VIROLOGY 185, 354–364 (1991)

Characterization of the Genome of Rice Tungro Bacilliform Virus: Comparison with *Commelina* Yellow Mottle Virus and Caulimoviruses

RONGDA QU, MAITRAYEE BHATTACHARYYA, GARY S. LACO, ALEXANDRE DE KOCHKO, B. L. SUBBA RAO, MARIA B. KANIEWSKA, J. SCOTT ELMER,* DEAN E. ROCHESTER, CHRISTINE E. SMITH,* AND ROGER N. BEACHY[1]

*Department of Biology, Washington University, St. Louis, Missouri 63130; and *Monsanto Company, Chesterfield, Missouri 63198*

Received February 15, 1991; accepted August 8, 1991

Rice tungro disease is caused by an infection of two different viruses, rice tungro spherical virus (a (+) sense RNA virus) and rice tungro bacilliform virus (RTBV) with a genome of circular double-stranded DNA. The genome of an RTBV isolate from the Philippines was cloned, sequenced, and found to be 8000 bp in length. It contains four open reading frames (ORFs) on a single strand, with ORF 1 having an internal termination codon (TAA). The 5′ and 3′ ends of a polyadenylated viral RNA transcript, of genome length, were mapped by primer extension and cDNA sequence analysis, respectively. The transcript is terminally redundant by 265–268 nucleotides. Purified virus particles contain two major proteins with molecular masses of 37 and 33 kDa, although only the 37-kDa protein was detected in the infected rice tissues. The N-terminal amino acid sequence of the 33-kDa protein was determined and its coding region was identified on the RTBV genome. The identity of the coat protein gene was further confirmed by expressing a region of the genome in *Escherichia coli,* the products of which reacted with anti-RTBV antibody. The unusually long ORF 3 of RTBV is predicted to encode a polyprotein of 194.1 kDa that includes: the coat protein(s), viral proteinase, reverse transcriptase, and ribonuclease H. The sections of the polyprotein show varying degrees of similarity to the counterparts of *Commelina* yellow mottle virus (a member of the proposed badnavirus group) and caulimoviruses. The functions of the other three ORFs are unknown.

INTRODUCTION

Rice tungro bacilliform virus (RTBV) contains a circular double-stranded DNA genome of approximately 8 kb (Jones *et al.,* 1991). Coinfection of rice plants by RTBV and rice tungro spherical virus (RTSV) produces a disease syndrome, known as rice tungro, which causes stunting along with yellow or orange discoloration of rice plants. This disease can result in substantial yield loss (Hibino *et al.,* 1978, 1979; Saito *et al.,* 1986).

RTBV is efficiently transmitted to members of the genus *Oryza* (rice) by the green leafhopper (*Nephotettix virescens*) when plants are coinfected with RTSV, presumably because RTSV provides a helper function (Hibino *et al.,* 1978; Hibino, 1983). Electron microscopy of RTBV-infected rice tissues showed that the RTBV particles were confined to the phloem. They were usually randomly dispersed in the cytoplasm of companion cells, or in sieve elements (Saito *et al.,* 1975). The virions of RTBV are approximately 35 nm in diameter with lengths from 150 to 350 nm (Hibino *et al.,* 1978) and have a buoyant density of 1.312 g/cm^3 (Saito *et al.,* 1986). The virus contains two major proteins with MW's of 37 and 33 kDa (Jones *et al.,* 1991).

In this paper we report the complete nucleotide sequence of a full-length clone of the RTBV genome and compare the sequence to that of *Commelina* yellow mottle virus (CoYMV) (Medberry *et al.,* 1990), cauliflower mosaic virus (CaMV) (Franck *et al.,* 1980), and other caulimoviruses at both the nucleotide and amino acid levels. We conclude that RTBV is likely a plant pararetrovirus even though it is substantially different from the caulimoviruses.

MATERIALS AND METHODS

Strategies for cloning and sequencing of the RTBV genome

Purified RTBV was kindly provided by Dr. H. Hibino of the International Rice Research Institute (Los Baños, Philippines). Viral DNA extracted from RTBV virions as described in Jones *et al.* (1991) was subsequently cloned into pUC119 (Vieira and Messing, 1987), or pBluescript (Stratagene, La Jolla, CA), by using the

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. M65026.

[1] To whom correspondence and reprint requests should be addressed at Division of Plant Biology, MRC-7, Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037.

0042-6822/91 $3.00

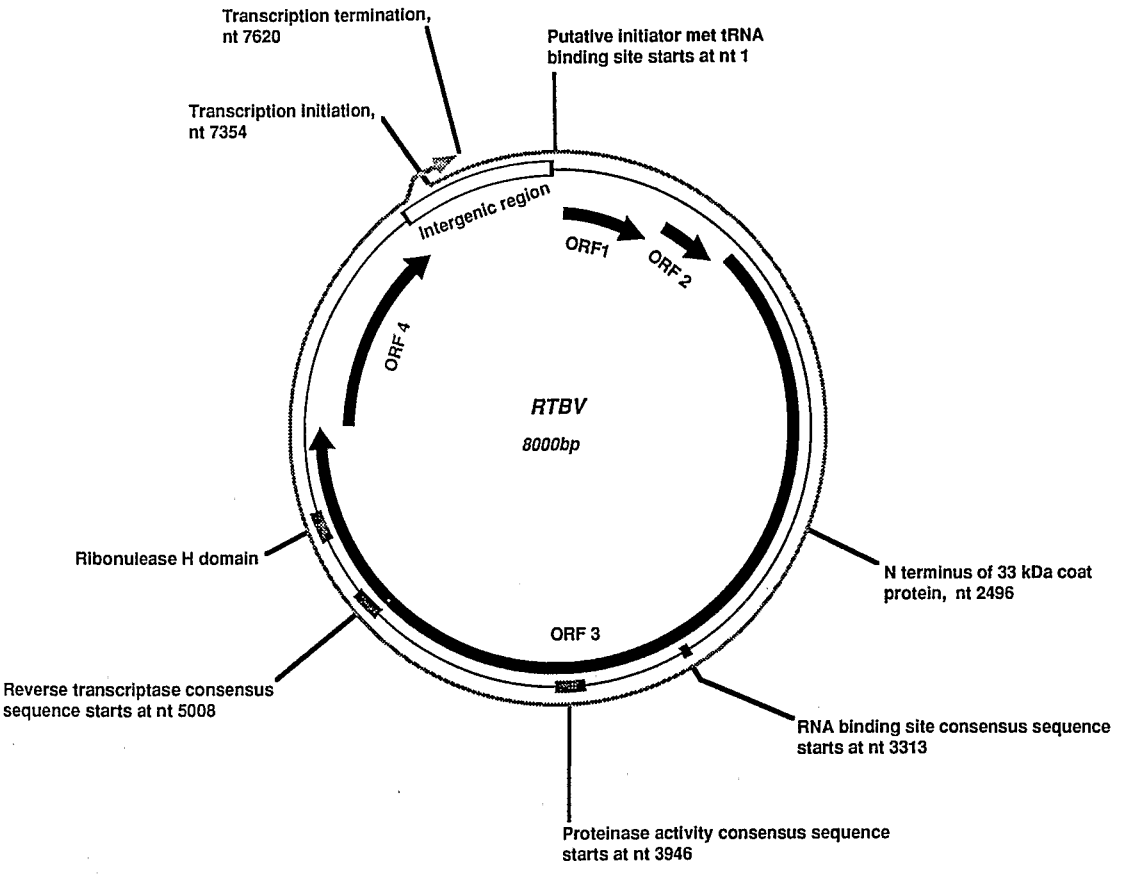


FIG. 1. Genome organization of RTBV based upon the nucleotide sequence of cloned viral DNA and the viral RNA analysis. The thin black line depicts the RTBV genome with the major intergenic region framed. The black arrows represent the open reading frames. The gray line (with arrow) indicates the major transcript of RTBV. The recognized locations of biologically important sequences (or consensus sequences) are presented as small blocks.

unique *Bam*HI site in the genome. A restriction map was made with several enzymes. DNA fragments produced by digestion with *Bam*HI, *Eco*RI, and *Pst*I, were cloned into pUC118 or pUC119. Approximately 2.5 kb of the genome (from the *Bam*HI site at nucleotide 7186 to the *Pst*I site at nt 787, and from the *Eco*RI site at nt 4867 to the *Cla*I site at nt 5918) was sequenced by making a nested, overlapping set of Exonuclease III deletions following the manufacturer's protocol (Promega, Madison, WI). The remainder of the genome was sequenced using discrete restriction fragments subcloned into pUC118 or pUC119. In some cases, oligonucleotide primers were used to obtain sequences from internal regions of DNA subclones. Both strands of the full-length clone were completely sequenced. In addition, the sequences of each of the restriction sites used for subcloning were confirmed by sequencing across the sites.

The dideoxynucleotide sequencing method (Sanger *et al.*, 1977) was used with [^{35}S]deoxyadenosine-5′(α-thio)triphosphate (New England Nuclear, Boston, MA) as the labeled nucleotide (Beggin *et al.*, 1983). Double-stranded plasmid DNA was prepared by the boiling miniprep procedure followed by PEG precipitation (Sambrook *et al.*, 1989). The sequencing reagents and enzymes (Sequenase version 2.0 or T7 DNA polymerase) were purchased from U.S. Biochemical Corp. (Cleveland, OH) and Pharmacia (Piscataway, NJ), respectively, and used according to manufacturers' suggestions. Products of the reactions were run on 6% polyacrylamide gels containing 7.5 *M* urea. The gels were fixed, dried, and exposed to Kodak X-OMAT film for 12–72 hr.

To confirm an internal termination codon (nt 66–68) within ORF 1, a PCR product was obtained from purified viral DNA using the oligonucleotide primers 5′TGGTATCAGAGCGATGT3′ (nt 1–17, plus strand) and 5′CCCCAGACTGCTAAGTC3′ (nt 132–148, minus strand) and the GeneAmp kit (Perkin Elmer Cetus, Norwalk, CT). The 148-bp product was directly sequenced by the method described above with the first oligomer as the primer.

Computer analysis of sequencing data

The GCG (Genetics Computer Group) sequence analysis software package (Version 5.1) of the University of Wisconsin (Devereux *et al.*, 1984) was used to analyze DNA sequence, as well as the derived amino acid sequence data.

N-terminal microsequencing of RTBV coat protein(s)

RTBV coat proteins were purified by SDS/polyacrylamide gel electrophoresis (Laemmli, 1970) in preparation for microsequencing as modified by Hunkapiller *et al.* (1983). A virion suspension, containing 5 to 10 μg of viral coat proteins, was mixed with 0.5 vol of 3× sample buffer (Laemmli, 1970) and incubated at 60° for 10 min. The separating gel was polymerized overnight, and prerun for 30 min at 100V in separating gel buffer. Sodium thioglycolate (Sigma, St. Louis, MO) was added to the stacking gel as well as to the cathode buffer reservoir to a final concentration of 0.1 m*M*, before electrophoresis.

Following electrophoresis, the protein was transferred to Immobilon-P PVDF membrane (Millipore Corp., Bedford, MA) as described by Matsudaira (1987). The membrane was stained with 0.05% Serva Blue R (Serva, New York, NY) in 50% methanol for 20 min and destained in a solution of 50% methanol and 10% acetic acid for 30 min. The membrane was then air dried and stored at −20° prior to microsequencing of the stained proteins with a Model 470A gas phase sequencer (Applied Biosystems, Inc., Foster City, CA).

Protein blots and antibody recognition

One gram of rice tissue was frozen in liquid nitrogen and powdered in a mortar with pestle. Two milliliters of extraction buffer (0.1 *M* Tris–HCl, pH 6.8; 10 m*M* EDTA, and 5 m*M* DTT) was added to the ground tissue. The homogenate was then centrifuged at 37,000 *g* for 30 min with the resulting supernatant being used for protein analysis.

For the expression of RTBV coat protein(s) in *Escherichia coli* (strain DH5αF'), a *Dra*I–*Pst*I fragment (nt 1905–3888) of the RTBV genome was inserted into the *Sma*I and *Pst*I sites of pUC118. A colony containing recombinant plasmid was grown overnight in 3 ml of LB medium containing 100 μg/ml of ampicillin (Sambrook *et al.*, 1989). The bacteria were collected by centrifuging 1:5 ml of the culture. The cell pellet was then resuspended in 100 μl of distilled H_2O.

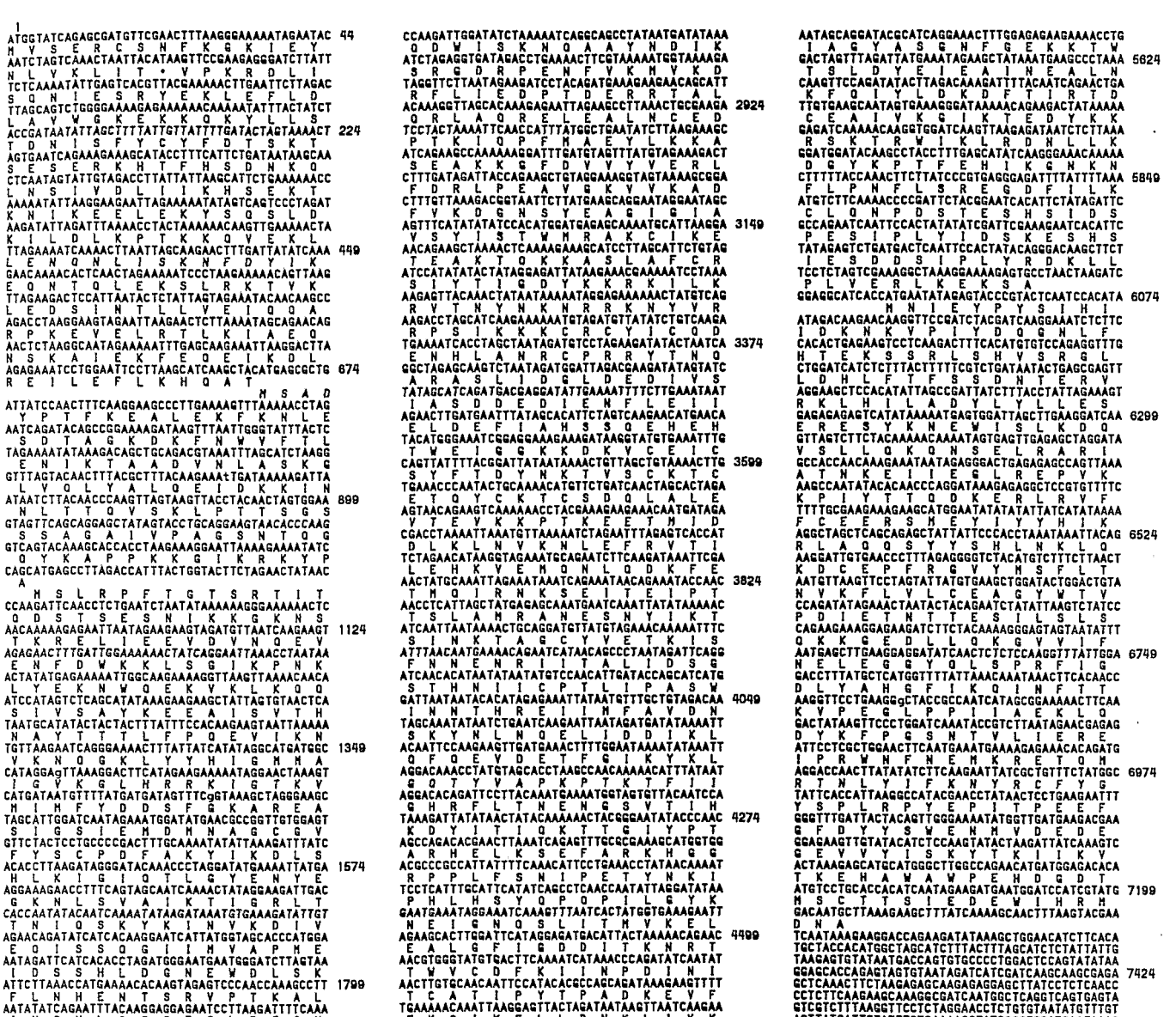
Two volumes of the supernatant from rice tissue, or an *E. coli* suspension, was mixed with 1 vol of 3× sample buffer and boiled for 5 min. The samples were next subjected to electrophoresis in 12.5% polyacrylamide gels containing 0.1% (w/v) of SDS (Laemmli, 1970). The proteins were then blotted onto nitrocellulose membrane (Towbin *et al.*, 1979) and reacted with antibody prepared in rabbits against purified RTBV virions using the ProtoBlot western blot alkaline phosphatase system (Promega).

Analyses of viral RNAs

Tissue for the isolation of viral RNAs was obtained frozen on dry ice from IRRI. The tissue was weighed before grinding in liquid nitrogen with a mortar and pestle. Total RNA was isolated from the pulverized infected tissue using a guanidinium thiocyanate buffer as described by Sambrook *et al.* (1989). Strandness of the transcript was determined by dot-blot hybridization using strand specific probes. Single strand probes were generated from a subclone of the genome (nt 2134–3888) by extension reactions using the Klenow fragment of DNA polymerase and M13/pUC primers. [α^{32}P]-dATP was added as the labeled nucleotide (Sambrook *et al.*, 1989).

Primer extension (Ghosh *et al.*, 1978) was used to determine the 5′ end of RTBV RNA isolated from infected tissue. Two primers complementary to the coding strand of RTBV were synthesized. The first primer is complementary to nt 7429–7448 of the virus genome. In addition, three nucleotides were added at the 5′ end to create a *Bgl*II site. The primer's sequence is 5′AGATCTTGCTCTCTTAGAAGTTT3′. The second primer has the sequence 5′GAGCCATTGATCGCCTTTGCTTC3′ which is complementary to nt 7479–7501 of the virus genome. One nmole of each oligonucleotide primer was end-labeled using 10 μCi of [γ^{32}P]-ATP (New England Nuclear) as described (Maniatis *et al.*, 1982). Following end labeling, the primer was separated from unincorporated [γ^{32}P]-ATP by gel filtration using a 10-ml bed volume Sephadex G-50 column (Sigma). The fractions containing the end-labeled primer, as determined by paper chromatography (Wu

FIG. 2. Nucleotide sequence of RTBV genomic DNA (plus strand with the numbering beginning as described in the text) and the derived amino acid sequences of the 4 ORFs. In order to make the coding region of the ORF 1 intact, the nucleotide sequence is presented to start at nt 8000 (or −1) and to end at nt 7999. The predicted amino acid sequence which matches the N-terminal sequence of the 33-kDa coat protein is underlined.



et al., 1974), were pooled and precipitated with ethanol.

Five to ten micrograms of total RNA extracted from rice tissues infected with both RTBV and RTSV, or RTBV predominantly, was resuspended in 6 μl of distilled H_2O. They were then heated to 90° for 5 min and cooled on ice. Fifty fmoles of end-labeled primer, 1 μl of 20× reverse transcriptase buffer (400 m*M* Tris–HCl, pH 8.3; 150 m*M* $MgCl_2$, 150 m*M* KCl, 40 m*M* DTT), and 2 units of RNasin (Promega) were then added. The reaction mix was next incubated at 42° for 30 min. Three microliters of an extension mixture, containing 1 m*M* of dATP, dCTP, dGTP, and dTTP, and 20 units of AMV reverse transcriptase (Seikagaku, Osaka, Japan), was added to the annealed reaction mixture. It was then incubated at 48° for 40 min. A second 3-μl aliquot of the extension mixture was subsequently added, with the reaction continuing at 48° for an additional 40 min. The reaction was stopped by adding sequencing gel-loading buffer (Sambrook *et al.*, 1989). One half of the reaction was loaded on a sequencing gel next to a sequencing reaction of the cloned full-length RTBV DNA using the same batch of end-labeled primers.

The 3′ end of the RTBV transcript was determined in the following manner. One microgram of total RNA, extracted from RTBV infected rice tissues, was used as a template for reverse transcription. An oligo dT primer linked to a *Bam*HI nucleotide adaptor was used to prime the first strand cDNA. The synthesis of cDNA was completed as described above, after which the reaction was diluted to 100 μl. One microliter of the diluted reaction product along with the oligo dT primer and a second primer (5′CCAGAACATGATGGAGAC3′, nt 7134–7151) were used for a PCR reaction with the GeneAmp kit (Perkin Elmer Cetus). The product was then subjected to restriction digestion with *Hin*dIII (nt 7214) and *Bam*HI (site created by first primer). The resulting DNA fragment was cloned into pUC119 and subsequently subjected to DNA sequence analysis.

RESULTS

DNA sequence and genome organization

Jones *et al.* (1991) reported that the genome of RTBV is a circular double-stranded DNA of approximately 8 kb. The RTBV genome was linearized by *Bam*HI, which cuts at a single site in the genome, and cloned into pUC119 and pBluescript. The resulting full-length clones (pBSR45, 63, 63A, 65, 95, and 105) showed identical restriction digest patterns (data not shown). The clone pBSR63A was chosen to derive a restriction map, and the genomic sequence, of RTBV.

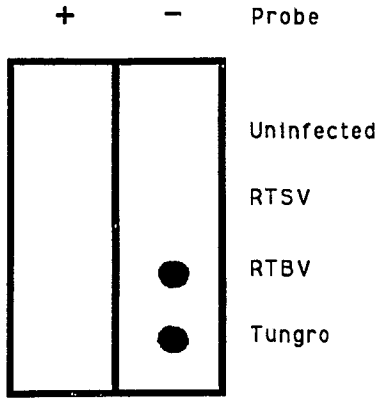


FIG. 3. Dot blot of total RNA extracted from RTBV-infected and uninfected rice tissues showing strand specificity. Ten micrograms of RNA extracted from each of the following samples were used for the blot: uninfected tissues (Uninfected), and tissues infected with RTSV only (RTSV), RTBV predominantly (RTBV), and both viruses (Tungro). The probes were either from the plus (+) or the minus (−) strand of a restriction fragment of RTBV DNA (nt 2134–3888).

The complete nucleotide sequence of the insert in pBSR63A was determined as described under Materials and Methods. The sequence, given in Fig. 2, consists of 8000 bp and has a G + C content of 33.7%. Computer analyses revealed that only one strand (plus strand) has extensive coding capacity (data not shown). The plus strand of the genome contains four open reading frames (ORFs) greater than 300 nucleotides and a major intergenic region (Figs. 1 & 2, and Table 1), whereas the longest ORF in the minus strand encodes only 82 amino acids. Identity of the coding strand of the RTBV genome was confirmed by dot blot hybridization (Fig. 3).

Determination of the RTBV transcript termini

Preliminary northern blot analysis of polyadenylated RNA isolated from infected rice tissues showed that RTBV had a major transcript of approximately 8 kb (data not shown). Primer extension experiments utilizing two different oligomers as primers were carried out to determine the location of the initial 5′ nucleotide in the RNA isolated from these tissues. Each primer produced two products that began in the major intergenic region of the genome. The major initiation site was at nucleotide 7354, and the minor at 7356 (Fig. 4A). Two TATA-like sequences were located upstream of the transcription initiation sites: one (TATAAT) was at nt 7343 while the other (TATAAA) was at nt 7267.

A PCR reaction was designed to determine the 3′ end of the RTBV transcript such that the product would indicate which of the two potential polyadenylation signals (AATAAA at nt 7247 and nt 7598) was really used

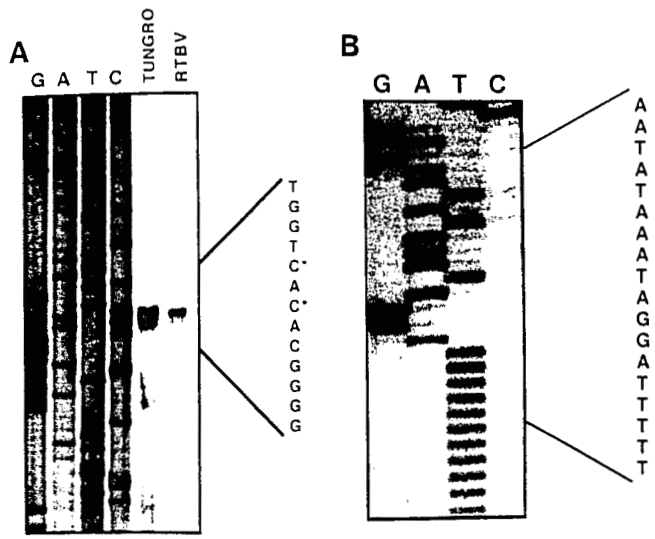


FIG. 4. Determination of initiation (A) and termination (B) sites of RTBV major transcript. (A) Primer extension reaction products were subjected to electrophoresis adjacent to sequencing reactions of RTBV DNA primed with the same primer. The sequencing gel is presented such that the complement to the transcript can be read directly in a 5′ to 3′ direction by reading up the gel. The two right lanes show the products of the primer extension reaction using total RNA extracted from rice tissues infected by both RTBV and RTSV (TUNGRO), or by RTBV predominantly (RTBV). The four left lanes show the sequencing reaction of RTBV DNA using the same primer. The nucleotide sequence of the minus strand of RTBV DNA containing the transcription initiation sites is indicated on the right side of the figure with asterisks (*) showing the major and minor transcription initiation sites. (B) The cDNA's PCR product was cloned and sequenced to determine the 3′ end of the RTBV transcript. The sequencing reaction products were subjected to electrophoresis. The sequencing gel is presented such that the cDNA can be read directly in a 5′ to 3′ direction by reading up the gel. The cDNA sequence, including oligo dT, is indicated on the right side of the figure.

for polyadenylation of the RTBV transcript. A PCR product of approximately 600 bp was obtained. Sequencing data from the cloned PCR product showed that the transcript terminated at nt 7620 or 7621 (nt 7621 is an A) (Fig. 4B). The presumed polyadenylation signal AATAAA starts at nucleotide 7598, 22 or 23 nucleotides upstream of the poly A tract. Thus, as in the case of CoYMV and caulimoviruses, the RTBV transcript is longer than genome length, and is terminally redundant by 265 to 268 nucleotides.

Putative primer binding site

A computer-assisted search identified a region of 18 nucleotides, TGGTATCAGAGCGATGTT, of which 16 are complementary (underlined) to the 3′ terminus of plant cytoplasmic $tRNA_i^{met}$ (Sprinzl *et al.*, 1987). This sequence may serve as a primer binding site for reverse transcription from the full-length transcript as proposed in the cases of caulimoviruses and CoYMV (Pfeiffer and Hohn, 1983; Guilley *et al.*, 1983; Hohn *et al.*, 1985; Medberry *et al.*, 1990). We have numbered the RTBV genome sequence to begin at the 5′ end of the putative primer binding site, as done in earlier studies with caulimoviruses and CoYMV (Hull *et al.*, 1986; Richins *et al.*, 1987; Hasagawa *et al.*, 1989; Medberry *et al.*, 1990).

Analysis of coding regions

The open reading frames predicted from the sequence of the RTBV genome are presented in Table 1. ORF 1 begins at nt 8000, even though it includes a TAA termination codon at amino acid position 23 and continues uninterrupted for an additional 198 amino acids without an initiation codon downstream from the stop codon. The result was confirmed by sequencing this region in other independent full-length clones of the genome (pBSR45, 65, 95, and 105), as well as a PCR product derived from the viral DNA (see Materials and Methods). The predicted 222 amino acid long product does not share similarities with ORF 1 of CoYMV, or with ORF I and VII of CaMV. Furthermore, a search of the NBRF Protein database did not find similarities with other protein sequences.

ORF 2 starts at nt 665 and ends at nt 997, encoding a protein of 110 amino acid residues with a molecular weight of 11.9 kDa. No similarity was found with ORF 2 of CoYMV, or with ORF I and II of CaMV. Again, a computer search of the database did not find similarities between ORF 2 and other proteins.

Between ORF 1 and ORF 2, and between ORF 2 and ORF 3, the sequence ATGA was found (Fig. 2). Identical sequence was noted in the genome of carnation etched ring virus (CERV) by Hull *et al.* (1986). Thus the termination codon of the previous ORF shares two nucleotides (TG) with the initiation codon of the subsequent ORF.

The long ORF 3 spans from nucleotide 994 to 6020 and encodes a putative 194.1-kDa polyprotein of 1675 amino acid residues. Located within the predicted ORF 3 product is the amino acid sequence, TVNPIEASKDRNFERIGSSYKKNFY (underlined in Fig. 2), that was found at the amino terminus of the 33-kDa coat protein (see below). The predicted ORF 3 product was also found to have the consensus sequences of retroviral proteinase, reverse transcriptase, and ribonuclease H (see below). For convenience, the polyprotein encoded by ORF 3 may be divided into three major sections which show various degrees of similarities to counterparts of CoYMV, and the products of ORF IV and ORF V of CaMV. These sections are designated as

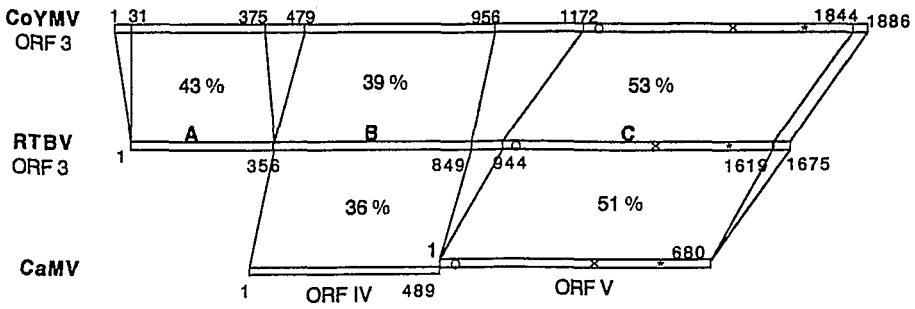


FIG. 5. Comparison of RTBV ORF 3 sections with sequences of CoYMV and CaMV. The three major sections of RTBV ORF 3 are labeled A, B, and C and are delineated by the amino acid residue numbers. The similarities of amino acid sequence between corresponding sections are shown by percentages. The locations of consensus sequences are indicated as follows: O, proteinase; X, reverse transcriptase; and *, ribonuclease H.

A, B, and C, respectively (Fig. 5). The boundaries of the sections described below are estimates based on the computer alignment of CaMV ORF products with the RTBV ORF 3 polyprotein.

Section A begins at aa 1 of the RTBV ORF 3 polyprotein and ends at aa 355. The predicted polypeptide shows 43% similarity to its counterpart in CoYMV (aa 31 through 375 of ORF 3). The function of this section is unknown.

Section B of ORF 3 encodes 494 amino acid residues (aa 356 through 849) including the coat protein(s) coding sequence. As described above, the amino acid sequence from the N-terminus of the 33-kDa virus protein was determined. That sequence was found at aa 502 through 526 of ORF 3. The most highly conserved region within section B among RTBV, CoYMV, and CaMV is aa 776 through 789 which matches the consensus sequence $CX_2CX_4HX_4C$ (Fig. 6) of the RNA binding site for retroviruses (Covey, 1986; Fuetterer and Hohn, 1987).

Attempts to obtain the amino acid sequence at the N-terminus of the 37-kDa protein failed. This is presumably due to a modification at the N-terminus of the protein. A Western blot of proteins extracted from infected rice tissues showed that a protein comigrating with the 37-kDa coat protein was the major protein recognized by the anti-RTBV antibody; no 33-kDa protein was detectable (Fig. 7). This result implies that the 37-kDa protein is the major coat protein of RTBV, while the 33-kDa coat protein is a processed product of the 37-kDa protein which may arise during virus purification or storage.

To further confirm that section B of ORF 3 encodes the coat protein(s), a *DraI–PstI* fragment of the RTBV genome, representing nucleotides 1905–3888, was inserted into pUC118. *E. coli* transformed with the resultant plasmid produced several proteins that were

RTBV	CRCYICQDENHLANRCP
CoYMV	CKCYICGQEGHYANQCR
CaMV	CRCWICNIEGHYANECP
CERV	CRCWVCNIEGHYANECP
FMV	CRCWICTEEGHYANECP
SoyCMV	CQCWLCHEEGHYANECP

FIG. 6. Comparison of proposed RNA binding sites among known plant pararetroviruses. The conserved amino acid residues are underlined.

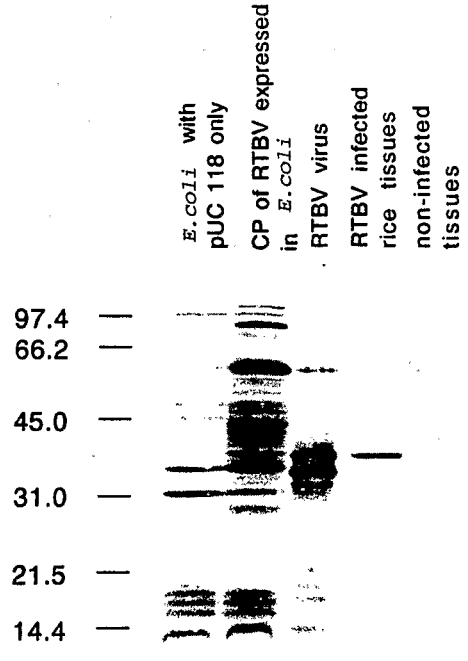


FIG. 7. A Western blot showing RTBV coat protein(s) from purified virus particles, infected rice tissues, and its gene expression in *E. coli*. The sources of the proteins are indicated above the lanes. The molecular weight standards are shown on the left side of the figure.

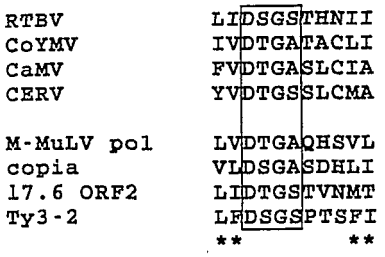


FIG. 8. Comparison of proteinase sequences amongst representative plant pararetroviruses, retroviruses, and retrotransposons (Toh *et al.*, 1985a; Krausslich and Wimmer, 1988; Hansen *et al.*, 1988). The consensus sequence is framed, and the positions conserved for hydrophobic residues are indicated with asterisks. M-MuLV, Moloney murine leukaemia virus; copia and 17.6, retrotransposons in *Drosophila;* Ty3-2, a retrotransposon in yeast.

recognized on a Western blot by the anti-RTBV antibody (Fig. 7). The large number of immuno-reacted polypeptides implies that either the protein is proteolytically degraded in *E. coli* or that there are internal initiation site(s) of transcription and/or translation.

Section C is composed of 676 amino acid residues (aa 944 through 1619) which shares 51% similarity with the ORF V product of CaMV and 53% similarity with a section of ORF 3 in CoYMV (aa 1172 through 1844). Consensus sequences characteristic for some enzymes, namely, proteinase, reverse transcriptase (RT), and ribonuclease H (RH), were identified. The sequence LIDSGSTHNII (aa 985 through 995) shows homology with the active site of retroviral proteinases (Fig. 8) and is therefore predicted to encode a proteinase. The sequence YIDDILI, which matches the reverse transcriptase consensus sequence (Fuetterer and Hohn, 1987) lies 354 amino acid residues downstream (aa 1339 through 1345) of the proteinase sequence motif (Fig. 9). In addition to reverse transcriptase activity, retrovirus replicases also have a RNase H (RH) activity which selectively degrades RNA in RNA-DNA hybrids during reverse transcription. A sequence (aa 1486 through 1608) which shares homology with the proposed RH sequence of CaMV (Doolittle *et al.*, 1989) was also located. A potential relationship between ORF 3 of RTBV and its counterparts in CoYMV, and CaMV, is proposed in Fig. 5.

An intergenic region of 20 nucleotides links ORF 3 and ORF 4. ORF 4 (nt 6042 through 7209) encodes a protein of 389 amino acid residues with a MW of 46.2 kDa. No equivalent ORF was found in CoYMV. Although the genomic location of ORF 4 in RTBV is similar to that of ORF VI in CaMV, the predicted product shares no significant similarity to the one of CaMV ORF VI, nor to other protein sequences in the NBRF Protein database.

DISCUSSION

Rice tungro disease, which is responsible for a substantial yield loss in rice, is caused by coinfection of two viruses: rice tungro spherical virus (RTSV) and rice tungro bacilliform virus (RTBV) (Hibino *et al.*, 1979; Jones *et al.*, 1991). RTSV contains a single-stranded RNA of about 12 kb in length (P. Shen and R. Beachy, unpublished results). The RTBV genome consists of 8000 nucleotides of ds DNA that is organized into four open reading frames and one major intergenic region. DNA sequence analysis and strand-specific hybridization reactions revealed that only one strand of the DNA genome is transcribed. It is believed to encode proteins essential for virus replication.

Identification of a putative primer ($tRNA_i^{met}$) binding site, the redundant termini of the transcript, the homologous amino acid sequence of ORF 3 to the reverse transcriptase of CaMV and other caulimoviruses, and the similarity of genome organization to the *gag–pol* structure of retroviruses (Jacks, 1990), strongly sug-

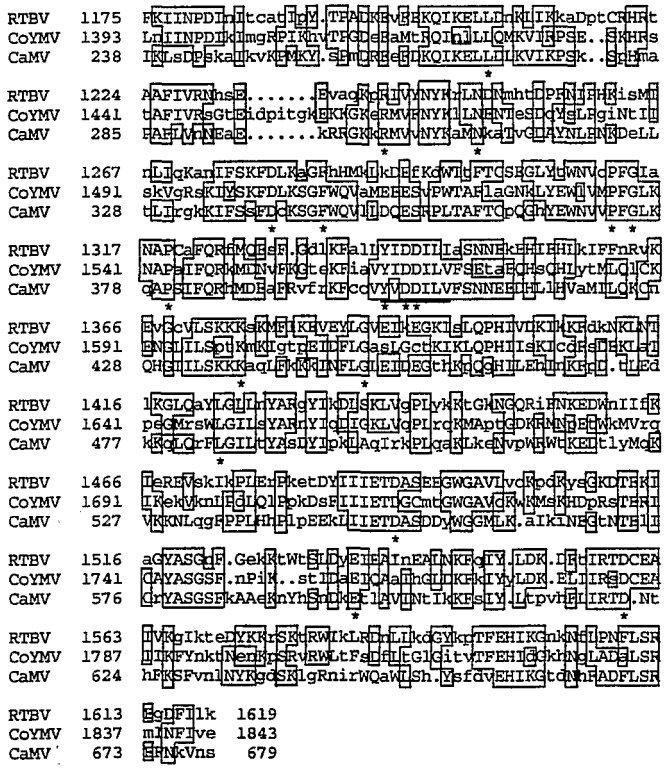


FIG. 9. Alignment of sequences containing consensus sequences of the reverse transcriptase and ribonuclease H between RTBV, CoYMV, and CaMV. The identical amino acid residues are framed. The similar substitutions are labeled with capital letters. The reverse transcriptase consensus sequence is underlined, and the asterisks underneath denote the amino acid residues conserved in retroviruses studied so far (Toh *et al.*, 1985b; Doolittle *et al.*, 1989). Peptide aa 542-672 of CaMV was identified as ribonuclease H sequence by Doolittle *et al.* (1989).

TABLE 1

PREDICTED OPEN READING FRAMES IN THE RTBV GENOME, AND COMPARISON WITH THOSE IN CoYMV

	RTBV					CoYMV		
	Nucleotides			Amino Acids		Amino Acids		
ORFs	Start[a]	Stop[a]	Length	Length	Mol. wt.	Length	Mol. wt.	Homology
1	8000	666	666	222[b]	26.2	200	23.3	28%
2	665	995	330	110	11.9	135	14.8	28%
3	994	6019	5025	1675	194.1	1886	215.5	46%
4	6042	7209	1167	389	46.2	—	—	

[a] The nucleotide number is given, assuming that the translation starts at the first in frame ATG and terminates at the first in frame stop codon, within each open reading frame, except for ORF 1.

[b] ORF 1 has a stop codon, TAA, corresponding to aa 23. The length and the molecular weight listed here are based on the presumption that the stop codon TAA is read through as CAA (encoding Gln, see Feng *et al.,* 1989).

gests that reverse transcription is involved in the replication of RTBV. Thus, RTBV is very likely a plant pararetrovirus (Temin, 1989).

The RTBV genome is different from caulimoviruses by virtue of a long open reading frame (ORF 3) that is predicted to encode a polyprotein of 194.1 kDa. Different sections of the polyprotein share various degrees of similarity with the corresponding sections of the polyprotein of CoYMV and the ORFs IV and V of CaMV. ORF 3 is approximately equivalent to the *gag* and *pol* genes in retroviruses and retrotransposons (Jacks, 1990) and likely encodes the coat protein(s), proteinase, reverse transcriptase, and ribonuclease H.

In retroviruses, caulimoviruses, and most retrotransposons, the *gag* and *pol* genes are in different reading frames or in the same reading frame with a termination codon between the two genes. Translational readthrough of the stop codon by suppression of termination, or by ribosomal frameshifting, has been shown to be the major mechanism responsible for the expression of the *pol* gene and for regulating the ratio of *gag* and *pol* products in retroviruses and most of the retrotransposons (Jacks, 1990). CaMV does not employ these strategies since the translation of *pol* (ORF V) is internally initiated (Penswick *et al.,* 1988). In contrast, the organization of the RTBV genome, in which the equivalents of the *gag* and *pol* genes are in the same open reading frame, is similar to that of CoYMV, a few retrotransposons including copia (Mount and Rubin, 1985), and the retrotransposon-like elements from plants such as Tnt1 (tobacco, Grandbastien *et al.,* 1989) or Ta1 (*Arabidopsis thaliana,* Voytas and Ausubel, 1988). This raises the question of how the synthesis of *gag* and *pol* products is regulated. In copia the major 5-kb mRNA is spliced to produce a 2.1-kb mRNA containing *gag,* but not *pol,* coding sequence (Miller *et al.,* 1989; Boeke and Corces, 1990). The unspliced 5-kb mRNA produces the *gag–pol* fusion protein which is processed to yield the mature *gag* and *pol* proteins. In the case of RTBV several possible strategies could be employed to regulate the amounts of coat protein and replicase (including reverse transcriptase and ribonuclease H domains) such as: mRNA splicing, internal transcription initiation (Plant *et al.,* 1985; Hasagawa *et al.,* 1989), use of alternate termination sites, translational regulation (for example, by ribosomal frameshifting), regulation of polyprotein processing, or protein turnover.

The termination codon within ORF 1 remains a puzzle to us. Its presence would affect the translational efficiency of the downstream ORFs, even though the codon can be read through (Pelham, 1978; Feng *et al.,* 1989), unless translation is reinitiated at the beginning of ORF 2 or ORF 3 (Dixon and Hohn, 1984). It will be important to confirm this observation by analyzing additional isolates of RTBV, and by *in vitro* transcription and translation studies.

Among the RNase H sequences of *E. coli,* retroviruses, and retrovirus-like entities, four acidic amino acid residues are highly conserved (Doolittle *et al.,* 1989). A recent report on the three-dimensional structure of RNase H from *E. coli* showed that these four acidic residues clustered together and might constitute a Mg^{2+}-binding site that was essential for enzyme activity (Katayanagi *et al.,* 1990). Three of these four residues are conserved in the RNase H sequence of RTBV (Asp1490, Glu1535, and Asp1559); the fourth, Asn1608, is a similar substitution to the usual Asp. However, further investigation is needed to determine whether the plant pararetroviruses encode their own RNase H activities, or employ host RNase H (Gordon *et al.,* 1988).

Thus far, amongst the known higher plant viruses, caulimovirus is the sole group containing double-

stranded DNA (Gronenborn, 1987). On the basis of standard criteria, i.e., the transmission vector, and size and shape of the virion, RTBV is quite different from caulimoviruses. While caulimoviruses are icosahedral and transmitted by aphids, RTBV is found to have bacilliform particles of various size which are transmitted by the green leafhopper (*Nephotettix virescens*) (Hibino, 1983). In addition, RTBV is not serologically related to CaMV (unpublished data). Our study of the genome suggests that RTBV may be distantly related to the caulimovirus group. On the other hand, RTBV and CoYMV share similarities in size, shape of virion particles, and genome organization (Lockhart, 1990; Medberry *et al.*, 1990). However, the CoYMV genome is smaller by 511 bp than that of RTBV and lacks ORF 4. Furthermore, the transmission of RTBV by leafhopper requires a helper virus (RTSV) (Hibino, 1983) whereas CoYMV is transmitted by a mealybug (B. Lockhart, personal communication). Recently, Lockhart (1990) proposed the establishment of a new group for nonenveloped bacilliform plant viruses that contain circular double-stranded DNA, termed badnaviruses. We propose to include RTBV in this group, although only when additional viruses of this type are characterized will it be possible to determine whether RTBV belongs to the suggested group, or if it represents a member of a third group of plant ds DNA viruses.

ACKNOWLEDGMENTS

The authors are grateful to Dr. H. Hibino of IRRI for kindly providing RTBV, anti-RTBV antibody and the infected and noninfected rice tissues. We thank Dr. N. Olszewski of University of Minnesota for providing the CoYMV nucleotide sequence, and Dr. R. Hull of John Innes Institute for sending the RTBV manuscript before publication. We are in debt to Monsanto Company for providing oligonucleotide primers. We also thank Drs. P. Shen and C. Fauquet for thoughtful discussion during the preparation of the manuscript. This research is supported by Rockefeller Foundation, Monsanto Company, and the Center for Plant Science and Biotechnology at Washington University.

Note added in proof. The paper by Hay *et al.* dealing with the same subject was published (*Nucleic Acids Res.* **19,** 2615–2621) while this manuscript was in revision. In addition to the two nucleotide difference in length, more than a hundred single nucleotide substitutions were found between the two RTBV sequences. Since the virus samples were from the same source (IRRI), the differences may reflect heterogeneity in the virus isolates.

REFERENCES

BEGGIN, M. D., GIBSON, T. J., and HONG, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.

BOEKE, J. D., and CORCES, V. G. (1989). Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.* **43,** 403–434.

COVEY, S. N. (1986). Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res.* **14,** 623–633.

DEVEREUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12,** 387–395.

DIXON, L. K., and HOHN, T. (1984). Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: Evidence from deletion mutagenesis. *EMBO J.* **3,** 2731–2736.

DOOLITTLE, R. F., FENG, D.-F., JOHNSON, M. S., and MCCLURE, M. A. (1989). Origins and evolutionary relationships of retroviruses. *Quart. Rev. Biol.* **64,** 1–30.

FENG, Y.-X., LEVIN, J. G., HATFIELD, D. L., SCHAEFER, T. S., GORELICK, R. J., and REIN, A. (1989). Suppression of UAA and UGA termination codons in mutant murine leukemia viruses. *J. Virol.* **63,** 2870–2873.

FRANCK, A., GUILLEY, H., JONARD, G., RICHARDS, K., and HIRTH, L. (1980). Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* **21,** 285–294.

FUETTERER, J., and HOHN, T. (1987). Involvement of nucleocapsids in reverse transcription: a general phenomenon? *Trends Biochem. Sci.* **12,** 92–95.

GHOSH, P. K., REDDY, V. B., SWINSCOE, J., CHOUDARY, P. V., LEBOWITZ, D., and WEISSMAN, S. M. (1978). The 5′ terminal leader sequence of late 16S mRNA from cells infected with simian virus 40. *J. Biol. Chem.* **253,** 3643–3647.

GORDON, K., PFEIFFER, P., FUTTERER, J., and HOHN, T. (1988). *In vitro* expression of cauliflower mosaic virus genes. *EMBO J.* **7,** 309–317.

GRANDBASTIEN, M. A., SPIELMANN, A., and CABOCHE, M. (1989). Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature (London)* **337,** 376–380.

GRONENBORN, B. (1987). The molecular biology of cauliflower mosaic virus and its application as plant gene vector. *In* "Plant DNA Infectious Agents" (J. Schell and Th. Hohn, Eds.), pp. 1–29. Springer-Verlag/Wien, New York.

GUILLEY, H., RICHARDS, K. E., and JONARD, G. (1983). Observations concerning the discontinuous DNAs of cauliflower mosaic virus. *EMBO J.* **2,** 277–282.

HANSEN, L. J., CHALKER, D. L., and SANDMEYER, S. B. (1988). Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Mol. Cell. Biol.* **8,** 5245–5256.

HASAGAWA, A., VERVER, J., SHIMADA, A., SAITO, M., GOLDBACH, R., VAN KAMMEN, A., MIKI, K., KAMEYA-IWAKI, M., and HIBI, T. (1989). The complete sequence of soybean chlorotic mottle virus DNA and the identification of a novel promoter. *Nucleic Acids Res.* **17,** 9993–10013.

HIBINO, H., ROECHAN, M., and SUDARISMAN, S. (1978). Association of two types of virus particles with penyaki habang (tungro disease) of rice in Indonesia. *Phytopathology* **68,** 1412–1416.

HIBINO, H., SALEH, N., and ROECHAN, M. (1979). Transmission of two kinds of rice tungro-associated viruses by insect vectors. *Phytopathology* **69,** 1266–1268.

HIBINO, H. (1983). Relations of rice tungro bacilliform and rice tungro spherical viruses with their vector *Nephotettix virescens*. *Ann. Phytopath. Soc. Japan* **49,** 545–553.

HOHN, T., HOHN, B., and PFEIFFER, P. (1985). Reverse transcription in CaMV. *Trends Biochem. Sci.* **10,** 205–209.

HULL, R., SADLER, J., and LONGSTAFF, M. (1986). The sequence of carnation etched ring virus DNA: Comparison with cauliflower mosaic virus and retroviruses. *EMBO J.* **5,** 3083–3090.

HUNKAPILLER, M., LUJAN, E., OSTRANDER, F., and HOOD, L. E. (1983). Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol.* **91,** 227–236.

JACKS, T. (1990). Translational suppression in gene expression in retroviruses and retrotransposons. *Curr. Top. Microbiol. Immunol.* **157,** 93–124.

JONES, M., GOUGH, K., DASGUPTA, I., SUBBA RAO, B. L., CLIFFE, J., QU, R., SHEN, P., KANIEWSKA, M. B., DAVIES, J. B., BEACHY, R. N., and HULL, R. (1991). Rice tungro disease is caused by an RNA and a DNA virus. *J. Gen. Virol.* **72,** 757–761.

KATAYANAGI, K., MIYAGAWA, M., MATSUSHIMA, M., ISHIKAWA, M., KANAYA, S., IKEHARA, M., MATSUZAKI, T., and MORIKAWA, K. (1990). Three-dimensional structure of ribonuclease H from *E. coli. Nature (London)* **347,** 306–309.

KRAUSSLICH, H.-G., and WIMMER, E. (1988). Viral proteinases. *Annu. Rev. Biochem.* **57,** 701–754.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227,** 680–685.

LOCKHART, B. E. L. (1990). Evidence for a double-stranded circular DNA genome in a second group of plant viruses. *Phytopathology* **80,** 127–131.

MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

MATSUDAIRA, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262,** 10035–10038.

MEDBERRY, S. L., LOCKHART, B. E. L., and OLSZEWSKI, N. E. (1990). Properties of *Commelina* yellow mottle virus's complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. *Nucleic Acids Res.* **18,** 5505–5513.

MILLER, K., ROSENBAUM, J., ZBRZEZNA, V., and POGO, A. O. (1989). The nucleotide sequence of *Drosophila melanogaster* copia-specific 2.1-kb mRNA. *Nucleic Acids Res.* **17,** 2134.

MOUNT, S. M., and RUBIN, G. M. (1985). Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol. Cell. Biol.* **5,** 1630–1638.

PELHAM, H. R. B. (1978). Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature (London)* **272,** 469–471.

PENSWICK, T., HUBLER, R., and HOHN, T. (1988). A viable mutation in cauliflower mosaic virus, a retroviruslike plant virus, separates its capsid protein and polymerase genes. *J. Virol.* **62,** 1460–1463.

PFEIFFER, P., and HOHN, T. (1983). Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* **33,** 781–789.

PLANT, A. L., COVEY, C. O., and GRIERSON, D. (1985). Detection of a subgenomic mRNA for gene V, the putative reverse transcriptase gene of cauliflower mosaic virus. *Nucleic Acids Res.* **13,** 8305–8321.

RICHINS, R. D., SCHOLTHOF, H. B., and SHEPHERD, R. J. (1987). Sequence of figwort mosaic virus DNA (caulimovirus group). *Nucleic Acid Res.* **15,** 8451–8466.

SAITO, Y., ROECHAN, M., TANTERA, D. M., and IWAKI, M. (1975). Small bacilliform particles associated with penyakit habang (tungro-like) disease of rice in Indonesia. *Phytopathology* **65,** 793–796.

SAITO, Y., HIBINO, H., OMURA, T., and INOUE, H. (1986). Rice tungro virus. *Tech. Bulletin, the Tropical Agriculture Research Center, Japan.*

SAMBROOK, J., FRITSCH, E. F., and MANIATIS, T. (1989). "Molecular Cloning: A Laboratory Manual." (2nd ed.). Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

SPRINZL, M., HARTMAN, T., MEISSNER, F., MOLL, J., and VORDERWULBECKE, T. (1987). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **15,**(Suppl.) r53–r188.

TEMIN, H. M. (1989). Retrons in bacteria. *Nature (London)* **339,** 254–255.

TOH, H., ONO, M., SAIGO, K., and MIYATA, T. (1985a). Retroviral protease-like sequence in the yeast transposon Ty1. *Nature (London)* **315,** 691.

TOH, H., KIKUNO, R., HAYASHIDA, H., MIYATA, T., KUGIMIYA, W., INOUYE, S., YUKI, S., and SAIGO, K. (1985b). Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and *pol* gene product of moloney murine leukaemia virus. *EMBO J.* **4,** 1267–1272.

TOWBIN, H., STAEHELIN, T., and GORDON, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76,** 4350–4354.

VIEIRA, J., and MESSING, J. (1987). Production of single-stranded plasmid DNA. *Methods Enzymol.* **153,** 3–11.

VOYTAS, D. F., and AUSUBEL, F. M. (1988). A copia-like transposable element family in *Arabidopsis thaliana. Nature (London)* **336,** 242–244.

WU, R., PADMANABHAN, R., and BAMBARA, R. (1974). Nucleotide sequence analysis of bacteriophage DNA. *Methods Enzymol.* **29,** 231–253.