Complete sequence of maize stripe virus RNA4 and mapping of its subgenomic RNAs

Layne Huiet, James H. Tsai and Bryce W. Falk

1Department of Plant Pathology, University of California, Davis, California 95616 and 2FLREC, Fort Lauderdale, Florida 33314, U.S.A.

The complete nucleotide sequence of maize stripe virus RNA4 was determined and found to consist of 2227 nucleotides containing two significant open reading frames. One, in the 5' end of the viral RNA, encodes the major non-capsid protein of Mr 19815. The other is located in the 5' end of the viral complementary RNA and could encode a protein of Mr 31900. This protein has not been identified previously and has been designated NS4, a non-structural protein. RNA-RNA hybridization detected subgenomic RNAs encoding these proteins, a characteristic of RNA possessing an ambisense gene organization.

Introduction

The tenuivirus group consists of five viruses, rice stripe virus (RStV), maize stripe virus (MStV), rice hoja blanca virus, rice grassy stunt virus and European wheat striate mosaic virus, all of which have a plant host range restricted to members of the Poaceae (Gingery, 1988). They are transmitted to plants only by delphacid planthoppers, which are also asymptomatic hosts for the viruses. Infectious virus preparations isolated from infected plants contain fine-stranded nucleoprotein filaments (Gingery, 1988; Ishikawa et al., 1989). Tenuivirus-infected plants also contain two virus-specific proteins, one of approximate Mr 32000 to 35000 which has been identified as the nucleocapsid (N) protein, and another, which is not RNA-associated, of approximate Mr 19000 referred to as the major non-capsid protein (NCP). For MStV it has been shown that NCP is extremely abundant in the plant host, but is not detectable in the insect host (Falk et al., 1987).

The MStV genome consists of five ssRNAs with a total size of approximately 18 kb (8-8 kb, 3-4 kb, 2-3 kb, 2-2 kb and 1-3 kb) (Falk & Tsai, 1984). Each RNA is of predominantly one polarity, although RNA hybridization experiments have shown that small amounts of RNA of the opposite polarity, or viral complementary RNA (vcRNA), are present in the nucleoprotein particles, presumably packaged separately (Falk et al., 1989). Recently, the complete sequences of MStV RNA3 and RStV RNA3 have been obtained (Huiet et al., 1991; Kakutani et al., 1991; Zhu et al., 1991), revealing that RNA3 encodes two proteins, the N protein and a non-structural protein (NS3) in an ambisense gene organization similar to that of the S RNAs of the phleboviruses and tomato spotted wilt tospovirus (Ihara et al., 1984; de Haan et al., 1990). Also, the complete nucleotide sequence of RStV RNA4 has shown that it encodes two proteins in an ambisense gene organization (Kakutani et al., 1990). Previously we have shown that the open reading frame (ORF) encoding MStV NCP is located in RNA4 (Huiet et al., 1990). Here we present the complete sequence of MStV RNA4 and an analysis of its expression via subgenomic RNAs.

Methods

cDNA cloning and DNA sequence analysis. Many of the cDNA clones had been isolated previously, but additional clones were generated by cloning random-primed cDNA into the vector pBluescriptII SK + as described in Falk et al. (1989). Overlapping clones were identified by DNA-DNA hybridization and restriction endonuclease mapping (Huiet et al., 1991). The 3' end of the RNA was cloned by amplification of cDNA synthesized using a primer which is complementary to a conserved 10 base sequence (5' GUCUUUGUGU 3') at the end of tenuivirus RNAs. Viral RNA (2 μg) isolated from nucleocapsids was polyadenylated in 100 μl of 50 mM-Tris-HCl pH 8.0, 10 mM-MgCl₂, 250 mM-NaCl, 0.5 mg/ml bovine serum albumin, 1 mM-DTT, 2.5 mM-MnCl₂, 1 mM-ATP using 2.5 units of poly(A) polymerase (BRL) at 37 °C for 20 min. The RNA was then phenol-extracted and precipitated with 10 μg of glycogen (Boehringer Mannheim) and ethanol. A sample of this polyadenylated RNA (300 ng) was mixed with 1 μg of a primer of the sequence 3' CTGAAACACATTTTTTTT 5' in 10 μl sterile water, heated at 70 °C for 10 min and quickly chilled on ice. The cDNA synthesis reaction was carried out using a cDNA synthesis kit from BRL, in a total reaction volume of 20 μl. This was diluted to 100 μl using 80 μl of 1 × polymerase chain reaction (PCR) buffer (100 mM-Tris–HCl pH 8.3, 500 mM-KCl, 15 mM-MgCl₂, 0.1% gelatin), 1 μg of second strand primer (5' TCAGATTTCGGTTTATCCTC 3') and 1 μl Taq polymerase. This reaction was carried out
for 1 min at 94 °C, 1 min at 37 °C, 1 min at 72 °C, and repeated for 25 cycles using a Perkin-Elmer Cetus DNA Thermo Cycler. DNA sequencing was performed using both the dideoxynucleotide chain termination method using modified T7 DNA polymerase (US Biochemicals) and the chemical degradation method (Tabor & Richardson, 1987; Maxam & Gilbert, 1980). Modifications were as described previously (Huiet et al., 1991). The entire sequence was determined on both strands and approximately 60% was obtained from at least two independent cDNA clones. The 3' end region, cloned by PCR, was sequenced from several independent clones.

Computer analysis. DNA sequence analysis and protein comparisons were performed using the software DNA Strider and DNA Molly for the Apple Macintosh. RNA structural analysis and protein database searches were carried out on a VAX 8600 using the programs from the Wisconsin GCG software package.

RNA hybridization and sequencing. MStV nucleoprotein RNA and total RNA were isolated according to the method of Falk & Tsai (1984). For RNA blot hybridization, RNA was denatured with glyoxal, fractionated by agarose gel electrophoresis and transferred to Nytran membranes using methods described previously (Falk et al., 1989). RNA–RNA hybridization was done at 55 °C in 50% formamide 3 x SSC, 10 x Denhardt's solution, 0.1% SDS, 50 μg/ml ssDNA, 10 μg/ml poly(A) for 16 to 24 h. The filters were washed several times at room temperature in 3 x SSC, 0.5% SDS. They were then treated with RNase A to reduce non-specific binding of the RNA probe as described below. The filters were rinsed for 5 min in 3 x SSC and then incubated in 3 x SSC plus 1 μg/ml RNase A for 15 min at room temperature. RNA hybridization probes were transcribed in vitro as described in Falk et al. (1989). Hybridization was quantified using an AMBIS radiolysis system. Direct sequencing of the 5' end of the RNA was performed using dideoxynucleoside triphosphates and avian myeloblastosis virus (AMV) reverse transcriptase (Life Sciences) as described previously (Huiet et al., 1990). A specific oligonucleotide, 5' CAGGCCTGTAATAGGGCCAA Y, complementary to nucleotides 88 to 107 of the sequence, was used as a primer.

Results

DNA sequence analysis of a 1030 bp cDNA clone has previously shown that the gene encoding NCP is located in the 5' end of MStV RNA4 (Huiet et al., 1990). This clone was used in DNA–DNA hybridization experiments to isolate overlapping cDNA clones representing the entire RNA sequence. The RNA4 3'—terminal region was cloned by using PCR with sequence-specific primers as described in Methods. The 5' end sequence was determined by directly sequencing the viral RNA using a specific primer in conjunction with AMV reverse transcriptase and dideoxyxynucleotides. The complete 2227 nucleotide sequence of RNA4 is shown in Fig. 1 and the predicted ORFs are shown in Fig. 2.

There are two ORFs in the RNA4 sequence. One ORF of 528 nucleotides located in the 5' end of the RNA had previously been identified as that encoding NCP. The second ORF, of 852 nucleotides, is located at the 3' end of the RNA sequence, in the complementary sense. The second ORF would encode a protein of Mr, 31 903 which has not been identified and it has been designated NSc (non-structural protein from RNA4). Analysis of the NSc protein for any distinguishing features revealed no large hydrophobic or hydrophilic regions, and no asparagine-linked glycosylation sites. A search revealed no significant homology to any proteins in the PIR and SwissProt databases. RNA4 has an overall A + U content of 64% which increases to 71% in the intergenic region between nucleotides 600 and 1300. Within this region there are three long regions in which the dinucleotide AU is repeated 11, 14 and 21 times. Two of these repeats occur in the region 700 to 800 nucleotides from the 5' end and could potentially base pair to form a hairpin structure.

To determine whether subgenomic mRNAs are produced from RNA4, we carried out RNA–RNA hybridization experiments using ssRNA probes representing the 5' end and the 3' ends of both the viral RNA and vcRNA (Fig. 3a). The results of these experiments are shown in Fig. 3. All samples were identical (duplicates) and were fractionated on the same gel. Panels (b) and (c) were hybridized to single-stranded probes from the 3' end of the vcRNA and the 5' end of the viral RNA respectively. In both (b) and (c) there is hybridization to the 2.2 kb genomic RNA present in the viral RNA and in the MStV-infected Zea mays total RNA. In (b) there is also hybridization to a subgenomic RNA present only in total RNA samples from infected Z. mays plants. In these samples the level of hybridization to the subgenomic RNA was twofold that to the genomic RNA, as measured by an AMBIS 2-D radioactivity counter. Neither probe hybridized to RNA extracted from healthy Z. mays plants. Fig. 3(d) and (e) show the hybridization results using probes from the 5' end of the vcRNA and the 3' end of the viral RNA, respectively. In addition to the viral RNA, we detected an approximately 1000 nucleotide subgenomic RNA only in (e). These results indicated that two subgenomic RNAs are produced which represent RNA4. One subgenomic RNA represents the 5' end of the viral RNA and would contain the NCP coding region. The other subgenomic RNA represents the 5' end of the vcRNA and contains the NSc ORF.

Discussion

DNA sequence analysis has revealed that MStV RNA4 contains two ORFs in an ambisense gene organization identical to that found in RNA3. The ORF in the 5' end of the viral RNA has been identified previously as encoding NCP (Huiet et al., 1990). The 5' end of the vcRNA encodes a protein, NSc, which has not been identified previously. RNA–RNA hybridization results have demonstrated that two subgenomic RNAs repre-
Fig. 1. Complete nucleotide sequence of MStV RNA4 including its two predicted protein-encoding regions; the amino acid sequence of the NCP encoded in the 5' end of the RNA is shown below the corresponding nucleotide sequence. The NS4 amino acid sequence (derived from the complementarity sequence) is encoded in the 5' end of the vcRNA. Asterisks indicate termination codons for each ORF. The nucleotide sequence is shown as DNA.
Fig. 2. The predicted ORFs for both polarities of MStV RNA4. Vertical lines indicate stop codons, half lines indicate locations of ATG triplets. The three possible reading frames for RNA4 (1, 2 and 3) and for vcRNA4 (−1, −2 and −3) are shown.

Fig. 3. (a) Schematic representation of both polarities of RNA4 and the plasmids used. (b to e) RNA–RNA hybridization analysis of MStV RNA4. Panels (b) and (c) were hybridized to ssRNA transcripts generated from plasmid pMS58 (nucleotides 3 to 1030), panels (d) and (e) were hybridized to transcripts from plasmid pMS505 (nucleotides 1389 to 2125). Transcripts used in (b) and (d) are of vcRNA polarity, transcripts used in panels (c) and (e) are of viral RNA polarity. Samples are duplicates which were glyoxalated, fractionated on the same 1–5% agarose gel and transferred to nylon membranes. Lanes 1, RNA isolated from viral nucleoprotein particles; lanes 2, total RNA from MStV-infected Z. mays plants; lanes 3, total RNA from healthy Z. mays plants. g, Genomic RNA; sg, subgenomic RNA.

Fig. 4. Comparison of the two RNA4-encoded proteins from MStV and RStV indicating exact matches. (a) Alignment of NS4 proteins, (b) alignment of NCPs.

senting RNA4 are produced in MStV-infected Z. mays plants. One subgenomic RNA represents the 5' end of the viral RNA and would contain the NCP coding region. The other subgenomic RNA represents the 5' end of the vcRNA and contains the NS4 coding region. Therefore these subgenomic RNAs are not generated by internal initiation, but from the 3' end of the vcRNA for the NCP subgenomic RNA, and from the 3' end of the viral RNA for the NS4 subgenomic RNA. This mechanism for the generation of subgenomic RNAs is found for all RNAs which possess an ambisense gene organization.

In these experiments no subgenomic RNA, including that encoding NCP, was detected in infected P. Peregrinus maidis (data not shown). However in infected Z. mays plants NCP is present in large amounts and therefore it is not surprising that the subgenomic RNA containing the NCP coding region is produced in large amounts (at a ratio of 2:1 to genomic RNA). Previous studies have not been able to detect the presence of NCP in infected P. maidis and the RNA hybridization results may indicate that NCP is not expressed owing to the lack of a detectable subgenomic RNA. Alternatively, it may be synthesized but rapidly degraded and therefore no NCP is produced. So far, other subgenomic RNAs detected in infected Z. mays plants also have not been detected in P. maidis (data not shown).
Comparison of the MStV RNA4 sequence with that previously reported for RStV RNA4 shows that MStV RNA4 is 90 nucleotides longer. Comparison of the 5' and 3' end sequences shows that the first 20 nucleotides at both the 5' and 3' ends are identical in MStV and RStV RNA4; 11 nucleotides at the 5' end and 10 at the 3' end of MStV RNA4 are identical to the 5' end and 3' end sequences of both RStV and MStV RNA3. The 5'- and 3'-terminal regions are complementary for 17 of 18 nucleotides, as is the case for the other tenuivirus RNAs which have been sequenced (Huitet et al., 1991; Kakutani et al., 1990, 1991; Zhu et al., 1991).

Comparison of the predicted amino acid sequences of MStV NS4 and NCP with those of the same proteins encoded by RStV RNA4 reveals 76% and 73% identity respectively (Fig. 4). This result correlates with the lack of cross-reactivity between RStV NCP and MStV NCP antiserum (Gingery et al., 1983). The percentage identity between these proteins is higher than that between the N proteins of these viruses, which is 65% (unpublished data). It is interesting to speculate as to the possible function of the NS4 protein. Its location in the 5' end of RNA4 vRNA, just as the N protein is located in the 5' end of RNA3 vRNA, may indicate an importance in the early stages of the infection process.

The authors would like to thank Lih Chin for help with the computer analysis. This work was supported in part by USDA Grant 90-37262-5232. This is paper number R-02222 from the Florida Agricultural Experiment Station for J. H. Tsai.

References


(Received 2 January 1992; Accepted 30 March 1992)