1998

Peanut yellow spot virus is a member of a new serogroup of Tospovirus genus based on small (S) RNA sequence and organization

Satyanarayana Tatineni
USDA-ARS, Satyanarayana.Tatineni@ars.usda.gov

S. Gowda
University of Florida

K. Lakshminarayana Reddy
International Crops Research Institute for the Semi-Arid Tropics

S. E. Mitchell
University of Georgia

W. O. Dawson
University of Florida

See next page for additional authors

Arch Virol (1998) 143: 353–364

Archives of Virology

Peanut yellow spot virus is a member of a new serogroup of Tospovirus genus based on small (S) RNA sequence and organization

T. Satyanarayana[1], **S. Gowda**[1], **K. Lakshminarayana Reddy**[2], **S. E. Mitchell**[3], **W. O. Dawson**[1], and **D. V. R. Reddy**[2]

[1]Citrus Research and Education Center, The University of Florida, Lake Alfred, Florida, U.S.A.
[2]International Crops Research Institute for the Semi-Arid Tropics-Asia Center, Patancheru, India
[3]Plant Genetic Resources Conservation Unit, USDA, The University of Georgia, Griffin, Georgia, U.S.A.

Accepted August 25, 1997

Summary. Peanut yellow spot virus (PYSV) represents a distinct tospovirus species based on serology and nucleic acid hybridization. The sequence of the S RNA was 2 970 nucleotides with 22 nucleotide long inverted repeats (with three mismatches) at the termini. The coding was ambisense with a long open reading frame (ORF) in each strand. The 5′-large ORF (1 440 nucleotides in the viral sense (v) strand) encoded a protein with a predicted size of 53.2 kDa that was identified as the nonstructural (NSs) protein based on 16–21% sequence identity and 42–48% sequence similarity with other tospoviruses. A 3′ ORF (741 nucleotides) in the virus complementary (vc) sense encoded a 28.0 kDa protein that was identified as the nucleocapsid (N) gene based on immuno-blot analysis of the in vitro expressed protein with PYSV polyclonal antiserum. The predicted N protein had 24–28% amino acid sequence identity and 44–51% sequence similarity with the members of other serogroups. In contrast to other tospoviruses, a third ORF (204 nucleotides) occurred in the vc strand, which could encode a protein with a predicted size of 7.5 kDa with two strong hydrophobic regions. The low degree of homology of N and NSs protein sequences with other serogroup members coupled with an additional ORF suggests that PYSV should be classified as a distinct species of the *Tospovirus* genus. This conclusion also is supported by the absence of serological cross reaction with other serogroups, and biological characteristics including thrips transmission, symptoms and host range.

Introduction

The members of *Bunyaviridae* infecting plants are included under the genus tospovirus [20]. The virions are quasi-spherical, enveloped and contain three linear ssRNA species, denoted small (S) RNA, medium (M) RNA and large (L) RNA [18, 28]. The RNAs form pseudo-circular structures associated with a nucleocapsid (N) protein and a few copies of a large protein, the putative viral RNA polymerase [29]. The two glycoproteins (c. 78 kDa (G1) and c. 59 kDa (G2)) presumably required for thrips transmission, are embedded in the viral envelope [19]. The complete genomic nucleotide sequences of tomato spotted wilt (T SWV), impatients necrotic spot (INSV) and peanut bud necrosis viruses (PBNV) have been determined ([4, 6, 12, 16, 24, 25, 30], Gowda et al. unpubl.). The L RNA is of negative polarity and encodes a putative viral RNA polymerase, whereas the S and M RNAs are ambisense. The M RNA contains two ORFs, one in the viral (v) sense that encodes a nonstructural (NSm) protein and the other in the viral complementary (vc) sense that encodes a precursor protein for G1 and G2 envelope glycoproteins. The S RNA also has two ORFs, the one in the v sense strand encodes a nonstructural (NSs) protein and that in the vc sense strand encodes the nucleocapsid (N) protein.

The genus *Tospovirus* consists of four serogroups based on the serological cross reactivity and sequence homology of the N protein. Serogroup I consists of TSWV [4, 17], which reacts weakly with the antibodies to serogroup II members [TSWV-B (Brazil isolate); groundnut ring spot virus (GRSV) and tomato chlorotic spot virus (TCSV)] [3, 21] do not react with antibodies to serogroup III members (INSV) [15]. The serogroup IV members, PBNV and watermelon silver mottle virus (WSMV) are serologically distinct from the members of serogroups I, II, and III [1]. The N protein of TSWV (serogroup I) showed 77–78% sequence identity with GRSV and TCSV (serogroup II) [3], 55% sequence identity with INSV (serogroup III)[6] and 33–35% sequence identity with PBNV and WSMV (serogroup IV) [31, 32].

Peanut yellow spot virus (PYSV) causes yellow leaf spots in peanuts which later coalesce and become necrotic. It was reported as a distinct tospovirus based on thrips transmission, serology and host range [22]. Recently PYSV was separated into a distinct species and proposed to be included in a newly established serogroup of the genus *Tospovirus* based on lack of serological cross reactivity and nucleic acid hybridization with TSWV, INSV and PBNV [26]. Although the incidence of PYSV in peanut fields can reach 90%, the crop losses due to this virus have not been determined (D. V. R. Reddy, unpubl.). In this paper, we show that the nucleotide sequence and genome organization of the S RNA segment of PYSV differs considerably from other tospoviruses.

Materials and methods

Purification of virus nucleocapsids and extraction of RNA

Peanut (*Arachis hypogaea* L.) plants exhibiting typical yellow spot symptoms of PYSV were collected from field plots at the ICRISAT-Asia center, India. The virus was isolated and

further maintained as described by Reddy et al. [22]. Extraction of virus RNA from purified nucleocapsids was done essentially as described by Satyanarayana et al. [26].

cDNA synthesis and cloning

The total RNA of PYSV was fractionated in 1.0% low melting point agarose under denaturing conditions [2]. The S RNA was eluted from the gel [23] and used as a template for making a cDNA library. Random primed first- and second-strand syntheses were performed as described by Gubler and Hoffman [9] using SuperScript Choice system (Gibco-BRL) followed by addition of Eco RI adaptors and cloned into pGEM-7f Z at Eco RI site. The 3′ end of PYSV S RNA was obtained by reverse transcription followed by polymerase chain reaction (RT-PCR) using an oligonucleotide 5′ AGAGCAATC which is complementary to the 3′ end of tospoviruses RNAs reported so far, and an oligonucleotide 5′CCTATAGCTGCCTGAAGCTCTTC, identical to nucleotides 2 170–2 192 of PYSV S RNA. The 5′ end of PYSV S RNA was also obtained by RT-PCR of purified RNA using an oligonucleotide, 5′AGAGCAATC, conserved at 5′ end of tospovirus nucleotides and an oligonucleotide, 5′CACCATCAAGGATACTTCTAATATGA, complementary to the nucleotides 1 286 –1 311 of PYSV S RNA. The gel purified RT-PCR products were cloned into pUC 119 at Sma I site for sequencing.

Nucleotide sequence and analysis

The sequencing of the double-stranded DNA was done with an automated DNA sequencer (Perkin-Elmer/Applied Biosystems, Model 373A). The nucleotide sequence and the deduced amino acid sequences were assembled and analyzed using the GCG program package [7]. CLUSTAL V [11] was used for comparative analysis including multiple alignments of tospoviruses N and NSs protein amino acid sequences and calculation of genetic distances. Distance was defined as (1- identity)/100, sequence gaps were not considered in the calculation, and no correction was made for multiple substitution. Phenograms were constructed by cluster analysis employing UPGMA [27] using Phylip (v.357c) with randomized order of data entry.

Expression of the nucleocapsid protein in Eschericia coli *and immuno-blot analysis*

The PYSV necleocapsid (N) gene was amplified from the first-strand cDNA synthesized using an oligonucleotide, 5′TTCGCCATGGCTACAAAAGGCGTTGTAAAG, complementary to nucleotides 2 871–2 890, contained an Nco I site (underlined) to place the ATG in the proximity of pET-15b. As a consequence, the second codon in the N gene was changed from TCT (ser) to GCT (ala). The other oligonucleotide used in PCR amplification of N gene was: 5′ ATCACTCGAGTCACAGCTCTCCGTTACCTATAG, with Xho I site (underlined) identical to 2 154 to 2 176. The PCR products were digested with Nco I and Xho I and cloned into the respective sites of pET-15b. The overnight culture of the N gene construct in pET-15b in *E. coli* (BL 21)DE 3 was diluted 1:100 in fresh LB medium containing ampicillin (50 μg/ml). Cells were grown for 4 h at 37 C and subsequently induced by adding IPTG to a final concentration of 0.4 mM. After electrophoresis in 12% polyacrylamide gels [14], proteins were electro-blotted onto nitrocellulose membrane and probed with PYSV polyclonal antiserum specific to nucleocapsids [22].

Results and discussion

cDNA cloning and sequencing of the PYSV S RNA

Among the cDNA clones generated by random priming of the S RNA from purified nucleocapsids, two overlapping clones representing nucleotids 1 193 to 2 265, and 1 559 to 2 164 were selected for sequencing as they were relatively large and specifically hybridized with the PYSV S RNA on Northern Blots (data not shown). The 5′ and 3′ ends of PYSV S RNA were obtained by RT-PCR using a 9-mer conserved tospoviral sequence present in all the tospoviral RNAs sequenced so far and PYSV S RNA specific primers (see Materials and methods). The 5′ and 3′ ends of the L RNA of INSV [30] and PBNV (Gowda et al. unpubl.) have been amplified and sequenced using the similar approach. The clones obtained by RT-PCR using specific primers were selected based on the presence of overlapping sequences.

Characteristics of the PYSV S RNA sequence

The complete nucleotide sequence of the PYSV S RNA was 2 970 nucleotides (nt) long (Fig. 1), which was similar to the size estimated by relative electrophoretic migration in denaturing agarose gels [26]. It had a base composition of 32% A, 32% U, 19% C and 17% G. The 5′ and 3′ ends of the viral RNA contained two untranslated regions, 57 and 75 nt respectively, with 22 nt complementary sequences (with three mismatches) at the termini which could form a strong panhandle structure characteristic of the members of *Bunyaviridae*.

The sequence analysis of PYSV S RNA revealed three ORFs with an ambisense arrangement. There was a large ORF of 1 440 nt located on the 5′ end of the V sense strand, initiating with an AUG codon at nucleotide 58 and terminating at an UGA stop codon at nucleotide 1 498 (numbered from the 5′ end of the viral strand) (Fig. 1). This ORF encoded a 480 amino acid polypeptide with a predicted Mr of 53.4 kDa. A search of the EMBL protein database revealed sequence similarity with corresponding nonstructural (NSs) proteins of PBNV and WSMV, and thus appeared to be the NSs protein. The deduced amino acid sequence of this ORF contained two potential N-glycosylation sites (Fig. 1). Another ORF was located near the 5′ end of the vc strand, initiating with an AUG codon at nucleotide 2 894 and terminating at an UGA stop codon at nucleotide 2 154 (numbered from the 5′ end of the v strand) (Fig. 1). This ORF could encode a protein of 246 amino acids with a predicted Mr of 28.0 kDa, which appeared to be the N protein, based on the sequence comparison to the N proteins of other tospoviruses (see below). The calculated Mr of 28.0 kDa of the putative N protein corresponded to the experimentally determined value for the N protein from purified nucleocapsids [26].

To confirm the identification of the 3′ -small ORF on vc strand as N gene based on the sequence analysis, we cloned the entire ORF into pET- 15b vector, creating a recombinant plasmid pET-15bN. After induction with IPTG, *E. coli* cells harboring this plasmid produced an abundant protein of 28 kDa, which is absent in uninduced *E. coli* cells. This protein was similar to the size expected for

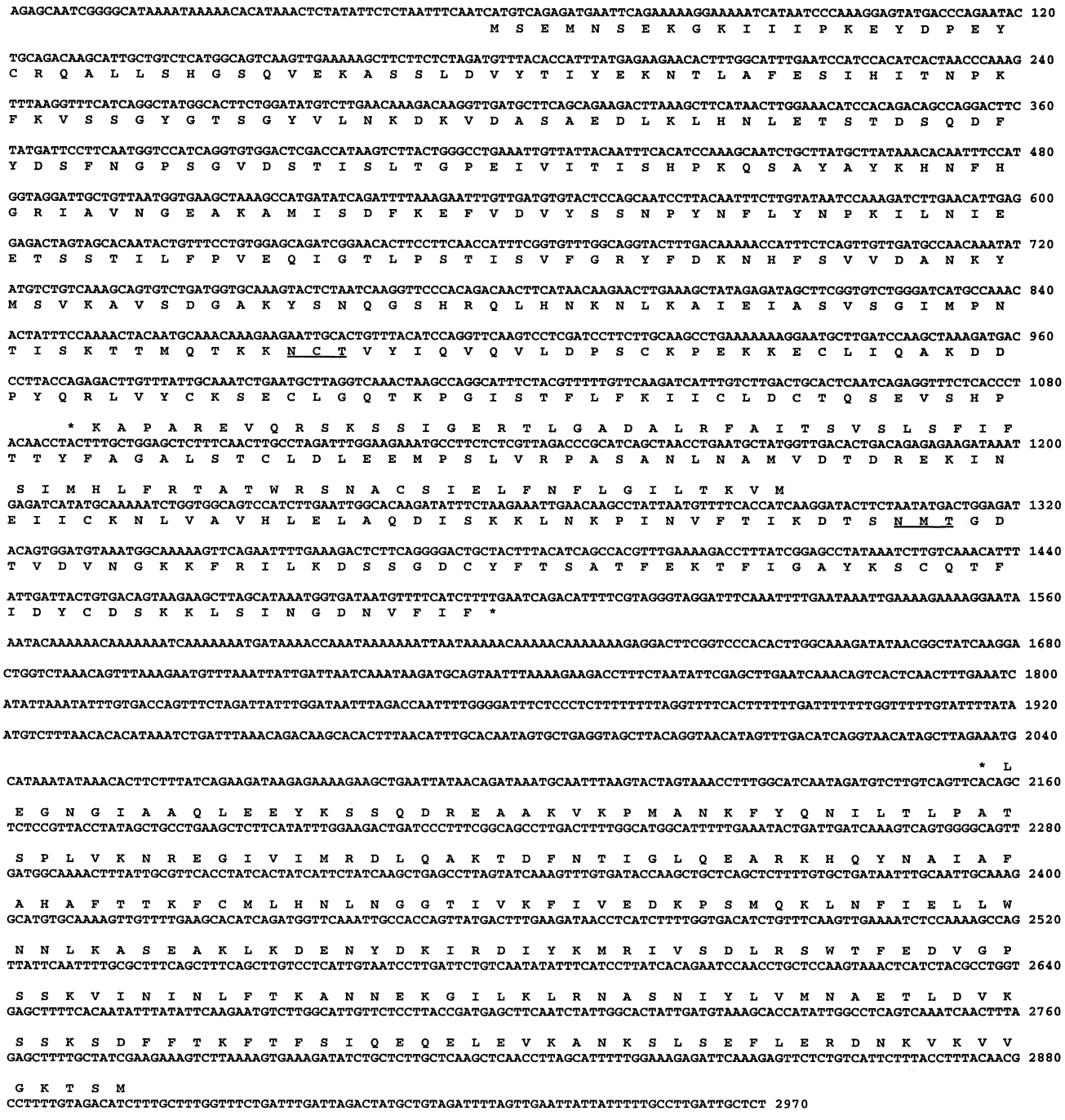

Fig. 1. The complete nucleotide sequence of PYSV S RNA is presented as DNA sequence in 5′ to 3′ virus-sense polarity. The deduced amino acid sequence of the protein encoded by the virus-sense RNA is shown above the sequence. The sequence of the proteins encoded by the virus-complementary sense RNA strand is shown above the sequence. Potential N-glycosylation sites are underlined. Asterisks indicate the stop codons

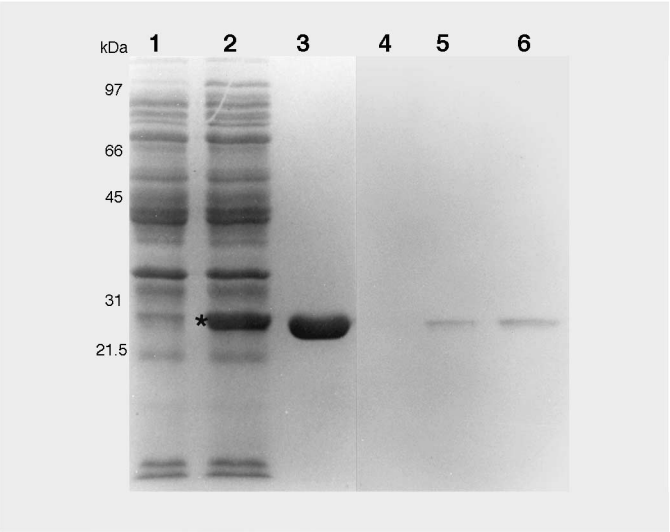

Fig. 2. SDS-PAGE (12%) analysis (*1–3*) and electro-blot immunoanalysis (*4–6*) of the PYSV N protein ORF after expression in *E. coli* (BL 21) DE 3 utilizing polyclonal antibodies. *1, 4 E. coli* with pET-15bN (-IPTG); *2, 5 E. coli* with pET-15bN (+IPTG); *3, 6* purified PYSV nucleocapsids; Molecular weight markers are indicated on the left side and position of the in vitro expressed N protein is indicated by an asterisk

the PYSV N protein (Fig. 2) and additionally reacted with a polyclonal antiserum prepared against PYSV N protein (Fig. 2).

The ORFs coding for NSs and N genes were separated by a 713 nt A+U-rich intergenic region (Fig. 1). Size differences between the S RNA molecules of different tospovirus isolates have been attributed to the variation in length of the intergenic region [10]. The predicted structure and stability of the PYSV S RNA intergenic region between 1 545 and 1 921 nt(free energy −83.2 kcal/mol) was comparable to the intergenic regions of PBNV and TSWV S RNAs, having a calculated free energy of −96.6 and −108.2 kcal/mole, respectively [4, 24].

A small ORF in vc strand of PYSV S RNA was observed in addition to the NSs and N ORFs. This ORF is initiated with an AUG codon at nucleotide 1 292 and terminates with an UAG stop codon at necleotide 1 088 (numbering of v sense RNA). It could encode a protein of 68 amino acids with a predicted molar mass of 7.6 kDa (Fig. 1). Two hydrophobic regions were predicted by the program Peptide Structure [13] to occur in the protein sequence at amino acids 1–16 and 26–48. Furthermore, a search of EMBL protein database revealed no statistically significant homology to any other reported protein sequences. The presence of an additional ORFs comparable to the size observed for PYSV has been reported for Maguari Bunyavirus S RNA [8] and for S RNA of WSMV (tomato isolate) [10]. However, evidence for existence of this protein in vivo has so far not been demonstrated.

Table 1. Comparison of the N and NSs proteins of different tospovirus isolates of serogroups I, II, III, IV and V

		Serogroup I		Serogroup II			Sero-group III	Serogroup IV		Sero-group V
		TSWV-BR 01	TSWV-L3	TSWV-B	GRSV	TCSV	INSV	PBNV	WSMV	PYSV
Sero-group I	TSWV-BR 01	**100**	98(99)	79(91)	78(90)	77(89)	55(69)	33(51)	35(51)	24(45)
	TSWV-L3	89(92)	**100**	80(91)	79(90)	78(89)	55(68)	34(51)	36(53)	24(45)
Sero-group II	GRSV	–[a]	–	–	**100**	81(89)	54(69)	34(53)	33(52)	26(50)
	TCSV	–	–	–	–	**100**	55(71)	33(53)	33(54)	25(49)
Sero-group III	INSV	52(69)	54(72)	53(70)	–	–	**100**	30(53)	32(49)	24(44)
Sero-group IV	PBNV	27(51)	22(45)	28(54)	–	–	22(46)	**100**	86(94)	26(51)
	WSMV	19(43)	23(47)	19(44)	–	–	22(47)	83(93)	**100**	25(50)
Sero-group V	PYSV	16(42)	18(43)	19(42)	–	–	19(44)	21(48)	20(48)	**100**

The percent identities and similarities (in parenthesis) of the N protein (above diagonal) and the NSs protein (below diagonal) as deduced from the nucleotide sequence data using the GAP program (gap weight 3.0, length weight 0.1) of the GCG sequence analysis software package

The sequence data were taken from the following publications: GRSV and TCSV [3], INSV [6], PBNV [24], TSWV-BR 01 [4], TSWV-L3 [17], WSMV [10, 31]

[a]Sequence data are not available for comparison

Sequence comparison of the PYSV N and NSs proteins

The identity and similarity values based on the amino acid sequences of the tospovirus N and NSs proteins were completely consistent with the known serological grouping. The amino acid sequence of PYSV N protein exhibited 24–28% identity and 44–51% similarity with the viruses of serogroups I–IV (Table 1). This value is in contrast to 77–79% identity reported between serogroups I and II, 55% identity observed between serogroups I and III and 33–35% identity reported between serogroups I and IV (Table 1). Thus, PYSV appears to represent the most distantly related taxonomic known member of the genus *Tospovirus*. The PYSV N protein consists of 246 amino acids, which is 12, 16 and 29 or 30 amino acids shorter than that reported for the members of serogroups I and II (258 amino acids), serogroup III (262 amino acids), and serogroup IV (275 or 276 aminoacids), respectively [3, 4, 6, 21, 24, 31]. Although the N protein of PYSV shared low degree of identity with those of other tospoviruses, some consensus residues are present (Fig. 3a). The C-terminal region (60 amino acids) of the PYSV N protein is highly variable as there are no consensus amino acid residues observed (Fig. 3a).

a

```
PBNV       MSNVKQLTEKKIKELLAGGSADVEIETEDSTPGFSFKAFYDTNRNL--EITFTNCLNILKCRKQIFAACKSG-KYVFCGKTI---VATNTDVGPDDWTFKRTEAFIRTKMASMVEKSKND 114
WSMV       MSNVKQLTEKKIKELLAGGTADVEIETEDSTPGFSFKAFFYNNKNI--EITFTNCLNILKCRKQIFAACKSG-KYNFCGKNI---VATSVDVGPEDWTFKRTEAFIRTKIVSMVEKSKNE 114
TSWV-BR01  MSKVK-LTKESIVALLTQG-KDLEFEEDQNLVAFNFKTFCLENIDQIKKMSVISCLTFLKNRQSIMKVIKQS-DFTFGKITIKK---TSDRIGGTDMTFRRLDSLIRVRLVE---ETGNS 111
TSWV-L3    MSKVK-LTKESIVALLTQG-KDLEFEEDQNLVAFNFKTFCLENLDQIKKMSVISCLTFLKNRQSIMKVIKQS-DFTFGKITIKK---TSDRIGATDMTFRRLDSLIRVRLVE---ETGNS 111
GRSV       MSKVK-LTKENIVSLLTQS-EDVEFEEDQNQVAFNFKTFCQENLDLIKKMSITSCLTFLKNRQSIMKVVKQS-DFTFGKVTIKK---NSERVEAKDMTFRRLDSMIRVKLIE---ETANN 111
TSWV-B     MSKVK-LTKENIVSLLTQS-ADVEFEEDQNQVAFNFKTFCQENLDLIKKMSITSCLTFLKNRQGIMKVVNQS-DFTFGKVTIKK---NSERVGAKDMTFRRLDSMIRVKLIE---ETANN 111
TCSV       MSKVK-LTRENIISLLTQA-GEIEFEEDQIKATFNFEDFCGENLDSIKKMSITSCLTFLKNRQSIMKVVNLC-DFTFGKITIKK---NSGRVGANDMTFRRLDSMIRVKLIE---ETGKA 111
INSV       MNKAK-ITKENIVKLLTQS-DSLEFEETQNEGSFNFTDFFTNNREKIQNMTTASCLSFLKNRQSIMRVIKSA-DFTFGSVTIKKTRNNSERVGVNDMTFRRLDAMVRVHLVGM--IKDNG 115
PYSV       MSTKGVVKVKNDRELFESLSKNAKVELEQEQISFTFKTFFDSKSSK-VDLTEANMVLYINSANRLKLIGKENNAKTFLNINIVK-----SSPGVDEFTWSRLDSVIRMKYIDRIKDY-NE 113
           *    .      *          *      * *  *          ..   . .    .          *   .*           . *  * .. .* .

PBNV       AAKQEMYNKIMELPLVAAYGLNVPASFDTCALRMMLCIGGPLPLLSSMTGLAPIIFPLAYYQNVKKEKLGI--KNFSTYEQVCKVAKVLSASQIEFKNELEEMFKSAVKLLSESNPGTAS 232
WSMV       AARQEMYGKIMELPLVAAYGLNVPASYDSCALRMMLCIGGPLPLLSSIRGLATIIFPLAYYQNIKKEKLGI--KNFSTYEQVCKVAKVLSASQVEFKDDLDVMFKQAVKILSESNTGTAS 232
TSWV-BR01  ENLNTIKSKIASHPLIQAYGLPLD---DAKSVRLAIMLGGSLPLIASVDSFEMISVVLAIYQDAKYKDLGIDPKKYDTKEALGKVCTVLKSKAFEMNEDQVKKGKEYAAILSSSNPNAKG 228
TSWV-L3    ENLNTIKSKIASHPLIQAYGLPLD---DAKSVRLAIMLGGSLPLIASVDSFEMISVVLAIYQDAKYKDLGIDPKKYDTKEALGKVCTVLKSKAFEMNEDQVKKGKEYAAILSSSNPNAKG 228
GRSV       ENLAIIKAKIASHPLVQAYGLPLD---DAKSVRLAIMLGGSIPLIASVDSLEMISVVLAIYQDSQVQELGIEPTKYNTKEALGKVCTVLKSKGFTMDDAQDNKGKEYAKILSSCNPNAKG 228
TSWV-B     ENLAIIKAKIASHPLVQAYGLPLA---DAKSVRLAIMLGGSIPLIASVDSFEMISVVLAIYQDAKYKELGIEPTKYNTKEALGKVCTVLKSKGFTMDDAQINKGKEYAKILSSCNPNAKG 228
TCSV       ENLAIIKSKIASHPLVQAYGLPLT---DAKSVRLAIMLGGSIPLIASVDSFEMISIILAIYQDAKYKDLGIEPSKYNTKEALGKVCTVLKSKGFTMDEEQVQKGKEYATILSSCNPNAKG 228
INSV       SALTEAINSLPSHPLIASYGLATT---DLKSCVLGVLLGGSLPLIASVLNFEIAALVPAIYQDAKHVELGIDMSKFSTKEAVGKVCTVLKSKGYSMNSVEIGKAKQYADILKACSPKAKG 232
PYSV       DKLKAESAKLNNW-LLEIFNLKQMSPKD--EVIFKVITGGNLNHLMCFKTTFAHAFAIANYQHKRAEQLGI--TNFDTKAQLDRMIVIGERNKVLPSTAPLTLINQYFK-------NAMP 221
                     .    *.  ..*      *       .  ** .  . .           * **  .   ***    . *   . ..  .                          .

PBNV       SISLKKYDEQVKYMDKAFSASLSMDDYGEHSKKKSSKAGPSLEL 276
WSMV       SISLKKYEDQVKYMDRVFSANLSVDDYGDHS-KKSSKPSTSLEV 275
TSWV-BR01  SVAMEHYSETLNKFYEMFGVKKQ-AKLAELA------------- 258
TSWV-L3    SIAMEHYSETLNKFYEMFGVKKQ-AKLTELA------------- 258
GRSV       SIAMDYYSDNLDKFYEMFGVKKE-AKIAGVA--------------258
TSWV-B     SIAMDYYSDNLEKFYEMFGVKKE-AKIAGVA--------------258
TCSV       SIAMEHYSEHLDKFYAMFGVRKE-AKISGVA--------------258
INSV       LAAMDHYKEGLTSIYSMFNATIDFGKNDSI-------------- 262
PYSV       KVKAAERDQ--SSKYEELQAAIG---NGEL---------------246
```

Fig. 3. Multiple alignment of the **a** N protein sequence of nine tospovirus isolates and **b** NSs protein sequence of seven tospovirus isolates of five serogroups. The alignment was done with the CLUSTAL V program. Dashes represent gaps introduced to reach optimal alignment. Perfectly conserved residues are indicated by an asterisk and highly conserved residues are indicated by a dot. The sequence data were taken from the following publications: GRSV and TCSV [3], INSV [6], PBNV [24], TSWV-BR 01 [4], TSWV-L3 [17], WSMV [31, 32]

b

```
PBNV      ----------MSTARSAASEFVKS---YG---TR-DNRAI-----NDCYSVFS-GE----GVNFLNLFMHNNAGI-KSAFSINDLGRN----EDIKIHEAEV-IDSCHDYHYFE---KFG 84
WSMV      ----------MSTAKNAASEFIKS---YG---TR-DNRAV-----NDCYSVFT-GE----GVNFLNLFMHNNAGI-KSAFSINDLGRN----EDVKIHEAEV-VDSCHDYHYFE---KFG 84
TSWV-L3   ----------MSSSVYESIIQTRAS-VWG---STASGKAV-----VDSYWIHELGT----GSLLVQTQLYSDSRS-KSSFGYT--AKV----GNLPCEEEEI-LSQHVYIPIFD---DID 86
TSWV-BR01 ----------MSSSVYESIIQTRAS-VWG---STASGKAV-----VDSYWIHELGT----GSQLVQTQLYSDSRS-KVVLWLY--CKV----GIFPVKKKRF-LSQHVYIPIFD---DID 86
TSWV-B    ----------MSSGVYESIIQTKAS-VWG---STASGKSI-----VDSYWIYEFPT----GSPLVQTQLYSDSRS-KSSFGYT--SKI----GDIPAVEEEI-LSQNVHIPVFD---DID 86
INSV      ----------MSSAMYETIIKSKSS-IWG---TTSSGKAV-----VDSYWIHDQSS----GKKLVEAQLYSDSRS-KTSFCYT--GKV----GFLPTEEKEI-IVR-CFVPIFD---DID 85
PYSV      -------MSEMNSEKGKIIIPKEYDPEYCRQALLSHGSQVEKASSLDVYTIYEKNTLAFESIHITNPKFKVSSGYGTSGYVLNK-DKVDASAEDLKLHNLETSTDSQDFYDSFNGPSGVD112
                     * .                         .  .      *  * .                      .              .                          *

PBNV      LDIIFCEHEMSLVVGKPGVKNTGCKFTMHNQIFNPNSDTLALTPGTVSENEFYEKIKIRSNGLLPSGWYQDECCKNNFYIATNGDLTLDYGFS-VMGKTTSYWRESISR-EKILSVKQKC 202
WSMV      LDITFCEHEMNLIVRKPGVKNTGCKFTMHNQIFNPNSNILNMTPGTISEDAFYDRSKITSSGLLPSGWCLDECYKNNFYIATNGDLTLDYGFS-VMGKTTSYWRESMSR-EKILSVKQKC 202
TSWV-L3   FSINIDDSVLALSVCSNTVNTNGVKHQGHLKVLSP--DQLHSIGSTMNRSDIKDRFQLQEKDIIPNDRYIEAANKGSLSCVKEHTYKIETCYNQALGKVNVLSPNRNVH-EWLYSFKPSF 203
TSWV-BR01 FSINIDNSVLALSVCSNTVNANGVKHQGHLKVLSP--AQLHSIESIMNRSDITDRFQLQEKDIIPNDKYIEAANKGSLSCVKEHTYKIEMCYNQALGKVNVLSPNRNVH-EWLYSFKPNF 203
TSWV-B    FSININDSFLAISVCSNTVNTNGVKHQGHLKVLSL--AQLHPFEPVMSRSEIASRFRLQEEDIIPDDKYISAANKGSLSCVKEHTYKVEMSHNQALGKVNVLSPNRNVH-EWLYSFKPNF 203
INSV      LNFSFSGNVVEILVRSNTTNTNGVKHQGHLKVLSS--QLLRMLEEQIAVPEITSRFGLKESDIFPPNNFIEAANKGSLSCVKEVLFDVKYSNNQSMGKVSVLSPTRSVH-EWLYTLKPVF 202
PYSV      STISLTGPEIVITISHPKQSAYAYKHNFHGRIAVNGEAK----AMISDFKEFVDVYSSNPYNFLYNPKILNIEETSSTILFPVEQIGTLPSTISVFGRYFDKNHFSVVDANKYMSVKAVS 228
                    . . .           *   *  ..                                                                      *.                 . *

PBNV      LPDN--TVPTNRLLSTSTVKGIQLGSELAPDTTVILSLKQNLNVDLKSQYRVSFHGIQ-EEGAFARTFCVPFENKSRMICLYAKTVVDNSN-ERTTLIIKVVTKTIES-----HLVIPIR 313
WSMV      FSDN--TVPTNRLLSTSTVKGIQIGSELASDTTVILSCKQNLNIDLKSQYRISFHGIQ-EEGAFARTFCVPFENKSRMISFYAKTVADNSN-ERTTLIIKVVTKTVDS-----HLVIPIR 313
TSWV-L3   --NQ--VESNNRTVNSLAVKSLLMSAENNIMPNSQAFVKASTDSHFKLSLWLRVPKVL-KQVSIQKLFKVAGD-ETNKTFYLSIVCIPNHNSVETALNISVICKH----------QLPIR 307
TSWV-BR01 --NQ--VESNNRTVNSLAVKSLLMSAENNIMPNSQ----ASTDSHFKLSLWLRVPKVL-KQVSIQKLFKVAGD-ETNKTFYLSIACIPNHNSVETALNITVICKH----------QLPIR 303
TSWV-B    --NQ--IESNNRTVNSLAVKSLLMATENNIMPNSQAFVKASTDSHFKLSLWLRIPKVL-KQIAIQKLFKFAGD-ETGKSFYLSIACIPNHNSVETALNVTVICRH----------QLPIP 307
INSV      --NQ--SQTNNRTVNTLAVKSLAMSATSDLMSDTHSFVRLNNNKPFKISLWMRIPKIM-KSNTYSRFFTLSDE-SSPKEYYISIQCLPNHNNVETVIEYNFDQSN----------LFLNQ 306
PYSV      DGAKYSNQGSHRQLHNKNLKAIEIASVSGIMPNTISKTTMQTKKNCTVYIQVQVLDPSCKPEKKECLIQAKDD-PYQRLVYCKSECLGQTKPGISTFLFKIICLDCTQSEVSHPTTYFAG 347
                     .* .     .* . . .                           .                    .                  .     .

PBNV      NHINCVKKIGARIGLVD------S---------IESD-PNYNQMIVKEL-LGVHTQFAINLSLVIKKPIIVFKMYDI-ELNNNHIDVYGRILNYQTDAEGNAYFLSKTLE--VLPKSLST 413
WSMV      NXINCEKKIGARVGLVD------F---------CDSD-PNYNQLIVKNL-LSVHTQFAINLSEVIQKPVIVFKMYDK-ELHNDYTEVSGRILNYQVDSERNVISSQRLLR--FCPKSLST 413
TSWV-L3   KCKAPFE-LSMMFSDLKEP---YN---------IVHD-PSYPQRIVHAL-LETHTSFAQVLCNNLQEDVIIYTLNNY-ELTPGKLDLGERTLNYSEDICKRKYFLSKTLE--CLPSNTQT 409
TSWV-BR01 KCKAPFE-LSMMFSDLKEP---YN---------IVHD-PSYPKGSVPMLWLETHTSLHKFFATNLQEDVIIYTLNNL-ELTPGKLDLGERTLNYSEDAYKRKYFLSKTLE--CLPSNTQT 406
TSWV-B    KSKAPFE-LSMIFSDLKEP---YN---------TVHD-PSYPQRIVHAL-LETHTSFAQVLCNKLQEDVIIYTINSP-ELTPAKLDLGERTLNYSEDASKKKYFLSKTLE--CLPVNVQT 409
INSV      LLLAVIHKIEMNFSDLKEP---YN---------VIHD-MSYPQRIVHSL-LEIHTELAQTVCDSVQQDMIVFTINEP-DLKPKKFELGKKTLNYSEDGYGRKYFLSQTLK--SLPRNSQT 409
PYSV      ALSTCLD-LEEMPSLVRPASANLN-------AMVDTDREKINEIICKNL-VAVHLELAQDISKKLNKPINVFTIKDTSNMTGDTVDVNGKKFRILKDSSGDCYFTSATFEKTFIGAYKSC 458
                   .      .                         *            * .  *        .  .   . .. .     .      .. .      *                  .

PBNV      LSYLGSIAPIQWEESLEHQHFIVKPE--------------------------------- 439
WSMV      LSYLGSIAPIQWKECLEHQHFVVKAE--------------------------------- 439
TSWV-L3   MSYLDSIQIPSWKIDFARGEIKISPQPVSVAKSLLKLDLSGIKKKGSKISETHASGSK 467
TSWV-BR01 MSYLDSIQIPSWKIDFARGEIKISPQSISVAKSLLKLDLSGIKKKESKVKEAYASGSK 464
TSWV-B    MSYLDSIQIPSWKIDFARGEIRISPQSTPIARSLLKLDLSKIKEKKSLTWETSSYDLE 467
INSV      MSYLDSIQMPDWKFDYAAGEIKISPRSEDVLKAISKLDLN------------------ 449
PYSV      QTFIDYCD--SKKLSINGDNVFIF---------------------------------- 480
           ...                    .
```

Fig. 3 (*continued*)

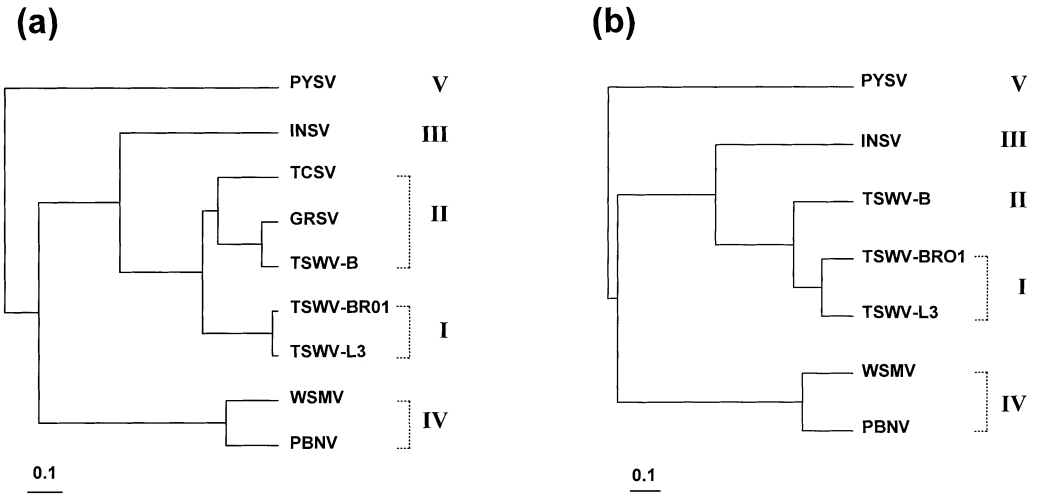

Fig. 4. Phenograms showing the genetic relationships among various tospovirus isolates based on the amino acid sequences of the **a** N and **b** NSs proteins. The scale at the bottom of the figure indicate genetic distance. Roman numerals on the right denote the various serogroups of *Tospovirus* genus

The PYSV NSs protein sequence revealed only 16–21% identity and 42–48% similarity with the members of serogroups I–IV. The PYSV NSs protein was 16, 13, 31 and 41 amino acids longer than the respective proteins of TSWV, TSWV-B, INSV and PBNV, respectively [4, 6, 21, 24]. Although the NSs protein of PYSV shared only 16–21% identity with those of other tospoviruses, still some consensus residues are observed (Fig. 3b).

The results of a clustering analysis of the aligned sequences of N and NSs proteins of tospoviruses are shown in Fig. 4. The phenograms suggest that PYSV should be included into a separate group based on N or NSs protein sequences. In conclusion, the observed low sequence identity of N (24–28%) and NSs (16–21%) proteins with serogroups I–IV members of Tospoviruses together with differences in host range, thrips transmission, serological cross reaction and nucleic acid hybridization analysis with TSWV, INSV and PBNV [22, 26] supports the inclusion of PYSV as a distinct member under new serogroup of the genus *Tospovirus*.

Acknowledgements

The nucleotide sequence data reported in this paper have been reported to the GenBank nucleotide sequence database and have been assigned the accession number AF013994. This paper is submitted as Journal Article No. JA 2215 by ICRISAT.

References

1. Adam G, Yeh SD, Reddy DVR, Green SK (1993) The 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
2. Bailey JM, Davidson N (1976) Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Anal Biochem 70: 75–85

3. De Avila AC, de Haan P, Kormelink R, Resende R de O, Goldbach RW, Peters D (1993) Classification of tospoviruses based on phylogeny of nucleoprotein gene sequences. J Gen Virol 74: 153–159
4. De Haan P, Wagemakers L, Peters D, Goldbach R (1990) The S RNA segment of tomato spotted wilt virus has an ambisense character. J Gen Virol 71: 1 001–1 007
5. De Haan P, Kormelink R, Resende R de O, Van Poelwijk F, Peters D, Goldbach R (1991) Tomato spotted wilt virus L RNA encodes a putative RNA polymerase. J Gen Virol 71: 2 207–2 216
6. De Haan P, de Avila AC, Kormelink R, Westerbroek A, Gielen JJL, Peters D, Goldbach R (1992) The nucleotide sequence of the S RNA of Impatiens necrotic spot virus, a novel tospovirus. FEBS Lett 306: 27–32
7. Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395
8. Elliott RM, McGregor A (1989) Nucleotide sequence and expression of the small (S) RNA segment of Maguari Bunyavirus. Virology 171: 516–524
9. Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. Gene 25: 263–269
10. Heinze C, Maiss E, Adam G, Casper R (1995) The complete nucleotide sequence of the S RNA of a new tospovirus species, representing serogroup IV. Phytopathology 85: 683–690
11. Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. CABIOS 8: 189–191
12. Kormelink R, de Haan P, Meurs C, Peters D, Goldbach R (1992) The nucleotide sequence of the M RNA segment of tomato spotted wilt virus, a bunyavirus with two ambisense RNA segments. J Gen Virol 73: 2 795– 2 804
13. Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. J Mol Biol 157: 105–132
14. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T-4. Nature 277: 680–685
15. Law MD, Moyer JW (1990) A tomato spotted wilt-like virus with serologically distinct N protein. J Gen Virol 71: 933–938
16. Law MD, Speck J, Moyer JW (1992) The M RNA of Impatiens necrotic spot tospovirus (Bunyaviridae) has an ambisense genomic organization. Virology 188: 732–741
17. Maiss E, Ivanova L, Breyel E, 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
18. Mohamed NA (1981) Isolation and characterization of subviral structures from tomato spotted wilt virus. J Gen Virol 53: 197–206
19. Mohamed NA, Randles JW, Francki RIB (1973) Protein composition of tomato spotted wilt virus. Virology 56: 12–21
20. Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (1995) Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien New York (Arch Virol [Suppl] 10)
21. Pang SZ, Slightom JL, Gonsalves D (1993) The biological properties of a distinct tospovirus and sequence analysis of its S RNA. Phytopathology 83: 728–733
22. Reddy DVR, Sudarshana MR, Ratna AS Reddy AS, Amin PW, Kumar IK, Murthy AK (1991) The occurrence of yellow spot virus, a member of tomato spotted wilt virus group, on peanut (Arachis hypogaea L) in India. In: Hsu HT, Lawson RH (eds) Virus-Thrips-Plant interactions of tomato spotted wilt virus. Proceedings of the USDA Agricultural Research Service, ARS-87, pp 77–88

23. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
24. Satyanarayana T, Mitchell SE, Reddy DVR, Brown S, Kresovich S, Jarret R, Naidu RA, Demski JW (1996a) Peanut bud necrosis tospovirus S RNA: complete nucleotide sequence, genome organization and homology to other tospoviruses. Arch Virol 141: 85–98
25. Satyanarayana T, Mitchell SE, Reddy DVR, Kresovich S, Jarret J, Naidu RA, Gowda S, Demski JW (1996b) The complete nucleotide sequence and genome organization of the M RNA segment of peanut bud necrosis tospovirus and comparison with other tospoviruses. J Gen Virol 77: 2 347–2 352
26. Satyanarayana T, Lakshminarayana Reddy K, Ratna AS, Deom CM, Gowda S, Reddy DVR (1996c) Peanut yellow spot virus: A distinct tospovirus species based on serology and nucleic acid hybridization. Ann Appl Biol 129: 237–245
27. Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco
28. Van Den Hurk J, Tas PWL, Peters D (1977) The ribonucleic acid of tomato spotted wilt virus. J Gen Virol 36: 81–91
29. van Poelwijk F, Boye K, Oosterling R, Peters D, Goldbach R (1993) Detection of the L protein of tomato spotted wilt virus. Virology 197: 468–470
30. van Poelwijk F, Prins M, Goldbach R (1997) Completion of the impatiense necrotic spot virus genome sequence and genetic comparison of the L protein within the family Bunyaviridae. J Gen Virol 78: 543–546
31. Yeh SD, Chang TF (1995) Nucleotide sequence of the N gene of watermelon silver mottle virus, a proposed new member of the genus tospovirus. Phytopathology 85: 58–64

Authors' address: Dr. D. V. R. Reddy, International Crops Research Institute for the Semi-Arid Tropics-Asia Center, Patancheru, 502 324, India.

Received June 17, 1997