

The Complete Nucleotide Sequence of the S RNA of a New *Tospovirus* Species, Representing Serogroup IV

Cornelia Heinze, E. Maiss, G. Adam, and R. Casper

First and fourth authors: Federal Research Center for Agriculture and Forestry, Institute for Biochemistry and Plant Virology, Braunschweig, Germany; second author: University of Hannover, Institute of Phytopathology and Plant Protection, Hannover, Germany; third author: University of Hamburg, Institute of Applied Botany, Department of Plant Protection, Hamburg, Germany.

Requests for reprints should be sent to the third author: DSM - Plant Virus Div. c/o BBA Messeweg 11/12, D-38104 Braunschweig, Germany.

The nucleotide sequence data reported in this paper will appear in the DDJB, EMBL, and GenBank databases under the accession number Z 46419.

The presented work is part of the C. Heinze's Ph.D. thesis.

We thank V. Bicknäse and R. Nieländer for careful technical assistance. We thank S. K. Green, AVRDC, for providing the virus isolate. C.

Heinze thanks PLANTA for financial support.

Accepted for publication 7 February 1995.

ABSTRACT

Heinze, C., Maiss, E., Adam, G., and Casper, R. 1995. The complete nucleotide sequence of the S RNA of a new *Tospovirus* species, representing serogroup IV. *Phytopathology* 85:683-690.

A tospovirus isolate Tospo-To, recently shown to be serologically related to some Taiwanese, Japanese, and Indian tospovirus isolates, was purified and its S RNA segment was cloned and sequenced. The complete nucleotide sequence of the S RNA, 3,536 nucleotides, contained inverted terminal repeats and two large open reading frames in ambisense orientation, separated by an intergenic region that is significantly larger than those of hitherto known tospoviruses. Amino acid sequence comparisons revealed 16 and 33% identities with the N_S and N protein

of tomato spotted wilt tospovirus for the viral sense and virus-complementary reading frames, respectively. In contrast to other tospoviruses, additional putative open reading frames besides those for N and N_S have been detected in the nucleic acid sequence of this new tospovirus isolate. The sequence comparisons for N protein amino acid sequences of several different tospovirus species and isolates indicated Tospo-To should be regarded as a new species in the genus *Tospovirus* as has already been suggested from serological data.

Additional keywords: INSV, nucleoprotein purification, sequence homologies, TSWV.

The genus *Tospovirus* of the family *Bunyaviridae* contains the only plant viruses that are transmitted in a circulative propagative transmission mode by thrips (29,31). Four different species have hitherto been recognized as members or tentative members of this genus (5): tomato spotted wilt (TSWV), impatiens necrotic spot (INSV), groundnut ringspot (GRSV), and tomato chlorotic spot virus (TCSV). These four viruses can clearly be differentiated based on serological cross-reactions of their N proteins (3,5,20) as well as based on amino acid sequence homology of the N proteins (4,10,21). Serological and electron-microscopy results have been published previously indicating the existence of even more tospovirus species (2,15,16,18,24,25,33); however, in some of these cases the available data are incomplete to warrant the assignment to a known species or the creation of a separate new species. The best-characterized tospovirus candidates among the above-mentioned isolates have been designated as belonging to groundnut bud necrosis virus (GBNV) (13) and watermelon silver mottle virus (WSMV) (32,33). These isolates were compared serologically (2,24,27,33), and of these, the new isolate from tomato, Tospo-To, as well as the isolates from watermelon in Taiwan and Japan and the groundnut isolate from India all have been shown to possess larger N proteins than other tospovirus species (2). In addition, it was clearly shown that the N proteins from these isolates are serologically related to each other and do

not react with antibodies specific to other tospovirus N proteins. Because of this close relationship the designation "GBNV" was suggested as the species name for all isolates that share a larger N protein and serological properties (2). Recently, the watermelon isolate from Taiwan, Tospo-W, was named watermelon silver mottle virus and was proposed as a new member in the genus *Tospovirus* because of the low level of homology that its N gene shares with those of other tospoviruses (32). In this paper we report the complete nucleotide sequence of the S RNA segment of the tomato isolate from Taiwan, Tospo-To.

MATERIALS AND METHODS

Virus source, propagation and purification. The virus isolate Tospo-To (DSM No. PV-0283) was supplied in infected tomato leaves by S. K. Green (AVRDC Taiwan). It was transferred to *Nicotiana benthamiana* Domin. by mechanical inoculation (2). Plants inoculated at the four-leaf stage developed necrotic lesions as primary symptoms after 2 weeks, subsequently followed by a mosaic and mottling of the newly formed leaves that rapidly developed into wilting and finally death of the plants. Inoculation of the isolate onto *Cucumis sativus* 'Riesenschäl' led to systemic infection.

Virions were purified from systemically infected *N. benthamiana* leaves harvested 2 to 3 days after symptom appearance. The purification of nucleoprotein followed the protocol of de Avila et al. (6) with either sucrose gradient centrifugation (6) or cesium sulfate gradient centrifugation as the final step (20). Purified nucleocapsids were either stored at -80°C or used immediately for

Corresponding author: G. Adam
E-mail: BT5A006@RRZ-CIP-1.RRZ.UNI-HAMBURG.DE

the extraction of RNA as described by Maiss et al. (22). Total RNA was extracted from infected *N. benthamiana* tissue according to the method of Verword et al. (30).

cDNA synthesis and cloning methods. A cDNA library of viral RNA isolated from purified unfractionated nucleoproteins was constructed. First-strand cDNA was synthesized by random priming with hexamer oligonucleotides and AMV (avian myeloblastosis virus)-reverse transcriptase according to Gubler and Hoffman (14). After second-strand synthesis, the resulting dsDNA was blunt-end cloned into the *Hinc*II-cut and dephosphorylated phagemid pT7T3-19U (Pharmacia, Freiberg, FRG).

Sequencing methods and strategies. Sequencing of cloned cDNA inserts was performed with single-stranded (ss) DNA or plasmid DNA with the chain termination technique (26), using the T7-sequencing kit (Pharmacia) or the TAQ cycle sequencing kit (USB), respectively.

The primers that were used for the determination of the 5' and 3' ends as well as for the verification of the intergenic region are listed in Table 1 and their positions indicated in Figure 1.

For the determination of the 3' and 5' ends of the S RNA we have used two different strategies. For 3' ends viral sense RNA from purified nucleoprotein was polyadenylated at the 3' ends according to the procedure of Carrington and Morris (7), and subsequently first-strand DNA synthesis was performed with the primer P8 (Table 1) using AMV-reverse transcriptase. After RNA digestion the first-strand DNA was amplified by polymerase chain reaction (PCR) using a pair of primers; the viral-sense primer P7 and P8. The resulting PCR product was cloned into

*Eco*RV-cut pBluescript SK- (Stratagene, La Jolla, Calif.) that was prepared for PCR-fragment cloning according to the method of Holton and Graham (17).

For the 5' ends we used total RNA from infected plants for first-strand cDNA synthesis with the antiviral-sense primer P2 (Table 1). The resulting cDNA was C-tailed at the 3' end with terminal transferase (Boehringer Mannheim, FRG) according to the manufacturer's instructions. Finally the C-tailed DNA was used for PCR with the antisense primer and the poly-G primer P1 (Table 1), and the PCR product was cloned into pBluescript SK- as described for the 3' ends.

Computer analyses of sequence information. Computer analyses of the resulting nucleotide sequences were performed with the program package GCG (11) on a VAX computer or with PC-GENE on a PC. Primers were designed using the program OLIGO (Nat. Biosciences, USA)

Expression of the N and Ns, ORFs in bacteria. In order to verify that the virus-complementary (v-c) ORF (open reading frame) codes for the N protein and that the sense-ORF codes for an Ns_s protein that is smaller than the tomato spotted wilt virus Ns_s protein, we have subcloned the corresponding coding sequences of the S RNA into the expression vectors pEMSK (see Results) and pQE 32 (Diagen, BRD), respectively.

Gel electrophoresis and immunoblot assays. Polyacrylamide gel electrophoresis (PAGE) and subsequent Western blot (immunoblot) assays were done as described by Adam et al. (1) using an antiserum prepared against purified N protein (DSM No. AS-0118) and goat anti-rabbit IgG labeled with alkaline phosphatase (Sigma Chemical Co., St. Louis, Mo.).

TABLE 1. Polymerase chain reaction primer sequences

No. ^a	Nucleotide Sequence from 5' to 3'	Position ^b
P1	5'-CGGAATTCCTGGG ₍₁₃₎ -3'	Bordering 5'-end
P2	5'-AACTTCGGCTTCATGGATTTGAC-3'	278-255
P3	5'-CCGGGAGAATCCTGAATTAC-3'	1215-1234
P4	5'-CCTGTAAGAATAATGGGATC-3'	1883-1902
P5	5'-CAACTGGTCATAGCATGTAG-3'	2076-2056
P6	5'-TGCTAAGCTTCTGTGGATG-3'	2719-2700
P7	5'-TGAATGTCCAGTCTCAGGT-3'	3192-3211
P8	5'-CGGGGATCTCGAGAAGCT ₍₁₇₎ -3'	Bordering 3'-end

^a Numbering refers to Figure 1.

^b Coordinates are given according to the numbering in Figure 2.

RESULTS

Sequence and genomic organization of the S RNA. Eleven clones, representing the S RNA segment but lacking the complete 3' and 5' ends were sequenced in both directions. The relative positions and orientations of the cloned inserts as well as that of the PCR-primers P1 through P8 are indicated in Figure 1. As expected, one ORF was found in the viral sense orientation starting at position 68 (Figs. 1 and 2), and terminating at position 1,387, thus coding for a protein of 49.7 kDa. A second ORF was positioned between nucleotides (nts) 2,645 and 3,470 (Figs. 1 and 2); however, in the v-c sense, it possessed a coding potential for a

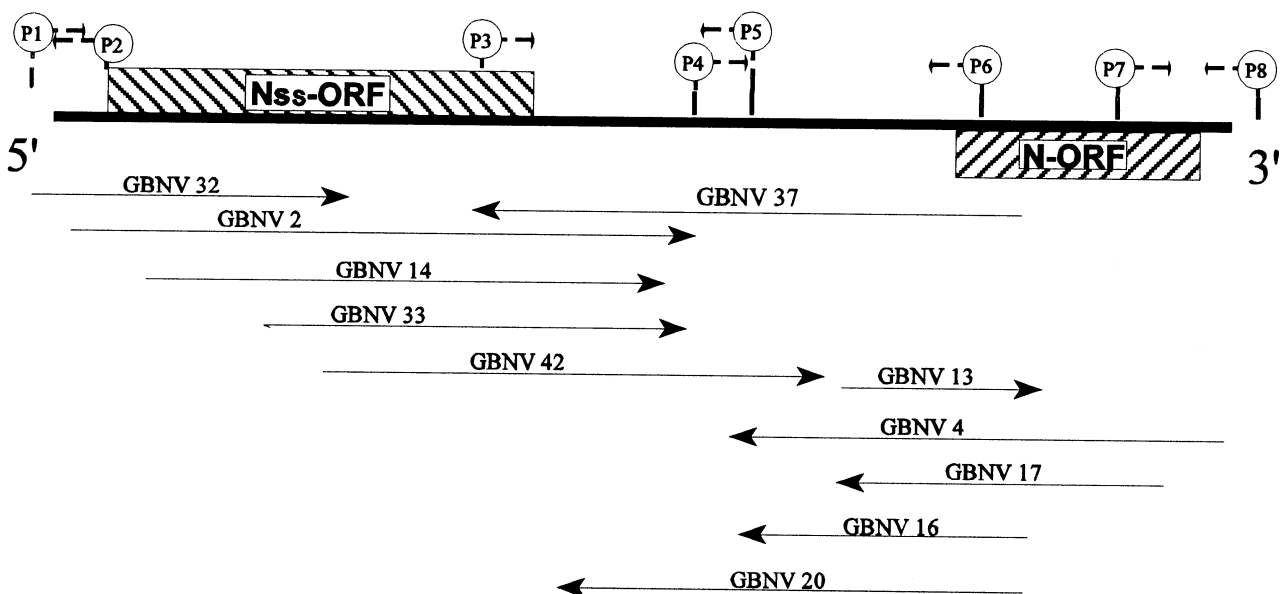


Fig. 1. The position and orientation of polymerase chain reaction (PCR) primers P1 through P8 used for sequencing are indicated above the thick horizontal line representing S RNA. The cDNA clones used for sequencing are shown below the horizontal line; arrows indicate the size and orientation of the cloned insert.

AGAGCAATCGAGGCAATAATAATTCTAGAAAAACAACGTTAGACTCGAGTAATACTATTTTCAGGCATGCTCTACTGCAAAGAATGCTGCTTCGGAATTC
 I K S Y G T R D N R A V N D C Y S V F T G E G V N F L N L F M H N
 ATCAAAAGTTATGGCACAAGAGACAACAGAGCAGTAAATGATTGCTACTCTGATTCACTGGAGAAGGTGTCAACTTTCTCAATTTGTTTCATGCATAATA
 N A G I K S A F S I N D L G R N E D V K I H E A E V V D S C H D Y H
 ATGCAGGTATTAATCTGCATTGAGCATCAACGATTTGGGCGAGGAATGAAGATGTCAAATCCATGAAGCCGAAGTGTGATTGATGCCATGATTATCA
Y F E K F G L D I T F C E H E M N L I V R K P G V K N T G C K F T
 TTTTTTGAAAAATTTGGACTAGACATAACATTTCTGCGAGCATGAAATGAACTTAATTTGTAAGGAAACACAGGTGTGAAGAACACAGGTCGCAAGTTCACA
 M H N Q I F N P N S N I L N M T P G T I S E D A F Y D R S K I T S
 ATGCACAATCAATCTTTAATCCAAATTCATATACTGAACATGACGCCGGAAACAATTTGAGAAGATGCTTTCTATGACAGAAGCAAATAACATCTA
 S G L L P S G W C L D E C Y K N N F Y I A T N G D L T L D Y G F S V
 GTGGACTACTTCCTCAGGATGGTGTGGATGAGTCTATAAGAATAATTTCTATATAGCTACCAATGGAGATCTTACCCTTGATTATGGGTTTTCTGT
M G K T T S Y W R E S M S R E K I L S V K Q K C F S D N T V P T N
 AATGGGAAAAACAACATCTTACTGGAGAGAGAGCATGTCAAGGGAAAAGATCTTGTGAGTAAAGCAAATAATGCTTTCTGACAACACTGTCCCTACCAAC
R L L S T S T V K G I Q I G S E L A S D T T V I L S C K Q N L N I
 AGATTGCTATCTACATCAACAGTCAAGGGCATCCAAATTTGGTTCTGAACAGCTTCTGATACCACAGTTATATATCATGCAAGCAAATTTAAACATG
 D L K S Q Y R I S F H G I Q E E G A F A R T F C V P F E N K S R M I
 ATTTGAAATCCCAATATCGTATTTTCATTCATGGTATTGAGAAAGAGGGCTTTGCGAAGGACTTCTGCGTCCCTTTGAAAATAAATCTAGAAATG
S F Y A K T V A D N S N E R T T L I K V V T K T V D S H L V I P
 CAGTTTTACGCAAAAATGTTGCTGACAACAGTAATGAGAGAACCCCTTATCATAAAGGTGGTTACAAGAGACTGTTGATTCCTCCTGTGCTTCTCT
 I R N H I N C E K K I G A R V G L V D F C D S D P N Y N Q L I V K
 ATAAGAAATCATATCAATTTGTAAGAAAAGAAATCGGAGCAAGAGTAGGCTTGTGATTCTGTGATAGTATCCCAATACAATCAATGATTGTCAAGA
 N L L S V H T Q F A I N L S E V I Q K P V I V F K M Y D K E L H N D
 ATCTGTTAAGTGTCATACTCAGTTGCTATCAATTTATCTGAAGTATTCAAAGCCAGTCATAGTGTCAAATGTATGATAAGAGCTCCCAATGA
Y T E V S G R I L N Y Q V D S E R N V Y F L S K T L E V L P K S L
 TTATACAGAAGTTCCGGGAGAATCCTGAATTACCAGGTAGATTGAGAACGGAATGTTATTTCTCTCAAAGACTCTTGAGGTTCTGCCTAAATCACTA

S T L S Y L G S I A P I Q W K E C L E H Q H F V V K A E *
 TCGACCCTGAGTTACTTAGGCAGCATTGCACCAATTCATGGAAGAGTGTGGAAACACCAGCACTTGTGTGAAAGCAGAATAAACTATTTGAAAC
 TCTTATGCTATATAATCTATAAAATAGTATATCGCATGTTGATGCTGCATTGTGCAATTTCAATTAATAAATAATTTAGTTATAAAATCCAAGTA
 TTTGATGCTTAGGTATTTTAAAGTTAAGTTAATAAAGTAAGTTTACTACTAGAATTGAATAAATAAATAGATAATAAATATATAAGTAAATGA
 AGCTAAAAACCAAAATCATAAAATTAAGTAAATAAATTTAGTCATGAATCTGAATGTTGTGCTGAGTGTAAATTAAGTCAAGTTAGTAGAT
 GAGTCTGTTGTGTCGATTAAGTAAATAACATATAAATAAGTAAATTAATAAATAAAGTAAATTAATAAATAAATAAATAAATAAATAAATAAATAA
 ATAAAAATGAAATAACAAAAAACCAAAAAGAGCTCGAAAGGGGCAATTTGGCTTCAATACCCTGAGCAATTTCAATTCACCTGTAAGAATAATGGGA
 TCATTTGGATCATGTGTTTTGTTATGATCATGATGAACAAGAGTTGCTTGAATTAATATAGTTGGAATCTTTAGGTCAATCGTGTAAGTTGCAA
 TGAAGGATCATTATGTTAGTTAGTTTACCCTTATCACATGATTAATACAAAATCTAACATGCTATGACCAGTGTGATCATGATATAGAATTCATGTAA
 ATTATATATCTGATTGATTTCGAGCAGCTTAGATCTGCATCTAATCTAAGATTCTTATCATGATTTGATCCTTTTATCTAACTATTTATGTTTCAA
 CCAAATCAATCTAAATCATTAAAACTTCTAATCTCAACAATTTGAAAGCTAAGTTTGTCTTACGATTGATCATTACAGTTAAATAAATAAATCATATTT
 ACAGAAAATTTGGCCATTGATTGCTCCGAAGAGGTCTTTTTGTTTTATTTTTGATTTTTTTGGTTTTTTGGAATTTTTATATATATATATAT
 TGTATTTATTTGTTTTAT
 TACTCGTATAATTAATAAGTATATAGTTATATAAGCACACAAAAAACTTAAACAACATAAATAGATTTATTAATATATAGTTAAGCATAAACA
 CAAACACACATGTGCACATATATGCACAGTTGACACTTCAAAGTTCACTTCAAAGAAGTCTAGGCTTTGAACTTTCTTTGAAATGATCACCAATC
V E L S T S P K S S K K S H D G Y D
 ATCCACAGAAAGGTTAGCACTGAAGACCTGTCCATGTAATTAATCTGATCCTCATATTTCTGAGCGAAATAGAGCTAGCTGTTCCAGGGTTACTCTCA
 D V S L N A S F V R D M Y K V Q D E Y K K L S I S S A T G P N S E
 CTGAGTATTTGACAGCCTGTTGAAACATAACATCCAAATCATCTTAAATCAACCTGTGAAGCAGAAAGAACCTTAGCCACCTTACAACCTTGCTCAT
 S L I K V A Q K F M V D L D D K F E V Q S A S L V K A A C K C V Q E Y
 AGGTGGAGAGTCTTTTATACTAACTTTTCTTTATATTTTGGTAAATAGCTAAAGGGAAATAATCGGTGCCAGTCCCTTATACTAGATAAAG
 T S F N K I G L K E K K I N Q Y Y A L P F I I P A L G R I S S L L
 AGGTAGAGGTCCTCAATACATAACATCATCTCAAGGCACAAGAAATCATAGATGCGGGCACATTCAACCCATAAGCAGCAACTAATGGTAGTTCCATG
 P L P G G I C L M M R L A C S D Y S A P V N L G Y A A V L P L E M
 ATTTTCCCATACATTTCTGTCTAGCAGCTTCACTTCTGCTTTTCTCCACCATACTAATTTTGTCTGATAAAAGCTTCTGCTCTTGAATGTCC
 I K G Y M E Q R A A E N K S K E V M S V I K T R I F A E T R K F T W
 AGTCTCAGGTCGACGCTCAACACTGTAGCAACAATATTTTACCAGAAAATTAATCTTACCCTCTTGCAGGCTGCAAAAATCTGTTTCTGCATTT
 D E P G V D V S T A V I N K G C F N Y K G S K C A A F I Q K R C K
 TAGGATATCAAGCAGTTGTGAAAGTTATCTATATTTTGTGTTGTGCAAAAAAGCTTTGAAGCTAAATCCGGGAGTGGAGTCTCTGTTTCAAT
 L I N L C N T F T I E I N K N N D F F A K F S F G P T S D E T E I
 TCAACATCTGCTGTACCACAGCCAAAAGTTCCTTGATTTTCTTTCTGTGAGCTGCTTAACTGTAGACATGGTGTTTACAAGGAAACGCTGATATGT
 E V D A T G G A L L E K I K K E T L Q K V N S M
 TGTTTGAAGTATATTTAGCGCTCGATTGCTCT

Fig. 2. The complete sequence of the GBNV S RNA segment and encoded proteins in DNA code and viral sense is given as determined from the cDNA library and additional cloning and sequencing of the terminal ends. The complementary 14 nucleotides at the 3' and 5' ends are underlined. For the two open reading frames of the N_s and N protein the deduced amino acid sequences are shown above or under the nucleic acid sequence, respectively. The start codons are underlined.

protein of 30.6 kDa. Despite the similarity to TSWV in genome organization, sequence comparisons of the computer-translated proteins of Tospo-To (Fig. 2) with those from TSWV (22) indicated no significant homologies with the N_s and N proteins for the viral sense and v-c ORFs, respectively. Nevertheless we assumed that the sense ORF represents the N_s , and the v-c ORF the N protein. Since no inverted repeats, characteristic for tospovirus RNAs, were present at the ends of the sequence determined from the cDNA library, we selected usable primer sites for cloning and sequencing the ultimate 3' and 5' ends (Table 1 and Fig. 1).

Using the strategies described in Materials and Methods, the exact ends of the S RNA were determined. When determining the 5'-end sequence from total RNA preparations from infected plants, several independent clones were obtained, some of which contained additional nucleotides that extended beyond the terminal repeat. This is probably due to the presence of a subgenomic mRNA for the putative N_s -protein to which the primer for first-strand synthesis also anneals and which contains nonviral sequences at its 5' end (19). The length of these nonviral sequences varied between different clones from 33 to 64 nucleotides.

Presence of other ORFs besides those for the putative N and N_s protein. Analyses of the complete sequence in all possible orientations with the program FICKETT (PC-GENE), revealed five additional ORFs besides those assumed to represent N and N_s . Three were detected on viral sense inside the N_s ORF: ORF₁ from position 84 to position 320; ORF₂ from position 111 to position 320; and ORF₃ from position 3,097 to position 3,535. Two additional ORFs were found on v-c sense, one of which,

ORF₄, was nested inside the N ORF starting from position 2,898 to position 2,581, whereas ORF₅ was located between positions 264 and 64. No ORF was detected inside the intergenic region.

Expression of the putative N protein ORF. The ORF on the v-c S RNA was subcloned into the expression vector pEMSK under the control of the $P_{(R)}$ -promotor. This expression vector is a derivative of the pEX2 vector (28) constructed by Maiss (*unpublished*). It contains T3 and T7 promoters, and the multiple cloning site from pBluescript II (Stratagene). The vector-encoded fusion protein, a truncated galactosidase, has a molecular weight of 42,000, which together with the inserted N-ORF should result in an expression product of 72,000.

After transformation of the resulting clone into *Escherichia coli* strain POP 2136 and induction of expression at 42°C, sodium dodecyl sulfate (SDS)-PAGE of the resulting protein extract from lysed bacterial cells revealed a prominent protein with a molecular weight of 72,000 (Fig. 3A), which is in good accordance with the expected size of the expression product. After transfer to nitrocellulose and subsequent treatment for immunodetection with N-specific antiserum the same band stained intensively (Fig. 3B), which further confirmed the identity of the translational product of this ORF as the N protein. The molecular weight of 30,600 for the computer-predicted translational product fits well with the observed mobility of purified N protein in SDS-PAGE, where a molecular weight of 32,000 was determined (2).

Expression of the putative N_s protein in bacteria. As already indicated above, the presumed ORF for the N_s protein was predicted to code for a protein of a smaller size than that of other

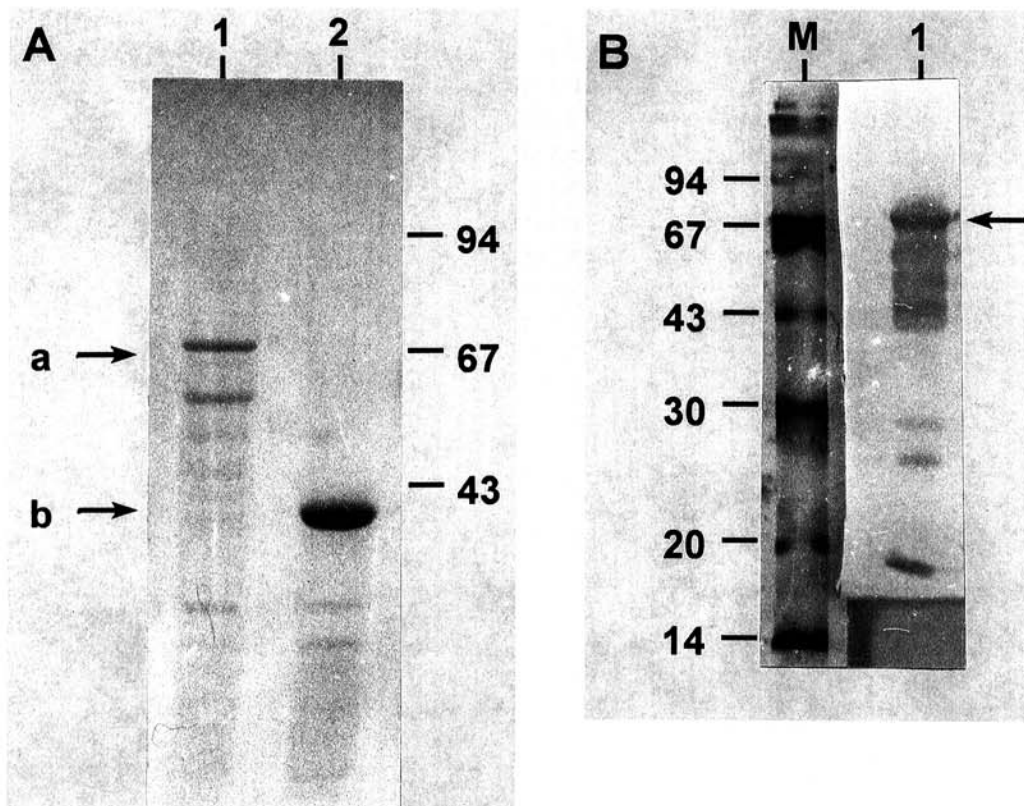


Fig. 3. Western blot (immunoblot) assay of the Tospo-To N protein after expression in *Escherichia coli*. The Tospo-To N protein in extracts from transformed cells was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis using 12.5% polyacrylamide gels along with biotinylated marker proteins (Pharmacia, LMW). One part of the gel was stained with Coomassie Blue (A) whereas the second part was transferred electrophoretically to nitrocellulose, treated with Tospo-To N-specific polyclonal antibodies (B) and subsequently with goat anti-rabbit alkaline phosphatase conjugate before staining with AS-MX phosphate, Fast Red. The positions and mol wt of the marker proteins are indicated on the right (A) and left (B) side, respectively. A, lane 1 extract from cells transformed with the expression vector plus Tospo-To N insert; lane 2, expression vector without insert. (a) depicts the N-containing fusion protein, (b) the expressed part of the truncated β -galactosidase gene without insert. B, Lane M contains the biotinylated marker proteins after staining with streptavidin alkaline phosphatase complexes; lane 1, immunostaining of the expressed N-containing fusion protein. The arrowhead depicts the position of the N-specific band, equivalent to (a) in A.

sequenced tospoviruses. To verify that the N_{s_s} protein of the Tospo-To strain is indeed smaller in size, the respective ORF as well as the corresponding sequence of the TSWV-L3 isolate were subcloned into the expression vector pQE 32 and 31, respectively, and transformed into *E. coli* M 15. After induction with isopropyl- β -D-thiogalactopyranoside, analysis of the expressed proteins revealed a newly synthesized protein band of 48 kDa for Tospo-To and 57 kDa for the TSWV-L3 N_{s_s} protein (Fig. 4). Since the expression vectors contained only the information for 10 additional amino acids besides the N_{s_s} ORFs, these values are in agreement with the predicted molecular weight values of 49,700 or 52,200 from the sequence data and thus confirm the reduced size of the N_{s_s} protein compared with other tospoviruses.

Homologies of the N and N_{s_s} proteins with other known tospoviruses. Comparisons of the amino acid sequences of several N and N_{s_s} proteins from different tospovirus isolates were made and the values for identities obtained are summarized in Table 2,

above diagonal. Regardless of the species, the identities for the N proteins were equal to or below 33% for the new sequence, indicating an even larger phylogenetic difference than that between INSV and TSWV, where identities between 50 and 60% were obtained (Table 2, above diagonal). In contrast to this, the comparison with an N protein sequence of the watermelon silver mottle virus that became available recently (32) revealed 99% identity, indicating a very close relationship.

For N_{s_s} protein only four additional sequences were available for comparison. The identities of their amino acid sequences are given below diagonal in Table 2: identical amino acids between Tospo-To and TSWV or INSV isolates were less than 20% whereas values between 50 and 84% identity were determined between TSWV and INSV (Table 2, below diagonal). However, when aligning the different available amino acid sequences for N_{s_s} proteins (Fig. 5) a remarkably conserved region of 24 amino acids at the C-terminus became evident (positions 398 to 421).

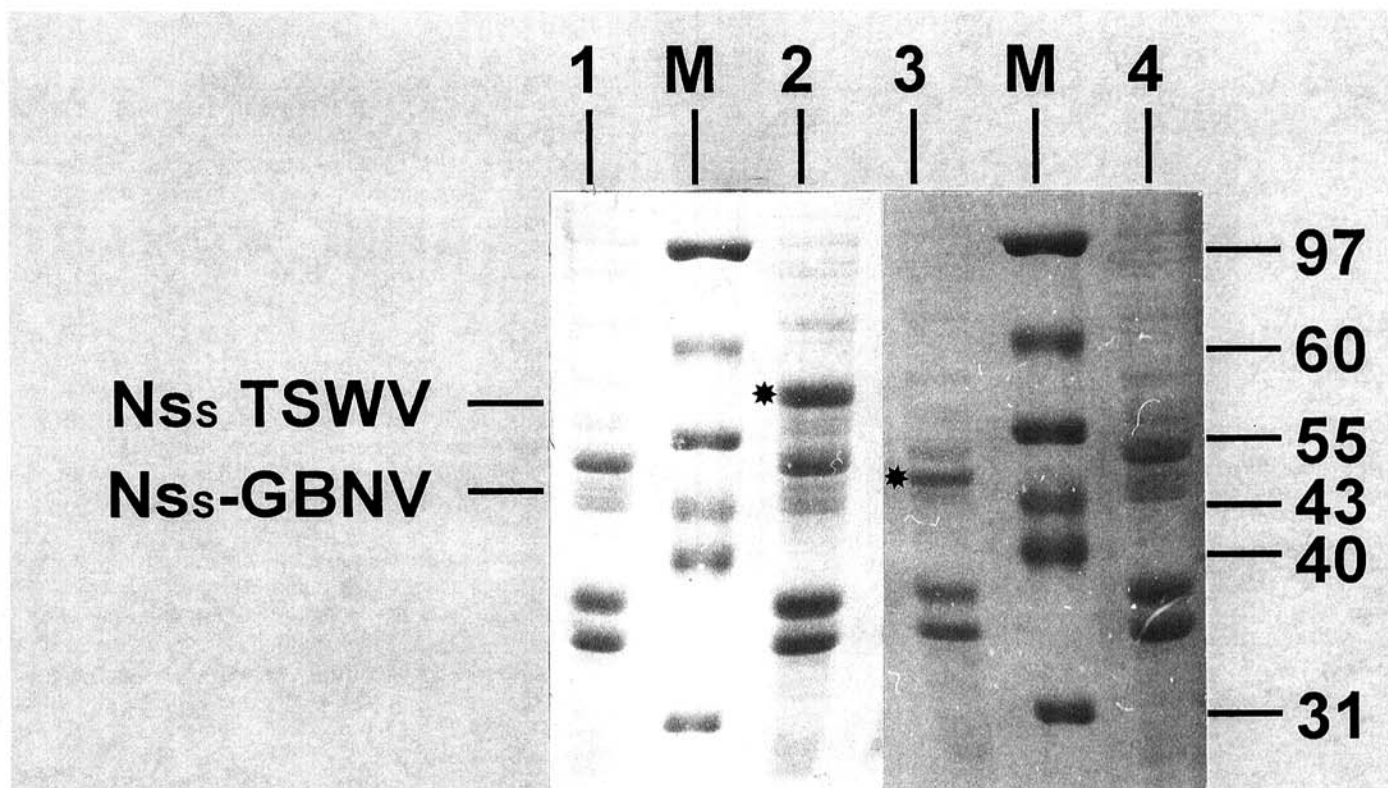


Fig. 4. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the Tospo-To and TSWV N_{s_s} proteins after expression in *Escherichia coli*. Protein extracts from transformed cells were analyzed by SDS-PAGE on 12.5% polyacrylamide gels along with molecular weight marker proteins, Promega (Madison, WI), (lane M). Proteins were stained with Commassie Blue. Lane 1, pQE 31 without insert; Lane 2, pQE 31 plus TSWV-L3 N_{s_s} protein; Lane 3, pQE 32 with Tospo-To N_{s_s} protein; lane 4, pQE 32 without insert. The asterisks in lanes 2 and 3 mark the expressed N_{s_s} proteins of TSWV and Tospo-To, respectively.

TABLE 2. Percentage of identities in amino acid sequences of N and N_{s_s} proteins from different *Tospovirus* isolates^a

	Tospo-To	WSMV	TSWV	TSWV-L3	TCSV	GRSV	TSWV-B	INSV-NL
Tospo-To ^b	100	99	32	33	26	33	32	30
WSMV		100	33	33	26	33	32	30
TSWV	16		100	98	76	77	79	54
TSWV-L3	16		92	100	78	79	80	55
TCSV	100	81	84	53
GRSV	100	95	54
TSWV-B	8		75	82	100	57
INSV-NL	19		50	55	100

^a The percentage of identical amino acids for the N (above diagonal) and N_{s_s} proteins (below diagonal) as deduced from the nucleotide sequences were determined with the program Fasta (obtained from W. A. Pearson, University of Virginia).

^b Virus names (references for sequencing data in parentheses): Tospo-To (this study); WSMV (32) = watermelon silver mottle virus; TSWV (10) = tomato spotted wilt, isolate B1; TSWV-L3 (22) = tomato spotted wilt, isolate L3; TCSV (4) = tomato chlorotic spot virus; GRSV (4) = groundnut ringspot virus; TSWV-B (23) = tomato spotted wilt virus isolate B; INSV-NL (8) = impatiens necrotic spot virus, isolate NL.

```

TSWV-BR01 SEDAYKRKYFLSKTLECLPSNTQTMSYLDSIQIPSWKIDFARGEIKISPQ
TSWV-L3   SEDICKRKYFLSKTLECLPSNTQTMSYLDSIQIPSWKIDFARGEIKISPQ
TSWV-B    SEDASKKKYFLSKTLECLPVNVQTMSYLDSIQIPSWKIDFARGEIRISPQ
INSV-NL07 SEDGYGRKYFLSQTLKSLPRNSQTMSYLDSIQMPDWKFDYAAGEIKISPR
GBNV      QVDSERNVYFLSKTLEVLPLKSLSTLSYLGSIAPIQWKECLEHQHFVVKAE
          * . . * * * * . * * . * . * * * . * * . . . . .

```

Fig. 5. Multiple alignment of N_s proteins from different tospoviruses was done with PC-GENE, using five published sequences (reference in parenthesis): TSWV-BR01 (10); TSWV-L3 (2); TSWV-B (23); INSV-NL07 (8); Tospo-To (*this manuscript*). The numbering for start and end of the alignment refers to the Tospo-To sequence. An asterisk below the sequence marks consensus amino acids, a dot indicates highly conserved amino acids.

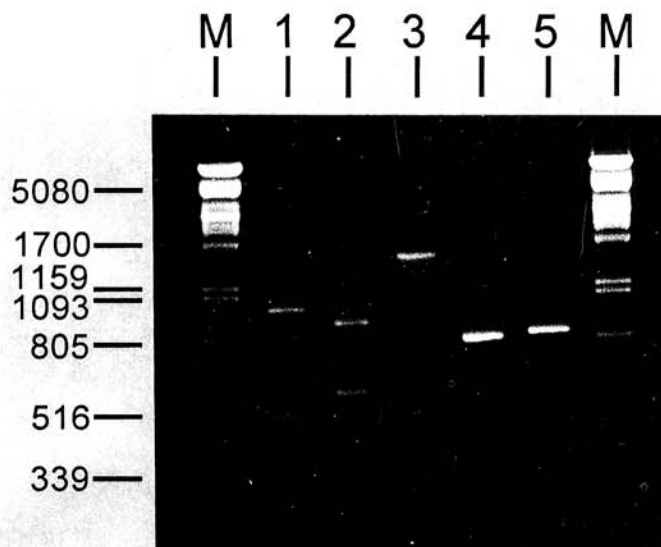


Fig. 6. Polymerase chain reaction products amplified from the intergenic region of Tospo-To were separated on 1.2% agarose gels, stained with ethidium bromide and visualized with UV-light. Lane M contains *Pst*I-cut λ DNA. Lanes 1 through 3, 1,500-bp fragment produced with primers P3 and P6; digested with either *Dra*I (lane 1) or *Bgl*III (lane 2), or undigested (lane 3). Lane 4, 840-bp fragment produced with primers P3 and P5; lane 5, 860 bp fragment produced with primers P4 and P6.

```

TSWV      3'- U C U C G U U A G C A C A G U
TSWV-B    3'- U C U C G U U A A C A C A G
GBNV      3'- U C U C G U U A G C U C C G
INSV      3'- U C U C G U U A A C A C A G

TSWV      5'- A G A G C A A U U G U G U C A
TSWV-B    5'- A G A G C A A U U G G G U C
GBNV      5'- A G A G C A A U C G A G G C
          * * * * * * * * * * *

```

Fig. 7. Alignment of the ultimate 14 3' and 5' nucleotides of S RNAs from several tospoviruses: (source of sequence information in parenthesis): TSWV (9); TSWV-B (23); INSV (21). An asterisk denotes identical nucleotides.

This corroborated our assumption that the viral-sense ORF represents a protein homologous to N_s proteins of other tospoviruses.

Verification of the length of the intergenic region. In order to verify the significantly larger intergenic region of the new isolate compared with that of other tospoviruses, we have selected two primers flanking, as well as two primers located within, this region (Fig. 1). Using purified total RNA from infected and healthy plants as template for first-strand synthesis and subsequent amplification by PCR we obtained three different PCR products: i) 1,500 bp (primer 3 + 6); ii) 860 bp (primer 3 + 5); and iii) 840 bp (primer 4 + 6) (Fig. 6). The size of all PCR products corre-

sponded to the predicted size based on the location of their primers on the S RNA, i.e., 1,505, 862, and 837 bp. These products were only obtained with RNA from infected plants. In addition we have cloned and subsequently sequenced the 1,500-bp fragment in both orientations. The resulting sequences matched exactly the sequence data in Figure 2. Digestion of the 1,500-bp PCR product with *Dra*I or *Bgl*III yielded two fragments as a result of cleavage at positions 2,222 and 2,134, respectively (Fig. 6). Thus it has been confirmed by several methods that the sequence determined for the S RNA in the intergenic region was correct and that the intergenic region is indeed larger than that from other sequenced tospoviruses.

DISCUSSION

The new tospovirus isolate, Tospo-To, that has been already separated from other known tospoviruses based on serological data (2,24,33), has to be considered as a new member of the genus *Tospovirus* based on the sequence data of the two ORFs of the S RNA. Our data not only substantiated that the recently reported larger N proteins for such type of isolates are actually due to a larger coding capacity of the N protein cistron (2,24,33), but also revealed that these isolates have significantly different N_s proteins. These differences are reflected in the drastically reduced percentage of identities that we found when comparing amino acid sequence data for the two main ORFs present on the S RNA segment with those of all known tospoviruses. Actually, the percentage of identities calculated for N as well as N_s proteins for Tospo-To are the lowest reported so far. Thus, criteria introduced by de Avila et al. (4,5) for the discrimination of tospovirus isolates are realized.

The genome organization that we found for the S RNA of Tospo-To is similar to all other tospoviruses. The S RNA contains all elements that have been described for other tospovirus S RNAs including 3' and 5' nontranslated regions that contain at their extreme ends inverted repeats of 14 highly conserved nucleotides, the first eight of which are identical to those of TSWV and INSV (Fig. 7), and two large ORFs in ambisense orientation separated by an extremely long intergenic region that can form a complex structure due to its high content of adenosine and thymidine.

A multiple alignment for all known N protein amino acid sequences revealed very little conservation. The statement of conserved positions of methionine residues (4) did not hold true when the new sequence was included. No particular region could be identified as highly conserved.

A multiple alignment of the four known N_s protein amino acid sequences revealed the presence of only one domain of 24 highly conserved amino acid residues at positions 398 to 421. Although no function has yet been assigned to the N_s protein of tospoviruses, one may postulate that this conserved region represents the most important functional domain. When folding this protein with the GCG program PEPTIDE STRUCTURE and predicting

the hydrophilicity as well as hydrophobicity the conserved area contained no turn, started with a hydrophilic region, and ended in a hydrophobic domain. The conserved region contained no amino acid motif indicative of a specific function when analyzed by PROSITE (PC-Gene).

The intergenic region of our isolate was shown to be significantly larger than those previously reported for other tospoviruses. This explains the increased length of the S RNA. Our experiments to amplify the complete intergenic region by PCR from total RNA of infected plants strongly support the contention that the region is not the resultant of a cloning artifact. Though not as pronounced, differences in the length of the intergenic region have already been reported for two isolates of TSWV (10,22). Besides its ability to form a complex-folded tertiary structure due to its high adenosine and thymidine content, the intergenic regions contain no common sequence motifs. Therefore, it is tempting to speculate that the sequence-dependent structure, but not a common sequence, is responsible for the suggested function of the intergenic region as regulator or terminator of transcription of subgenomic RNAs (10,22).

Our results obtained from the sequencing of the 5' ends of the viral RNA indicate that the Tospo-To also uses cap-snatching as a mechanism of generating subgenomic mRNA as already described for the type strain of TSWV (19).

In addition, and different from other tospovirus S RNAs, we found nested inside the ORFs for the two proteins, N and N_s, five other possible ORFs, one of which, ORF₄, has a high probability for expression even *in vivo*, since an internal initiation from the subgenomic mRNA appears possible when taking into account the surrounding sequence motifs. That such a genome organization is not unusual in the family *Bunyaviridae*, to which the genus *Tospovirus* belongs, is indicated by the report of Elliott and McGregor (12), although these authors also could not provide conclusive evidence for an expression *in vivo*. Whether such additional ORFs have anything to do with the altered host range of isolates belonging to GBNV-like species (18,27,33) remains to be determined in further studies.

Sequence data for another Taiwanese tospovirus isolate from watermelon (32) that recently became available indicated that this isolate is as close to our tomato isolate as the TSWV isolates BR01 and L3 are to each other (Table 2). The serological results published for this isolate (33) as well as the serological comparison published for different other new isolates (2) indicated that a close relationship exists not only between the two Taiwanese isolates and the watermelon isolate from Japan (18), but also among all three isolates and the isolate from groundnut in India designated GBNV (13). The strong serological cross-reactions observed with different antisera among the N proteins as well as the significantly larger size of the N proteins led to the assumption that all these isolates must belong together in one serogroup. The designation "GBNV" was proposed as the name for a new species that encompasses all three isolates (2). This designation, however, may be misleading, especially because the isolates from Japan and Taiwan did not originate from groundnut and because of their unique ability to infect cucurbits systemically, a feature not reported for the Indian isolate (24). However, host range and symptomatology are risky parameters for the classification of tospovirus isolates (5). We therefore emphasized the serological data reported from the comparison of these isolates when proposing GBNV as a new name. The name WSMV proposed by Yeh and Chang (32) for their watermelon isolate from Taiwan may be adopted instead if sequencing data of the Indian GBNV isolate substantiate the phenotypic differences. Until then we propose to group tospovirus isolates with N proteins of a molecular weight higher than 30,000 that show serological cross-reactions with either antisera against the groundnut, the watermelon, or the tomato isolate (2,18,24,33) into a new serogroup IV of the genus *Tospovirus*.

LITERATURE CITED

- Adam, G., Chagas, C. M., and Lesemann, D.-E. 1987. Comparison of three rhabdovirus isolates by two different serological techniques. *J. Phytopathol.* 120:31-43.
- Adam, G., Yeh, S.-D., Reddy, D. V. R., and Green, S. K. 1993. Serological comparison of tospovirus isolates from Taiwan and India with impatiens necrotic spot virus and different tomato spotted wilt virus isolates. *Arch. Virol.* 130:237-250.
- de Avila, A. C., de Haan, P., Kitajima, E. W., Kormelink, R., Resende, R. de O., Goldbach, R. W., and Peters, D. 1992. Characterization of a distinct isolate of tomato spotted wilt virus (TSWV) from *Impatiens* sp. in The Netherlands. *J. Phytopathol.* 134:133-151.
- de Avila, A. C., de Haan, P., Kormelink, R., Resende, R. de O., Goldbach, R. W., and Peters, D. 1993. Classification of tospoviruses based on phylogeny of nucleoprotein gene sequences. *J. Gen. Virol.* 74:153-159.
- de Avila, A. C., de Haan, P., Smeets, M. L. L., Resende, R. de O., Kitajima, E. W., Goldbach, R. W., and Peters, D. 1993. Distinct levels of relationship between tospovirus isolates. *Arch. Virol.* 128:211-227.
- de Avila, A. C., Huguenot, C., Resende, R. de O., Kitajima, E. W., Goldbach, R. W., and Peters, D. 1990. Serological differentiation of 20 isolates of tomato spotted wilt virus. *J. Gen. Virol.* 71:2801-2807.
- Carrington, J. C., and Morris, T. J. 1984. Complementary DNA cloning and analysis of carnation mottle virus RNA. *Virology* 139:22-31.
- de Haan, P., de Avila, A. C., Kormelink, R., Westerbroek, A., Gielen, J. J. L., Peters, D., and Goldbach, R. W. 1992. The nucleotide sequence of the S RNA of impatiens necrotic spot virus, a novel tospovirus. *FEBS Lett.* 306:27-32.
- de Haan, P., Wagemakers, L., Peters, D., and Goldbach, R. W. 1989. Molecular cloning and terminal sequence determination of the S and M RNAs of tomato spotted wilt virus. *J. Gen. Virol.* 70:3469-3473.
- de Haan, P., Wagemakers, L., Peters, D., and Goldbach, R. W. 1990. The S RNA segment of tomato spotted wilt virus has an ambisense character. *J. Gen. Virol.* 71:1001-1007.
- Devereux, J., Heuberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis for the VAX. *Nucleic Acids Res.* 12:387-395.
- Elliott, R. M., and McGregor, A. 1989. Nucleotide sequence and expression of the small (S) RNA segment of Maguari Bunyavirus. *Virology* 171:516-524.
- Ghanekar, A. M., Reddy, D. V. R., Iizuka, N., Amin, P. W., and Gibbson, R. W. 1979. Bud necrosis of groundnut (*Arachis hypogaea*) in India caused by tomato spotted wilt virus. *Ann. Appl. Biol.* 93:173-179.
- Gubler, V. G., and Hoffman, J. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25:263-269.
- Hall, J. M., Speck, J., Geske, S., Mohan, K., and Moyer, J. W. 1993. A tospovirus with a serological distinct nucleocapsid isolated from onion (*Allium cepa*). (Abstr.) Page 357 in: *Abstr. Int. Congr. Virol.*, 11th.
- Hayati, I., Kline, A. S., Kim, K. S., and Gergerich, R. C. 1990. A tomato spotted wilt virus from *Verbesina alternifolia*. *Phytopathology* 80:1060-1061.
- Holton, T. A., and Graham, M. W. 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Res.* 19:1156.
- Iwaki, M., Honda, Y., Hanada, K., Tochihara, H., Yonaha, T., Hokama, K., and Yokoyama, T. 1984. Silver mottle disease of watermelon caused by tomato spotted wilt virus. *Plant Dis.* 68:1006-1008.
- Kormelink, R., van Poelwijk, F., Peters, D., and Goldbach, R. W. 1992. Non-viral heterogeneous sequences at the 5' ends of tomato spotted wilt virus (TSWV) mRNAs. *J. Gen. Virol.* 73:2125-2128.
- Law, M. D., and Moyer, J. W. 1990. A tomato spotted wilt-like virus with serologically distinct N protein. *J. Gen. Virol.* 71:933-938.
- Law, M. D., Speck, J., and Moyer, J. W. 1991. Nucleotide sequence of the 3' non-coding region and N gene of the S RNA of a serologically distinct tospovirus. *J. Gen. Virol.* 72:2597-2601.
- Maiss, E., Ivanova, L., Breyel, E., and Adam, G. 1991. Cloning and sequencing of the S RNA from a Bulgarian isolate of tomato spotted wilt virus. *J. Gen. Virol.* 72:461-464.
- Pang, S.-Z., Slightom, J. L., and Gonsalves, D. 1993. The biological properties of a distinct tospovirus and sequence analysis of its S RNA. *Phytopathology* 83:728-733.
- Reddy, D. V. R., Ratna, A. S., Sudarshana, M. R., Poul, F., and Kumar, I. K. 1992. Serological relationships and purification of bud necrosis virus, a tospovirus occurring in peanut (*Arachis hypogaea* L.) in India. *Ann. Appl. Biol.* 120:279-286.
- Reddy, D. V. R., Sudarshana, M. R., Ratna, A. S., Reddy, A. S., Amin, P. W., Kumar, I. K., and Murphy, A. K. 1991. The occurrence of yellow spot virus, a member of tomato spotted wilt virus group, on peanut (*Arachis hypogaea* L.) in India. Pages 77-88 in: *Virus-Thrips-Plant Interactions of Tomato Spotted Wilt Virus*. H. T. Hsu and R. H. Lawson, eds. Proc. USDA Worksh., USDA Agric. Res. Serv., ARS-87.

26. Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
27. Sreenivasulu, P., Demski, J. W., Reddy, D. V. R., Naidu, R. A., and Ratna, A. S. 1991. Purification and some serological relationships of tomato spotted wilt virus isolates occurring on peanut (*Arachis hypogaea*) in the USA. *Plant Pathol.* 40:503-507.
28. Stanley, K. K., and Luzzio, J. P. 1984. Construction of a new family of high efficiency bacterial expression vectors: Identification of cDNA clones coding for human liver proteins. *EMBO J.* 3:1429-1434.
29. Ullman, D. E., German, T. L., Sherwood, J. L., Westcot, D. M., and Cantone, F. A. 1993. *Tospovirus* replication in insect vector cells: Immunocytochemical evidence that the nonstructural protein encoded by the S RNA of tomato spotted wilt tospovirus is present in thrips vector cells. *Phytopathology* 83:456-463.
30. Verword, T. C., Dekker, B. M. M., and Hockema, A. 1989. A small-scale procedure for rapid isolation of plant RNAs. *Nucleic Acids Res.* 17:2362.
31. Wijkamp, I. M. G., van Lent, J., Kormelink, R., Goldbach, R. W., and Peters, D. 1993. Multiplication of tomato spotted wilt virus in its insect vector, *Frankliniella occidentalis*. *J. Gen. Virol.* 74:341-349.
32. Yeh, S.-D., and Chang, T.-F. 1995. Nucleotide sequence of the N gene of watermelon silver mottle virus, a proposed new member of the genus *Tospovirus*. *Phytopathology* 85:58-64.
33. Yeh, S.-D., Lin, Y.-C., Cheng, Y.-H., Jih, C.-L., Chen, M.-J. and Chen, C.-C. 1992. Identification of tomato spotted wilt-like virus on watermelon in Taiwan. *Plant Dis.* 76:835-840.