The S RNA segment of tomato spotted wilt virus has an ambisense character

Peter de Haan,* Lia Wagemakers, Dick Peters and Rob Goldbach

Department of Virology, Agricultural University, Binnenhaven 11, 6709 PD Wageningen, The Netherlands

The complete nucleotide sequence of the S RNA of tomato spotted wilt virus (TSWV) was determined. The RNA is 2916 nucleotides long and has an ambisense coding strategy. The sequence contains two open reading frames (ORFs), one in the viral sense which encodes a protein with a predicted Mr of 52.4K and one in the viral complementary sense which encodes the viral nucleocapsid protein of Mr 28.8K. Both proteins are expressed by translation of two subgenomic RNA species that possibly terminate at a long stable hairpin structure, located at the intergenic region. The structure of this RNA segment resembles that of the arthropod-borne phleboviruses (family Bunyaviridae). The absence of significant sequence homology between TSWV and bunyaviruses infecting animals suggests that TSWV should be considered as a representative of a new genus within the Bunyaviridae.

Introduction

Among the plant viruses, tomato spotted wilt virus (TSWV) is unique in its particle morphology and genome structure. It has therefore been classified as the only representative of a distinct group, the tomato spotted wilt virus group (Be, 1970; Matthews, 1982). The virus has a broad host range and causes large yield losses in many economically important crops. It is transmitted exclusively by thrips species in a persistent manner (Sakimura, 1962; Paliwal, 1974).

TSWV virions are spherical enveloped particles, about 80 to 110 nm in diameter, that contain four different proteins: an internal nucleocapsid protein (N) of Mr 27,000 (27K), two membrane glycoproteins (G1 and G2) of 78K and 58K respectively, and a large protein (L) of approximately 120K (Mohamed et al., 1973; Tas et al., 1977). The genome consists of three linear ssRNA molecules approximately 3100 (S RNA), 5400 (M RNA) and 8200 (L RNA) nucleotides long, each complexed with N proteins to form circular nucleocapsids (van den Hurk et al., 1977; Mohamed, 1981; de Haan et al., 1989a). Based on this and other properties it has been suggested that TSWV should be considered as a member of the arthropod-borne Bunyaviridae (Milne & Francki, 1984; de Haan et al., 1989a). Recently the S and M RNA segments of TSWV have been cloned and their terminal sequences determined (de Haan et al., 1989b). This analysis has revealed that, like RNAs of the Bunyaviridae and other negative-stranded RNA viruses, TSWV RNAs have complementary ends.

To gain more insight into the molecular properties of TSWV and to study its possible relationship to the Bunyaviridae in more detail, information on the nucleotide sequence of the TSWV genome is required. In this paper we report the complete nucleotide sequence of the TSWV S RNA. The data show that this RNA molecule has an ambisense gene arrangement, with a putative 'non-structural' protein gene located on the viral RNA strand and the nucleocapsid protein gene located on the viral complementary RNA strand. Evidence will be provided that both genes are expressed by the synthesis of two subgenomic RNA molecules.

Methods

Cloning of cDNA and sequence determination. Complementary DNA to the S RNA segment of TSWV isolate CPNH1 was synthesized and cloned as described previously (de Haan et al., 1989b). DNA sequencing was performed by the dideoxynucleotide chain termination method (Sanger et al., 1977), after subcloning of restriction fragments in M13 mp18 or mp19 vectors (Yanisch-Perron et al., 1985). The RNA sequences of the 3' and 5' ends of the TSWV S RNA were determined as described before (de Haan et al., 1989b). Sequence data were stored and edited using the Staden programs (Staden, 1982).

RNA isolation and Northern blot analysis. Total RNA from healthy and TSWV-infected Nicotiana rustica var. America plants was isolated according to De Vries et al. (1982). RNAs in samples of 5 μg were separated by electrophoresis in 1% agarose gels, under denaturing conditions (Bailey & Davidson, 1976), transferred to nitrocellulose and hybridized to 32P-labelled DNA probes by standard methods (Maniatis et al., 1982).

In vitro transcription and cell-free translation. For in vitro expression purposes, viral cDNA was cloned in the transcription vector pSK +
1002  P. de Haan and others

transcripts were synthesized using T7 RNA polymerase according to downstream of the cDNA insert and messenger sense 'run-off' TSWV nucleocapsids, prior to electrophoresis, as described by Bernstein & Hruska (1981).

Results

Cloning and sequencing of the TSWV S RNA

The nucleotide sequences of 30 nucleotides at the 3' end of the S RNA and five nucleotides at the 5' end were determined after end-labelling of the RNA, followed by partial degradation with base-specific ribonucleases. These experiments revealed, in comparison to the previously reported terminal sequences (de Haan et al., 1989b), one extra A residue at the 5' end and one extra U residue at the 3' end of the S RNA segment. From the deduced sequence, a synthetic oligonucleotide (S1) was synthesized to be complementary to the 20 3' terminal nucleotides. This oligonucleotide was used for synthesis and cloning of cDNA, and for determination of the nucleotide sequence of approximately 200 nucleotides at the 3' end of the RNA, by primer extension (de Haan et al., 1989b). Based on the restriction map of S RNA-specific cDNAs, the clones 514, 520, 608 and 614 were selected for further sequence analysis (Fig. 1). The sequence of each DNA restriction fragment was determined from both strands. Finally, a second synthetic oligonucleotide (S3) was synthesized, corresponding to the most 5'-proximal sequence of cDNA clone 520, and the 5'-terminal sequence of the S RNA (5' AGAGCAA 3') was verified by primer extension.

The complete sequence of the TSWV S RNA (Fig. 2) is 2916 nucleotides long. This is approximately 200 to 500 nucleotides shorter than previously reported values, which were based on electrophoretic mobilities (Verkleij & Peters, 1983; de Haan et al., 1989b). Its base composition is 31.6% A, 32.9% U, 19.3% C and 16.0% G. The 3'- and 5'-terminal sequences are complementary over a stretch of 65 to 70 nucleotides and can be folded into a stable panhandle structure, with a free energy of $\Delta G = -254.1$ kJ/mol (Fig. 3a). An internal inverted complementary sequence of U-rich stretches followed by A-rich stretches is located between positions 1582 and 1834 (numbered from the 5' end of the viral RNA). This sequence can be folded into a stable hairpin structure with a $\Delta G = -452.7$ kJ/mol (Fig. 3b).

Fig. 1. Cloning strategy for the TSWV S RNA segment. S1 and S3 correspond to the synthetic oligonucleotides used for primer extension sequencing as represented by the open arrows. The numbers correspond to the cDNA clones used. Restriction enzymes are abbreviated as follows: Bg, BgII; E, EcoRI; H, HindIII.

Predicted gene products of TSWV S RNA

Analysis of the six different reading frames of the viral and viral complementary RNA strand revealed two long open reading frames (ORFs), one on each strand. No additional ORFs of significant length (i.e. ORFs encoding proteins larger than 5K) could be detected in any of the other reading frames (Fig. 4). The ORF on the viral RNA strand starts with an AUG codon at position 88 and terminates at an UAA stop codon at position 1481 (numbered from the 5' end of the viral strand), which corresponds to a protein of 465 amino acids and an $M_r$ of 52.4K. The amino acid sequence of this putative protein does not contain hydrophobic regions that might function as signal peptides or transmembrane domains, according to the hydropathy algorithms of Hopp & Woods (1981) and Kyte & Doolittle (1982) (data not shown). Although the sequence contains five potential N-glycosylation sites, it is not known whether these sites are indeed glycosylated in vivo. No significant homology could be detected to any other protein in the EMBL protein and nucleotide sequence database.

The ORF on the viral complementary RNA strand starts with an AUG codon at position 2763 and terminates at a UGA stop codon at position 1989 (numbered from the 5' end of the viral strand); it encodes a protein of 259 amino acids with an $M_r$ of 28.8K. The amino acid sequence also does not contain any strongly hydrophobic regions and there are no possible N-glycosylation sites present. The length of this ORF suggests that it encodes the nucleocapsid protein. In order to verify this, a cDNA fragment containing this ORF was expressed in vitro. To this end a 387 bp EcoRI cDNA fragment of clone 614 was fused to the 494 bp HindIII/EcoRI cDNA fragment of clone 520 and subsequently cloned in pSK+ (pTSWV-vcORF). Run-
Nucleotide sequence of TSWV S RNA

Fig. 2. The complete nucleotide sequence of TSWV S RNA (numbered from the 5' end) and the predicted gene products. The deduced amino acid sequence of the protein encoded by the viral RNA is written above the RNA sequence. The sequence of the protein encoded by the viral complementary strand is written below the RNA sequence. Potential N-glycosylation sites are underlined. The asterisks (*) indicate termination codons.
Fig. 3. The inverted complementary sequence at the 5' and 3' end (a) and the secondary structure at the intergenic region of TSWV S RNA (b). The nucleotide positions are numbered from the 5' end. Asterisks (*) represent gaps corresponding to unpaired nucleotides in the sequence.
transcripts were synthesized and translated in a rabbit reticulocyte lysate (Fig. 5). One discrete translation product was obtained (Fig. 5, lane 4), which comigrated with purified viral nucleocapsid protein (Fig. 5, lane 8), and moreover which reacted with an antiserum raised against purified nucleocapsids (Fig. 5, lane 7). Translation products of unrelated plant virus RNAs [tobacco mosaic virus (TMV) and cowpea chlorotic mottle virus (CCMV)] were not precipitated by this antiserum (Fig. 5, lanes 5 and 6). These results indicate that the ORF located on the viral complementary strand of the TSWV S RNA represents the nucleocapsid gene. When searching the EMBL protein and nucleotide sequence database, no significant homology could be detected to other nucleocapsid proteins of positive- or negative-stranded viruses, or to any other protein.

Detection of subgenomic RNA species in TSWV-infected plant cells

The ambisense gene arrangement of the S RNA predicts that the two ORFs are expressed by the formation of two subgenomic mRNA species, as found for other ambisense RNA molecules of members of the Arenaviridae and Phlebovirinae (Auperin et al., 1984; Ihara et al., 1984; Salvato, 1989). To identify possible mRNA species corresponding to the nucleocapsid protein and the putative non-structural protein, Fig. 6 shows Northern blots of total RNA isolated from healthy and infected plants, hybridized to 32P-labelled probes. The probe corresponding to the ORF on the viral strand (probe NSs) hybridized to the genomic S RNA and to a subgenomic RNA species of approximately 1.7 kb. With the probe corresponding to the nucleocapsid gene (probe N), a second smaller subgenomic mRNA (1.2 kb) was detected, in addition to full-length S RNA. These results indicate that the expression strategy of the TSWV S RNA is similar to that of the ambisense RNAs of arenaviruses and phleboviruses.

Discussion

The sequence data presented here show that the TSWV S RNA is 2916 nucleotides long. The discrepancy between this and the previously estimated size may be due to the tight association of nucleocapsid protein with this RNA, retarding the migration of RNA.

From the sequence it can be deduced that the TSWV S RNA has an ambisense gene arrangement. The Northern blot hybridization experiments demonstrate that the two genes on this RNA are expressed by two subgenomic RNA species. This genome strategy is also found for the S RNA segments of viruses of the genus phlebovirus (Ihara et al., 1984; Marriott et al., 1989) and for the S and L RNA segments of viruses of the family Arenaviridae (Auperin et al., 1984; Romanowski et al., 1985; Clegg & Oram, 1985; Salvato, 1989), but has not yet been reported for any plant virus. Together with morphological data (de Haan et al., 1989a), the structure and genetic organization of the S RNA provide strong evidence that
TSWV should be regarded as a member of the Bunyaviridae. Preliminary nucleotide sequence data demonstrate that the TSWV M RNA, like the corresponding RNA of the animal Bunyaviridae, is entirely of negative polarity (P. de Haan, unpublished results).

Compared to the ambisense RNA molecules of the animal phlebo- and arenaviruses, TSWV S RNA contains relatively long terminal untranslated regions. This may be related to the formation of a long stable panhandle structure at the termini of the genomic RNA molecule and may explain the circular appearance of isolated TSWV nucleocapsids (de Haan et al., 1989b). These non-coding sequences probably contain important signals for replication and transcription (Strauss & Strauss, 1988). Another putative regulatory element is located in the intergenic region of the RNA, where the sequence can be folded into a long hairpin structure. A similar hairpin structure has been found in the S RNA of Punta Toro virus (Emery & Bishop, 1987). For this virus it is known that the peak of the hairpin forms the transcription termination point for the two subgenomic RNA molecules produced from this RNA. The estimated sizes of the subgenomic RNA species of TSWV S RNA (1.2 and 1.7 kb) and the location of the hairpin structure are consistent with TSWV using the same mechanism for transcription termination. Investigations are currently in progress to study the regulation of the expression of both genes of TSWV S RNA and to map precisely the transcription initiation and termination sites.

TSWV S RNA appears to encode a non-structural protein which we propose to designate NSs, according to the nomenclature used for the Bunyaviridae. It does not show significant amino acid homology to any of the reported phlebovirus NSs proteins (Marriott et al., 1989). The protein has not yet been detected in TSWV-infected plant cells and its role in the viral infection process is so far unknown.

In conclusion, the data presented in this paper indicate that TSWV should be considered as a member of the Bunyaviridae. However in view of the lack of homology in both the nucleotide sequence and the derived amino acid sequence to any of the animal Bunyaviridae analysed so far, we propose to place TSWV into a new genus, which could be named phytophlebovirus. In view of its proposed taxonomic position the question needs to be answered as to whether TSWV replicates also in its insect vectors. Further studies on its genome organization and replication may reveal how this virus departed evolutionarily from other bunyaviruses as an insect–animal virus to become an insect–plant virus.

The authors wish to thank Yoshitaka Sano and Richard Kormelink for their help in some of the experiments.

References


(Received 13 November 1989; Accepted 29 January 1990)