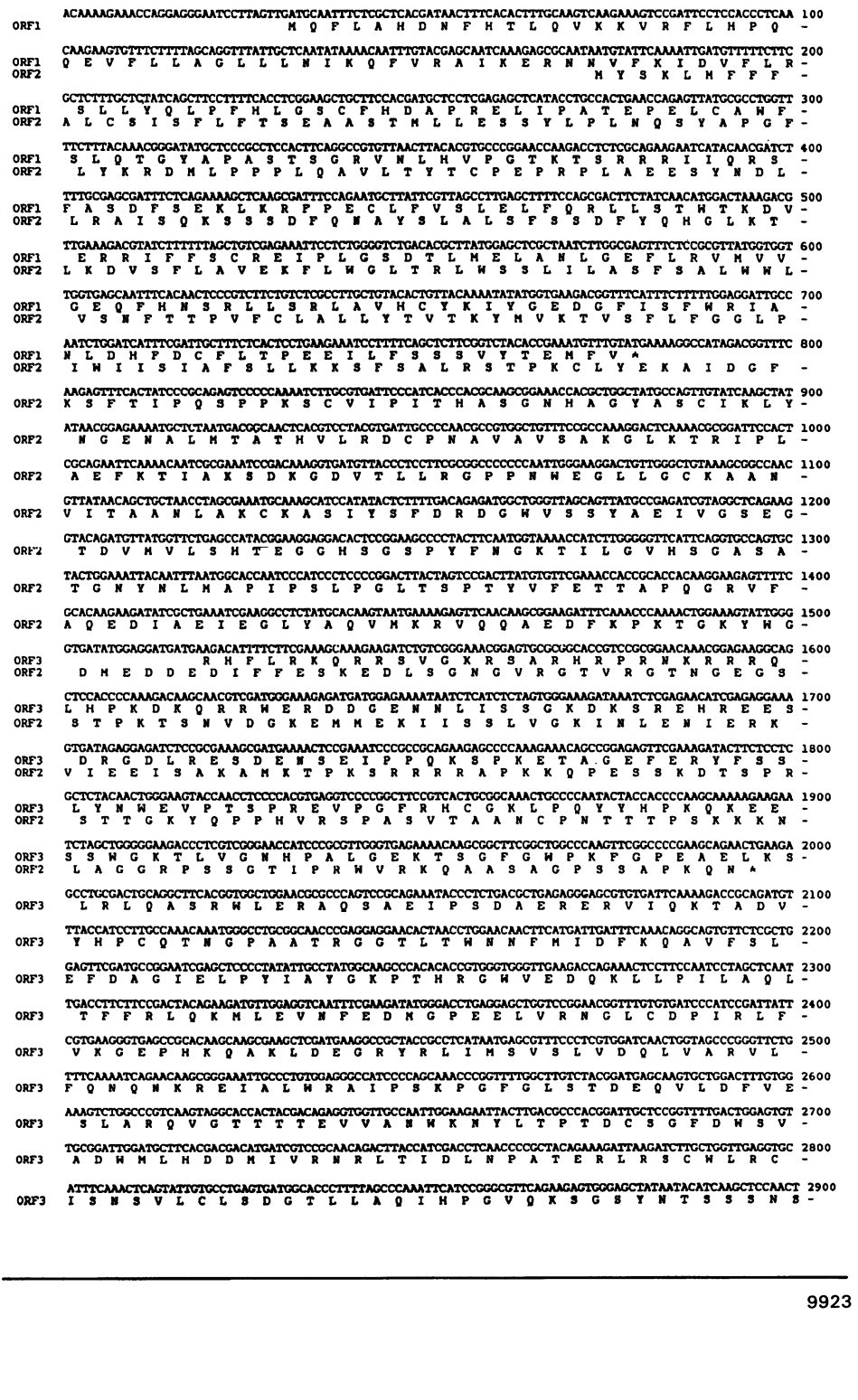
however, that the viral RNA in fact terminates with one or a few A residues, which would be indistinguishable from the polyA sequence added *in vitro* for cloning purposes.

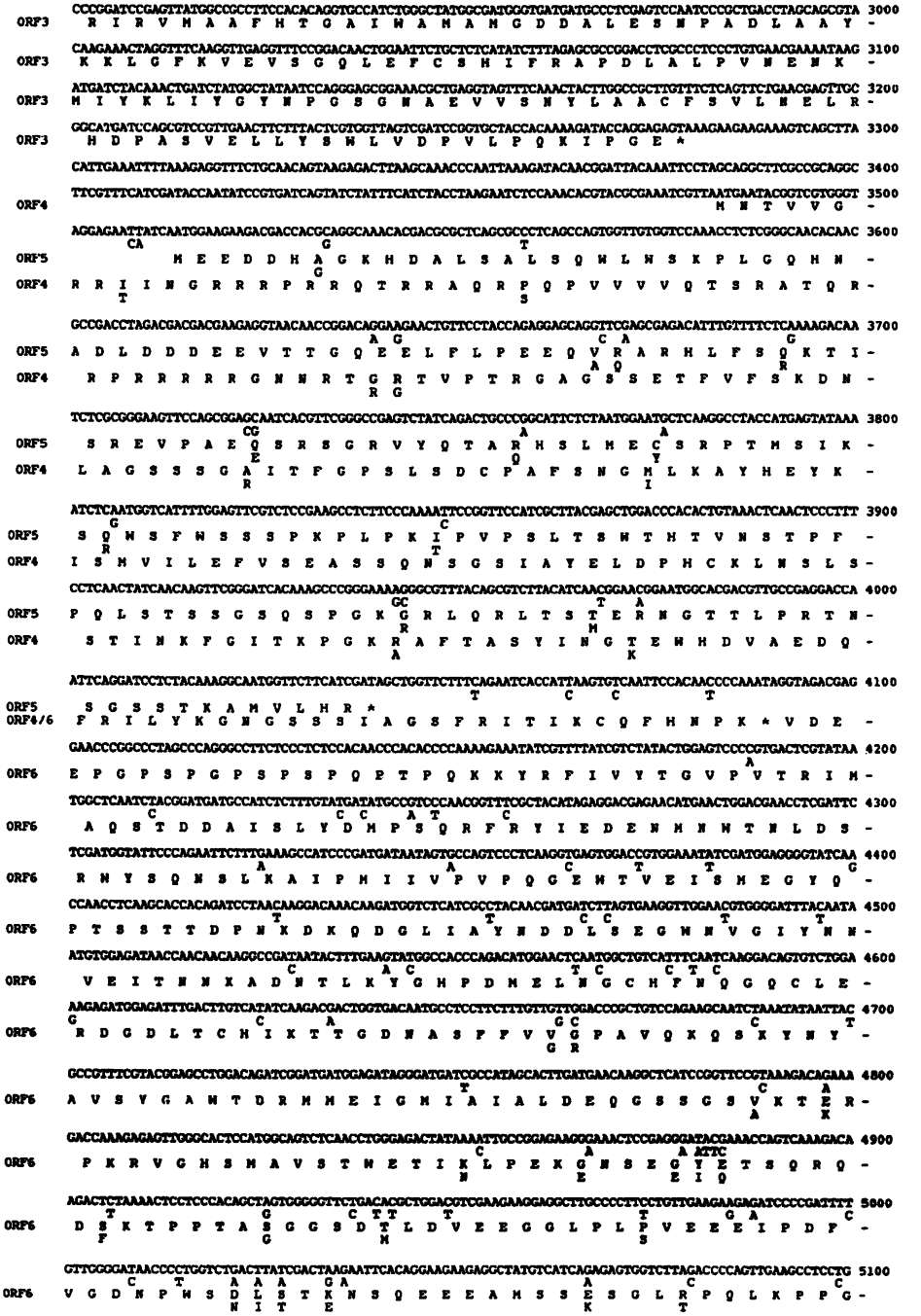
The complete sequence of BWYV(FL1) RNA is presented in Fig. 3 along with the amino acid sequence of the long open reading frames discussed below. The length of 5641 nt (excluding possible additional A residues at the 3'extremity) is in good agreement with the value of $1.85X10^6$ measured for BWYV RNA by gel electrophoresis (30). Where sequence data were obtained from several overlapping clones the observed sequences were, with one exception, identical indicating that there was little sequence heterogeneity within the population of viral RNA molecules used for cDNA synthesis. The exception was clone pBW69, which lacked two residues at positions 1152-1153 compared to plasmids pBW14, PBW47 and pBW66 (Fig. 1). The deletion in pBW69 is presumably a cloning artifact or has arisen from cDNA synthesis on a variant RNA molecule. In the latter case the putative variant RNA molecule is likely to be defective because the deletion would lead to premature termination of a long open reading frame. For this reason the variant sequence is not shown in Fig. 3.

Open reading frames. Fig. 4 presents the position of all AUG initiation and UAG, UGA and UAA termination codons in the three possible reading frames of the BWYV RNA sequence. The six longest open reading frames (ORFs) are identified by number. ORF 1 begins at AUG(32) and ends at UGA(779). It can encode a polypeptide of 29080 M_r (29 KD). Overlapping extensively with ORF 1 but in a different reading frame, ORF 2 begins at AUG(174) and ends at UGA(1995). It can encode a polypeptide of 66202 M_r (66 KD). The 5'terminal portion of ORF 3 overlaps the 3'terminus of ORF 2 for 474 nt. ORF 3 is unusual in that the first in-phase initiation codon in its sequence, AUG(2168), occurs 216 codons after the beginning of the open reading frame proper (that is, after the last in-phase stop codon upstream). The possible significance of the absence of AUGs near the 5'terminus of ORF 3 will be discussed below.

Separated from ORF 3 by a noncoding region of 202 nt, ORF 4 extends from AUG(3483) to UAG(4089) and can encode a polypeptide of 22468 M_r (22.5 KD). ORF 5, encoding a M_r 19478 (19.5 KD) polypeptide, falls entirely within ORF 4 but in another open reading frame. Finally, ORF 6 is separated from ORF 4 by a single amber termination codon and is in the same reading frame. It extends to UGA(5493). The first AUG in ORF 6 is at nt 4200. Translation initiation here would give rise to a polypeptide of 47484 M_r. If, on the other hand, the entire 51.5 KD ORF 6 sequence is expressed by translational suppression of the ORF 4 amber termination codon (see below) an ORF 4-ORF 6 fusion polypeptide of 74021 M_r would be produced.

Assuming that all the aforesaid ORFs are expressed as proteins, the noncoding sequence of BWYV(FL1) RNA consists of 379 nt (6.7% of the total sequence), 31 nt at the 5'extremity, 146 nt at the 3'terminus and 202 nt in the intergenic region separating ORFs 3 and 4. The longest ORF in the BWYV minus-strand sequence is 158 codons in length (nt





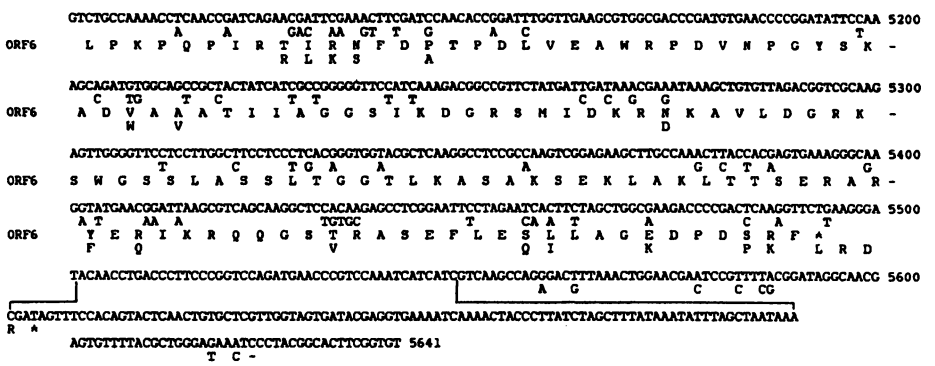


Figure 3. Sequence of BWYV(FL1) RNA. Sequence differences observed in the 3' terminal 2208 nt of BWYV(GB1) RNA are written below the FL1 nucleotide sequence. Deduced amino acid sequences of long open reading frames referred to in the text are indicated below the nucleotide sequence. Amino acid substitutions in isolate GB1 are also shown.

1460-1933). In view of the lack of evidence for expression of information encoded by the minus-strand of other plus-strand RNA plant viruses the biological significance of this ORF is doubtful.

Sequence analysis of BWYV from sugarbeet. In order to gain information about the sequence variability among different BWYV isolates several cDNA clones were prepared from RNA of BWYV(GB1), an isolate collected from sugarbeet in the German Democratic Republic. Sequence analysis of two cDNA inserts defined a sequence of 2208 residues extending from nt 3484 (numbering refers to the BWYV(FL1) sequence) to the 3'extremity (Fig. 3). The sequence includes ORFs 5 and 6 and all of ORF 4 but the A residue of the AUG initiation codon. Comparing BWYV(FL1) and GB1 reveals 145 point mutations in the coding regions. The base changes are not evenly distributed along the sequence. Thus the 525 nt region where ORFs 4 and 5 overlap contains only 17 base changes (3.2% of the positions) whereas 122 of 1413 positions (8.6%) in ORF 6 have changed (Fig. 5). Apparently the overlap of the ORF 4-ORF 5 sequences imposes constraints on the mutation rate. All 17 of the base changes in the region of open reading frame overlap result in amino acid changes in one or the other (or both) of the corresponding amino acid sequences with ORF 4 being more conserved. Thus the base changes result in 11 amino acid substitutions in the ORF 5 polypeptide but only 5 in the overlapping ORF 4 sequence. Even more striking is the uneven pattern of amino acid substitution in ORF 6 (Fig. 5). Thirty base changes occur in the 5'terminal 571 nt of this ORF but all are silent. The 115 base changes in the 3'terminal 60% of this ORF, on the other hand, result in frequent amino acid substitutions. Hence the BWYV ORF 6 polypeptide appears to consist of two domains, a highly conserved N-terminal domain and a more variable C-terminal region.

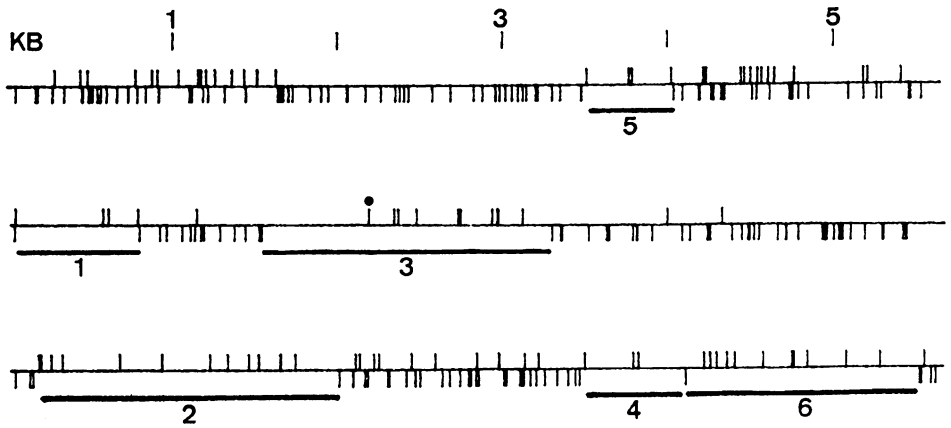


Figure 4. Open reading frames in the plus-strand of BWYV(FL1) RNA. Position of initiation codons (verticle lines pointing upward) and termination codons (verticle lines pointing downward) are shown for the three frames. Horizontal bars numbered 1-6 indicate the six long ORFs referred to in the text. The circle indicates the position of the first AUG in ORF 3. For reasons given in the text ORFs 3 and 6 are shown extending from one in-phase termination codon to the next. The other ORFs are shown beginning at the first in-frame initiation codon.

A notable difference between the lettuce and sugarbeet isolates occurs near the beginning of the 3'noncoding region. Here, the 47 nt sequence at positions 5502-5548 in BWYV(FL1) is replaced in GB1 by a completely unrelated sequence of 98 nt (Fig. 3). The remainder of the 3'noncoding regions of the two isolates is rather similar. We do not yet know to what extent the various sequence differences detected here reflect "normal" variation between geographically distinct BWYV isolates or whether some may represent adaptations to different hosts.

Homologies with other plant viruses. The sequence of a second luteovirus, the PAV isolate of BYDV, has recently been established (1). Comparison of this sequence to that of

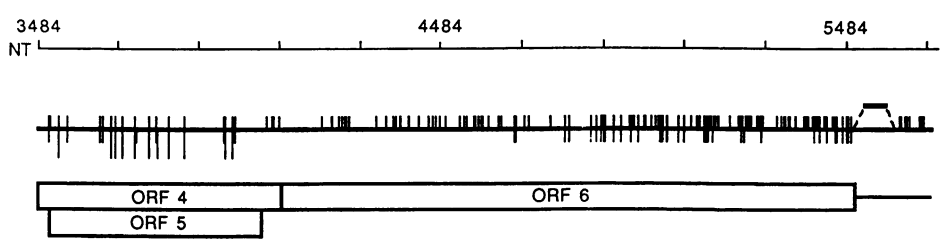


Figure 5. Map of nucleotide changes (lines pointing upward) and resulting amino acid substitutions (lines pointing downward) observed in the 3'terminal portion of the GB and FL1 sequences. Short lines pointing downward are amino acid differences in ORFs 4 and 6; longer lines pointing down are differences in ORF 5. Numbering refers to the BWYV(FL1) sequence.

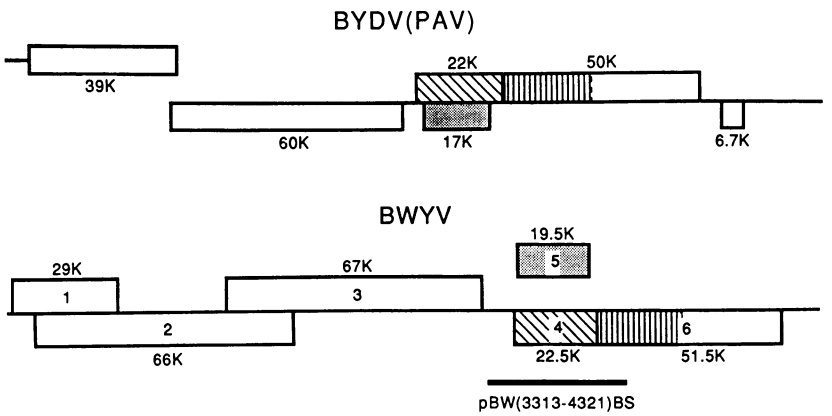


Figure 6. Position of long open reading frames in genomic RNAs of BYDV(PAV) (1) and BWYV. The size in kilodaltons (K) of polypeptides encoded by each ORF is shown. Similar shading indicates regions of amino acid sequence homology between the ORFs of the two viruses. The extent of the cDNA insert transcribed in pBW(3313-4321)BS is shown below the BWYV map.

BWYV reveals amino acid homology for some but not all of the open reading frames. Fig. 6 presents the genome organization of BYDV(PAV) and BWYV. Total genome length of the two viruses is very similar, 5641 nt for BWYV and 5677 nt for BYDV, although there is one less long ORF in BYDV. This difference is compensated by the longer 3'terminal sequence following the 50KD ORF in BYDV. A 6.7 KD ORF is also present in this 3' region of BYDV for which there is no equivalent in BWYV (Fig. 6).

The BYDV 22 KD polypeptide has been shown to correspond to the viral coat protein (2). Comparison of its amino acid sequence to that of the various BWYV ORFs reveals strong homology with the 22.5 KD polypeptide encoded by the similarly situated cistron in BWYV (Fig. 6). Computer-assisted comparison of the two sequences produced an optimal alignment in which the two sequences match (either identical or chemically similar amino acids) at 52% of the paired positions (Fig. 7A). This sequence similarity is undoubtedly responsible for the reported serological cross-reactivity between BWYV and BYDV(PAV) virions (31).

The BYDV 22 KD cistron, like the 22.5 KD cistron of BWYV, contains a second, shorter open reading frame in another phase (Fig. 6). The 17 KD polypeptide encoded by this ORF in BYDV may correspond to the VPg (2). There is considerable sequence homology (33%) between this species and its 19.5 KD analogue in BWYV (Fig. 7B). Finally, both genomes possess long open reading frames in phase with and separated from the adjacent 22-22.5 KD ORFs only by an amber termination codon (Fig. 6). Marked sequence homology exists between these two ORFs, primarily in their N-terminal regions which match in 43% of the paired positions (Fig. 7C). It is perhaps noteworthy that the region of homology

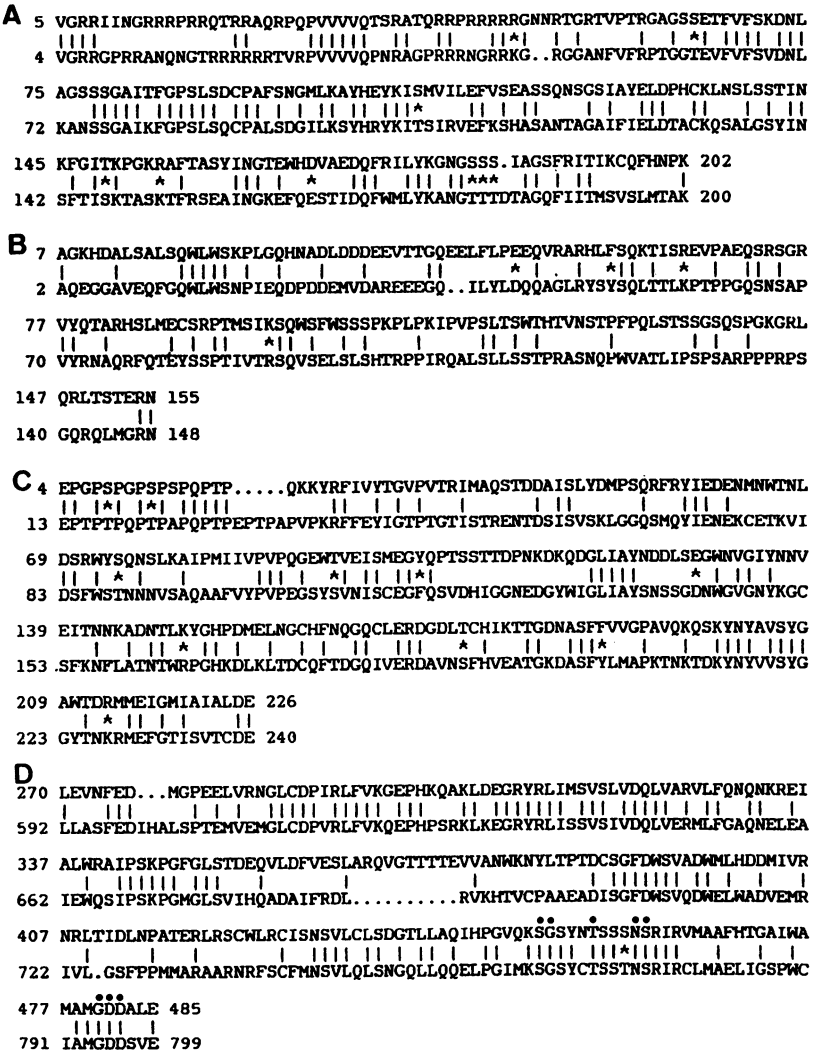


Figure 7. Amino acid sequence comparisons of BWYV(FL1) ORFs with ORFs of other plant viruses using BESTFIT (22). Vertical lines indicate identical amino acids, asterisks chemically similar amino acids (T and S, D and E, R and K, Y and F) and dots positions where gaps have been used to optimize the fit. Numbering refers to amino acid position in each ORF. (A) Alignment of the BWYV 22.5 KD polypeptide (ORF 4) with the BYDV(PAV) (1) 22 KD polypeptide; (B) alignment of the BWYV 19.5 KD polypeptide (ORF 5) with the BYDV 17 KD polypeptide; (C) alignment of the N-terminal region of BWYV ORF 6 with the BYDV 50 KD polypeptide; (D) alignment of BWYV ORF 3 with the putative RNA polymerase (ORF 2) of southern bean mosaic virus (26). Filled circles above the sequence indicate the conserved "core polymerase" sequence motif (33).

corresponds approximately to the portion of ORF 6 which is most highly conserved between the lettuce and sugarbeet isolates of BWYV.

In BYDV the sequence flanking the amber termination codon separating the 22 KD and 50KD ORFs has homology with that surrounding stop codons of several other viruses which are known to undergo supression during translation (1). The context of the analogous amber codon in BWYV (CCAAAUAGGUAGAC) is identical to that of BYDV and the corresponding ca. 50 KD ORFs of both luteoviruses may be expressed by translational readthrough (see below). Near the beginning of the BYDV 50 KD ORF there is an unusual 32 amino acid sequence in which every other residue is a proline (1). A similar but shorter sequence, EPGPSPGPSPSPQPTP, occurs near the beginning of the BWYV 51.5 KD ORF.

The BWYV 67 KD ORF and the BYDV 60 KD ORF both contain the amino acid sequence motif GXXXTXXXNX$_{18\text{-}50}$GDD (X is any amino acid) which occurs in non-structural proteins of virtually all plus-strand RNA viruses and is thought to represent a core sequence of the viral RNA-dependent RNA polymerase (32,33). We suggest a similar role for the BWYV ORF 3 sequence. When the 67 KD amino acid sequence of BWYV ORF 3 was compared to putative RNA polymerases of other plant RNA viruses the closest resemblance was found with that of southern bean mosaic virus. The region of greatest homology includes the "GDD" motif and sequence extending about 170 amino acids upstream. The optimal alignment matches the two sequences at 51% of the paired positions (Fig. 7D). In BYDV, on the other hand, the 60 KD ORF displays homology not with the southern bean mosaic virus polymerase but with that of carnation mottle virus (1).

No amino acid sequence homology of note was detected between the BWYV 29 KD and 66 KD ORFs and the various ORFs of BYDV or of any other published plant virus sequence. However, the sequence of a third luteovirus, PLRV, has recently been completed. It has closely similar genome organization to BWYV and significant amino acid sequence

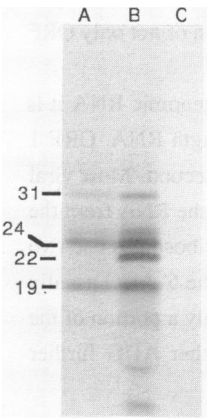


Figure 8. Cell-free translation products of pBW(3313-4321)BS runoff transcript fractionated by electrophoresis through a 12.5% polyacrylamide gel. (A) Translation products immunoprecipitated by BWYV antiserum; (B) translation products before immunoprecipitation; (C) translation products precipitated by nonimmune serum. Apparent M_r (KD) of major translation products are indicated to the left.

homology exists between all the analogous open reading frames in the two viruses except for BWYV ORF 1 and its counterpart in PLRV (M. A. Mayo, personal communication).

Possible mechanisms of genome expression. Based upon its size and the observed sequence homology with the BYDV coat protein the 22.5 KD polypeptide encoded by ORF 4 is the obvious candidate to be the BWYV coat protein. In order to test this prediction a Dra I-EcoR I restriction fragment containing ORF 4 was inserted into the transcription vector BS(-) so that transcription of plus-strand RNA was under control of the bacteriophage T3 RNA polymerase promoter. In addition to the ORF 4 sequence the cDNA insert contains 170 nt of the intergenic region preceding ORF 4, the entire ORF 5 sequence and the first 229 nt of ORF 6 (Fig. 6). The recombinant plasmid, pBW(3313-4321)BS, was linearized with EcoR I and a plus-strand runoff transcript was synthesized with T3 RNA polymerase. The transcript was in turn used to direct synthesis of ^{35}S-methionine-labelled polypeptides in rabbit reticulocyte lysate. Four abundant translation products were detected by SDS-polyacrylamide gel electrophoresis with apparent M_r of 31 KD, 24 KD, 22 KD and 19 KD (Fig. 8B). All but the 22 KD species were precipitated specifically by BWYV-antiserum (Fig. 8A). The 24 KD polypeptide probably represents the primary ORF 4 translation product; it comigrated with authentic BWYV coat protein obtained by dissociation of virions in polyacrylamide gel loading buffer. The 31 KD species is about the size expected for a fusion polypeptide containing ORF 4 and the cloned portion of ORF 6. Its existence supports the hypothesis (1) that the coat protein amber termination codon can undergo translational readthrough, at least *in vitro*. The origin of the 19 KD polypeptide is not known but it most probably represents a product of premature translation termination of the coat protein cistron. Finally, there is one species, with apparent M_r 22 KD, which is not recognized by the BWYV antiserum (Fig. 8). We suggest that this polypeptide is the ORF 5 translation product. As a rule internal cistrons of RNA viruses are expressed by means of 3'terminally nested subgenomic RNAs and there is evidence for such an RNA beginning near the coat protein cistron of BYDV (34). The cell-free translation experiment with synthetic transcript described above suggests that the existence of a comparable RNA in BWYV-infected plants could account for the expression of not only ORF 4 but ORFs 5 and 6 as well.

Just as ORFs 4 and 5 may both be translated from a single subgenomic RNA it is conceivable that BWYV ORFs 1 and 2 are both expressed from full-length RNA. ORF 1 begins at the first AUG codon in the RNA sequence and ORF 2 at the second. Most viral RNAs are expressed in a monocistronic fashion with ribosomes scanning the RNA from the 5'end until an AUG is encountered. Situations exist, however, in which ribosomes use two initiation sites on the same RNA (reviewed in 35). In such circumstances the 5' AUG usually has suboptimal codon context (as is the case for BWYV ORF 1) so that only a portion of the scanning ribosomes initiate at that site while the rest continue to another AUG further downstream.

Expression of BWYV ORF 3 may involve a different sort of mechanism. As noted above, AUG(2168), the first potential initiation codon in ORF 3, is separated from the beginning of the open reading frame proper by 216 codons (Fig. 4). Much of the reading frame preceding AUG(2168) overlaps the end of ORF 2 and we suggest that a -1 translational frameshift (see 36 for review) somewhere in the region of overlap permits translation of the ORF 3 sequence beyond this point as an ORF 2-3 frameshift fusion protein of approximately 115 KD. A similar mechanism has been proposed for expression of the analogous BYDV 60 KD ORF although in this case the overlap with the preceding cistron is only 13 residues in length (1).

Luteovirus taxonomy. With respect to luteovirus classification the comparison between the BWYV and BYDV(PAV) sequences is of considerable interest. We have seen that the two sequences contain regions of sequence homology as well as elements of dissimilarity. Thus BWYV ORFs 4, 5 and 6 display a clear evolutionary relationship with the corresponding cluster of cistrons in BYDV(PAV) both at the amino acid sequence level and in overall organization. The putative polymerase regions, on the other hand, are distinct from one another and appear to share common ancestry with other viruses, southern bean mosaic virus in the case of BWYV and carnation mottle virus for BYDV. It has been suggested (37) that RNA viruses may evolve by exchanging functional sequence modules with one another, a mechanism which could explain the juxtaposition of similar and dissimilar elements in BWYV and BYDV.

If the frameshift hypothesis for expression of the BWYV and BYDV polymerase ORFs is borne out by experiment this will represent a second parallel between the two viruses. In contrast, however, each virus posesses a sixth ORF (at the 5'extremity in BWYV and at the 3'extremity of BYDV) for which no counterpart exists in the other virus. Based on this difference and on the apparently different origin of the polymerase ORFs, we suggest that BWYV and BYDV(PAV) may be considered as distinct subtypes of the luteovirus group. As noted above, PLRV appears to have genetic organization similar to BWYV. Whether other members of the group will conform to one or the other of these two subtypes is a question that must await additional sequence analysis.

ACKNOWLEDGEMENTS The authors would like to thank M.A. Mayo for the exchange of unpublished sequence data, W.A. Miller for preprints and C. D'Arcy for information about Extractase. EMBL DATA Bank accession numbers for the BWYV(FL1) and BWYV(GB1) sequences are X13063 and X13062, respectively.

REFERENCES

1. Miller, W.A., Waterhouse, P.M. and Gerlach, W.L. (1988) Nucleic Acids Res. 16, 6097-6111.
2. Miller, W.A., Waterhouse, P.M., Kortt, A.A. and Gerlach, W.L. (1988) Virology 169, 306-309.
3. Matthews, R.E.F. (1982) Intervirology 17, 140-141.

4. Rochow, W.F. and Duffus, J.E. (1981) Luteoviruses and yellows diseases. In "Handbook of Plant Virus Infections and Comparative Diagnosis", E. Kurstak (ed.), pp. 147-170, Elsevier/North Holland Biomedical Press, Amsterdam.
5. Casper, R. (1988) Luteoviruses. In "The Plant Viruses; Polyhedral Virions with Monopartite RNA Genomes. Vol. 3", R. Koenig (ed.), pp. 235-258, Plenum, New York.
6. Mayo, M.A., Barker, H., Robinson, D.J., Tamada, T. and Harrison, B.D. (1982) J. gen. Virol. 59, 163-167.
7. Murphy, J.F. Clark, J.M. and D'Arcy, C.J. (1987) Phytopathology 77, 1705.
8. Govier, D.A. (1985) Ann. Appl. Biol. 107, 439-447.
9. D'Arcy, C.J., Hewings, A.D., Burnett, P.A. and Jedlinski, H. (1983) Phytopathology 73, 755-759.
10. Briest, E. and Kegler, H. (1987) Arch. Phytopathol. Pflanzenschutz 23, 3-11.
11. Kuhne, T., Proeseler, G., Richter, J., Stanarius, A. and Proll, E. (1985) Arch. Phytopathol. Pflanzenschutz 21, 3-12.
12. Heidecker, G. and Messing, J. (1983) Nucleic Acids Res.11, 4891-4906.
13. Bouzoubaa, S., Guilley, H., Jonard, G., Richards, K. and Putz, C. (1985) J. gen. Virol. 66, 1553-1564.
14. Maizels, N. (1976) Cell 9, 431-438.
15. Gubler, V. and Hofmann, B.J. (1983) Gene 23, 263-269.
16. Bouzoubaa, S., Quillet, L., Guilley, H., Jonard, G. and Richards, K. (1987) J. gen. Virol. 68, 615-626.
17. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-565.
18. Sanger, F., Miklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
19. Hattori, M. and Sakaki, Y. (1986) Anal. Biochem. 152, 232-238.
20. Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767-4771.
21. Deborde, D.C., Naeve, C.W., Herlocher, M.L. and Maassab, H.F. (1986) Anal. Biochem. 157, 275-282.
22. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395.
23. Jupin, I., Quillet, L., Ziegler-Graff, V., Guilley, H., Richards, K. and Jonard, G. (1988) J. gen. Virol., In press.
24. Ziegler, V., Richards, K., Guilley, H., Jonard, G. and Putz, C. (1985) J. gen. Virol. 66, 2079-2087.
25. Domier, L.L., Franklin, K.M., Shahabuddin, M. Hellmann, G.M., Overmeyer, J.H., Hiremath, S.T., Siaw, M.F.E., Lomonossoff, G.P., Shaw, J.G. and Rhoads, R.E (1986) Nucl. Acids Res.14, 5417-5430.
26. Wu, S., Rinehart, C.A. and Kaesberg, P. (1987) Virology 161, 73-80.
27. Nomoto, A., Omata, T., Toyoda, H., Kuge, S., Horie, H., Kataoka, Y., Genba, Y., Nakano, Y. and Imura, N. (1982) Proc. Natl. Acad. Sci. USA 79, 5793-5797.
28. Lomonossoff, G.P. and Shanks, M. (1983) EMBO J. 2, 2253-2258.
29. Greif, C., Hemmer, O. and Fritsch, C. (1988) J. gen. Virol. 69, 1517-1529.
30. Hewings, A.D. and D'Arcy, C.J. (1983) Phytopathology 73, 789-790.
31. Duffus, J.E. and Rochow, W.F. (1978) Phytopathology 68, 45-49.
32. Kamer, G. and Argos, P. (1984) Nucleic Acids Res. 12, 7269-7282.
33. Goldbach, R.W. (1986) Ann. Rev. Phytopath. 24, 289-310.
34. Gerlach, W.L., Miller, W.A. and Waterhouse, P.M. (1987) Barley Yellow Dwarf Virus Newsletter 1, 17-19.
35. Kozak, M. (1986) Adv. Virus Res. 31, 229-292.
36. Craigen, W.J. and Caskey, C.T. (1987) Cell 50, 1-2.
37. Zimmern, D. (1987) Evolution of RNA Viruses. In "RNA Genetics", J. Holland, E. Domingo and P. Ahlquist (ed.) CRC Press, Boca Raton, Fla.