Tomato spotted wilt virus L RNA encodes a putative RNA polymerase

Peter de Haan,* Richard Kormelink, Renato de Oliveira Resende, Frank van Poelwijk, Dick Peters and Rob Goldbach

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

The complete nucleotide sequence of the large (L) genome segment of tomato spotted wilt virus (TSWV) has been determined. The RNA is 8897 nucleotides long and contains complementary 3' and 5' ends, comprising 62 nucleotides at the 5' end and 66 nucleotides at the 3' end. The RNA is of negative polarity, with one large open reading frame (ORF) located on the viral complementary strand. This ORF corresponds to a primary translation product of 2875 amino acids in length, with a predicted Mr of 331500. Comparison with the polymerase proteins of other negative-strand viruses indicates that this protein most likely represents the viral polymerase. The genetic organization of TSWV L RNA is similar to that of the L RNA segments of Bunyamwera and Hantaan viruses, animal-infecting representatives of the Bunyaviridae.

Introduction

Based on its unique properties among other plant viruses, tomato spotted wilt virus (TSWV) has previously been classified as the single representative of a distinct virus group (Ie, 1970; Matthews, 1982). Recently, molecular data have provided evidence that TSWV should be considered as a member of the arthropod-borne Bunyaviridae, although unique in being able to infect plants (de Haan et al., 1989a, b, 1990).

Like the established members of the Bunyaviridae (Elliott, 1990), TSWV is characterized by spherical enveloped particles of approximately 80 to 110 nm in diameter. Two virus-encoded glycoproteins, denoted G1 (M, 78K) and G2 (M, 58K) are associated with the virus envelope (Tas et al., 1977). The internal pseudo-circular nucleocapsids consist of three species of ssRNA, denoted S RNA (2916 nucleotides), M RNA (approximately 5000 nucleotides) or L RNA (approximately 8000 nucleotides), which are tightly encapsidated with the nucleocapsid (N) protein (Mr 28.8K) (de Haan et al., 1989 b). In addition a few copies of a large (L) protein (approximately 200K) are present in the virus particle, and may represent the viral polymerase (Mohamed et al., 1973; Mohamed, 1981; Tas et al., 1977).

Recently, the genomic RNA segments have been cloned (de Haan et al., 1989b) and the complete nucleotide sequence of the S RNA has been determined from a set of overlapping cDNA clones (de Haan et al., 1990). TSWV S RNA encodes two proteins, the N protein and a non-structural (NSs) protein, in an ambisense gene arrangement. The N protein is expressed from a subgenomic mRNA species of approximately 1-2 kb, transcribed from the viral RNA strand, and the NSs protein (M, 52-4K) is expressed from an mRNA of approximately 1-7 kb, transcribed from the viral complementary RNA strand. The structure of TSWV S RNA conforms with that of the phleboviruses and uukuviruses, two genera of the family Bunyaviridae (Giorgi et al., 1991).

Here we report the complete nucleotide sequence of TSWV L RNA. It contains a single large open reading frame (ORF) in the viral complementary sense, which most probably corresponds to the viral polymerase gene. The genetic organization of the TSWV L RNA segment further strengthens our previous conclusion that this virus represents a plant-infecting member of the Bunyaviridae.

Methods

**Viruses and plants.** TSWV CNPH1 (now BR-01), a Brazilian isolate from tomato, was maintained in tomato by grafting and infected leaf tissue was stored in liquid nitrogen. *Nicotiana rustica* plants were either mechanically inoculated from this original virus stock, or from previously inoculated, systemically infected *N. rustica*. Virus was purified from infected *N. rustica* leaves according to Tas et al. (1977) and RNA was extracted as described previously (de Haan et al., 1989b).

**Synthesis, cloning and sequence determination of cDNA.** cDNA to TSWV RNA was synthesized and cloned as previously described (de Haan et al., 1989b). To obtain cDNA clones containing the 3' end of the L RNA, a 5 pg portion of genomic RNA was polyadenylated at the 3' end, using 1 unit of poly(A) polymerase (Bethesda Research Laboratories), according to Devos et al. (1976). First-strand cDNA synthesis was primed with oligo(dT), followed by second-strand synthesis according
to Gubler & Hoffman (1983). Double-stranded cDNA was made blunt-ended using T4 DNA polymerase and subsequently cloned into the Smal site of plasmid pUC19 (Maniatis et al., 1982).

DNA sequencing was performed by the dideoxynucleotide chain termination method (Sanger et al., 1977), on dsDNA templates (Zhang et al., 1988), or after subcloning of restriction fragments in M13mp18 or -mp19 vectors (Yanisch-Perron et al., 1985). Nucleotide and amino acid sequences were compiled and analysed using programs developed by the University of Wisconsin Genetics Computer Group (UWGCG).

Results

Cloning and sequence determination of the TSWV L RNA

Northern blot analysis of genomic RNA, purified from the original BR-01 virus stock, revealed that the previously reported restriction map of TSWV M RNA (de Haan et al., 1989b) actually represented that of a defective L RNA molecule of 4.7 kb in length. This defective RNA molecule was abundantly present in the TSWV BR-01 line used in this study and masked the authentic M RNA segment (5.0 kb). This TSWV line had been maintained by mechanical passage of the virus for many years.

In order to obtain cDNA clones corresponding to the full-length genomic RNA sequence, the original cDNA library (de Haan et al., 1989b) was screened again, and additional cDNA clones to TSWVL RNA could be aligned, yielding a restriction map covering approximately 8900 nucleotides (Fig. 1). The cDNA clones denoted 70, 266, 280, 299, 329, 420, 662, 669, 803, 808 and 810 were selected for sequence analysis. Since clones 280, 803, 806 and 808 hybridized only to the full-length L RNA and not to the defective L RNA molecule (results not shown), it can be assumed that the latter molecule is the result of an internal deletion in TSWV L RNA. The nucleotide sequences and origin of defective L RNA species in TSWV isolates will be discussed in a separate paper.

Direct dideoxynucleotide sequencing, using L RNA as a template and four different synthetic oligonucleotides as primers, was used to obtain the 5'-terminal sequence and to verify internal sequences (Fig. 1). To obtain cDNA clones containing the 3'-terminal sequences of the L RNA, genomic RNA was polyadenylated and cDNA was synthesized by priming first-strand cDNA synthesis with oligo(dT). Clones were subsequently selected, using a 830 bp EcoRI/SphI restriction fragment of cDNA clone 662 as a probe in a colony hybridization experiment. One of the selected clones, denoted 669, contained the sequence 5' ..ACCTGATTGCTCT(A)22 3', which is complementary to the sequence at the 5' end of the TSWV L RNA (5' AGAGCAAUC.. Y), as determined by primer extension sequencing (Fig. 1). These terminal sequences are also identical for the first eight nucleotides to the 3' and 5' termini of the S RNA (de Haan et al., 1990), indicating that the entire L RNA sequence was indeed included. The identification of clone 669 as an L RNA-specific cDNA clone was further confirmed by Northern blot hybridization (results not shown).

Characteristics of the TSWV L RNA

The complete nucleotide sequence of the TSWV L RNA is shown in Fig. 2. The RNA is 8897 nucleotides long,
with a base composition of 28.7% A, 37.8% U, 19.0% C and 14.5% G. The length is in rather good agreement with the previously estimated size, deduced from electrophoretic mobility (Van den Hurk, 1977; de Haan et al., 1989a). The L RNA exhibits complementarity between its 3' and 5' ends for 62 nucleotides at the 3' end to 66 nucleotides at the 3' end, similar in range to the complementary termini of the S RNA (de Haan et al., 1990). The resulting 'panhandle' structure (Fig. 3) has a free energy of $\Delta G = -217.1$ kJ/mol. Moreover, the 10 3'- and 5'-terminal nucleotides show a remarkable homology to that of RNA segment 3 of Thogoto virus, a tick-borne member of the Orthomyxoviridae (Clerx et al., 1983; Staunton et al., 1989) (Fig. 4).
Predicted gene product encoded by TSWV L RNA

Analysis of the six reading frames of the viral and viral complementary RNA strand revealed only one large ORF, located on the viral complementary RNA strand (Fig. 5). This ORF starts with an AUG codon at position 34 and extends to a UAA stop codon at position 8659, hence the non-coding regions of the plus-sense RNA are 33 bases long at the 5' end and 235 bases at the 3' end.

The amino acid sequence derived from this ORF is shown in Fig. 2. The sequence of the predicted gene product is 2875 amino acids long and has an estimated
Analysis of the amino acid sequence of the predicted protein reveals several short hydrophobic regions (Kyte & Doolittle, 1982) and a very acidic carboxy terminus, as can be seen by the large number of aspartic acid (D) and glutamic acid (E) residues (Fig. 2).

A search in the EMBL protein and nucleotide sequence database revealed that the predicted protein encoded by TSWV L RNA is homologous to the L proteins of the animal-infecting Bunyaviridae. Hence, it can be deduced that the L RNA segment of TSWV encodes the L protein. The discrepancy between the size reported here (331-5K) and the previously estimated size (200K) may be due to the gel systems used in those experiments (Mohamed et al., 1973; Tas et al., 1977).

Computer-assisted alignment of the predicted L protein of TSWV with that of Bunyamwera virus (Elliott, 1989) reveals one internal region (approximately 1000 amino acids long) with significant (27% identity) amino acid sequence homology (Fig. 6). Homology between TSWV and Hantaan virus L proteins, and between those of Bunyamwera and Hantaan virus, however, is lower and restricted to a shorter internal stretch of approximately 200 to 250 residues (Fig. 6).

For the animal-infecting Bunyaviridae it has been proposed that the L proteins represent the viral RNA polymerases. Proteins involved in transcription and replication of RNA viruses contain conserved signature sequences, such as putative polymerase, helicase or methyltransferase motifs (Kamer & Argos, 1984; Goldbach, 1987; Hodgman, 1988; Gorbalenya et al., 1989). The presence or absence of these motifs, together with other molecular characteristics such as genome structure...
Fig. 7. Amino acid sequence homology between the RNA polymerases of members of the Bunyaviridae and protein P1, the core polymerase of the influenza virus. Residues conserved in at least three sequences are indicated in bold. Sequence data were obtained from Yamashita et al., 1989; Candresse et al., 1990). A search for such conserved sequences in the (putative) polymerase proteins of members of the Bunyaviridae and Orthomyxoviridae reveals five types of short consensus sequences: GDX1-K, GXXNXXS, SDD, FX10-17K and EFXSXF (Fig. 7). These amino acid motifs are present in the region where the predicted L protein of TSWV shows sequence homology to Bunyamwera and Hantaan L proteins and to influenza A virus protein P1, the core polymerase of this virus (Braam et al., 1983; Krug et al., 1989). Hence, it is anticipated that the major ORF in TSWV L RNA represents the polymerase gene.

Discussion

Determination of the complete nucleotide sequence of the TSWV L RNA demonstrates that TSWV is a negative-strand RNA virus. The presented nucleotide sequence data confirm the previous conclusion, derived from the S RNA sequence, that TSWV should be considered as a member of the Bunyaviridae. Indeed, at the ICTV meeting during the Eighth International Congress of Virology in Berlin (1990), TSWV was accepted as the first member of a newly created genus, tospovirus, within the Bunyaviridae.

The TSWV L RNA segment is 8897 nucleotides long, which is significantly longer than the L RNAs of Bunyamwera (6875 nucleotides) and Hantaan viruses (6530 nucleotides) (Elliott, 1989; Schmaljohn, 1990). Additional domains may be present in the gene product of TSWV L RNA, which may reflect adaptation of this bunyavirus to plants.

TSWV L RNA contains complementary ends of 62 to 66 nucleotides in length. Hence, the RNA can be folded into a stable panhandle structure (Fig. 3), which may be involved in the appearance of circular nucleocapsids in virus particles (Peters et al., 1991), as also found for the Bunyaviridae (Raju & Kolakofsky, 1989). Moreover, these terminal sequences will play an important role in genome transcription and replication, since they contain the initiation signals for encapsidation and RNA synthesis (Krug et al., 1989; Parvin et al., 1989). An alignment of the 3'-terminal sequences of the RNAs from segmented negative-strand viruses is shown in Fig. 4. On the basis of terminal nucleotide sequence homology, the animal bunyaviruses can be clustered into three groups, the nairoviruses (Clerx-van Haaster et al., 1982), the uukun-/phleboviruses (Ihara et al., 1984, 1985; Ronnholm & Pettersson, 1987; Simons et al., 1990) and the hanta-/bunyaviruses (Schmaljohn et al., 1986, 1987; Clerx-van Haaster et al., 1982). Members of the Arenaviridae and Orthomyxoviridae have their own distinct terminal sequences (Fig. 4; Desselberger et al., 1980; Auperin et al., 1982). Strikingly, the termini of the TSWV RNAs show considerable sequence homology to that of RNA segment 3 of Thogoto virus, a member of the Orthomyxoviridae, which might reflect ancestral relationships between both virus families.

TSWV L RNA contains a single ORF in the viral complementary sense, corresponding to a protein with a predicted M, of 331.5K. Analysis of viral RNA species in infected plant cells indicates that this ORF is expressed by the formation of an mRNA of approximately genome length. No subgenomic RNA species derived from the L RNA could be detected (unpublished results). Remarkably, in several TSWV isolates, defective L RNA species appear when maintained under laboratory conditions. In line BR-01, which has been used for sequence determination of the L RNA, a deleted form of this RNA segment accumulates, which is approximately the size of the M RNA. The genesis and implications of these defective RNA molecules for virus multiplication are currently under investigation.

The predicted 331.5K protein encoded by TSWV L RNA most probably corresponds to the viral polymerase. Comparisons of (putative) RNA polymerases from TSWV, Bunyamwera, Hantaan and influenza A viruses reveal the presence of amino acid sequence motifs that are present in all polymerases showing RNA template
specificity and most likely form the active sites for RNA synthesis (Poch et al., 1989). The region in the predicted TSWV L protein, surrounding these 'polymerase' motifs, shows considerable sequence homology (approximately 27% identity) to the putative polymerase of Bunyamwera virus, but to a much lesser extent to that of Hantaan virus. All three L proteins in their turn share conserved amino acid motifs, in a stretch of 200 to 250 residues, with the PB1 polymerase subunit of influenza viruses (Fig. 7). These findings further underline the importance of these common signature sequences and, moreover, justify the assumption that TSWV L RNA indeed encodes the viral polymerase. Strikingly, on the basis of amino acid homology, TSWV is more closely related to Bunyamwera virus than Hantaan virus is to this prototype bunyavirus. It may be anticipated that the amino acid homology between TSWV L protein and those of phlebo- and uukuviruses is even higher, since these viruses are even more closely related to TSWV, sharing similarly organized ambisense S RNA segments (de Haan et al., 1990; Giorgi et al., 1991).

The data presented furthermore imply that, based on molecular properties, such as terminal sequences, and based on the exclusive host range and mode of transmission, TSWV is indeed a member of a new distinct genus (tospovirus) within the Bunyaviridae. The authors wish to thank Jan Gielen and Mart van Grinsven of Zaaunie Research, Enkhuizen, The Netherlands, for providing the synthetic oligonucleotides and for helpful discussions.

References


(Received 19 March 1991; Accepted 14 May 1991)