Journal of General Virology (1996), **77**, 2347–2352. Printed in Great Britain SHORT COMMUNICATION

The complete nucleotide sequence and genome organization of the M RNA segment of peanut bud necrosis tospovirus and comparison with other tospoviruses

T. Satyanarayana,[1]† S. E. Mitchell,[2] D. V. R. Reddy,[1] S. Kresovich,[2] R. Jarret,[2] R. A. Naidu,[1] S. Gowda[4] and J. W. Demski[3]

[1] Crop Protection Division, International Crops Research Institute for the Semi-Arid Tropics Asia Center (ICRISAT-IAC), Patancheru-502 324, India

[2,3] USDA Plant Genetic Resources Conservation Unit[2] and Department of Plant Pathology,[3] The University of Georgia, Experiment Station, Griffin, GA 30223, USA

[4] Citrus Research and Education Center, The University of Florida, Lake Alfred, FL 33850, USA

The M RNA of peanut bud necrosis virus (PBNV; synonym groundnut bud necrosis virus) is 4801 nucleotides in length. It comprised two ORFs in an ambisense organization and terminal inverted repeats. The 3′ large ORF (3363 nucleotides in the virus-complementary strand) encoded a protein with a predicted size of 127·2 kDa which was identified as the glycoprotein precursor (GP) of the G1 and G2 glycoproteins. A comparison of the deduced amino acid sequence of GP revealed 37% identity and 58–59% similarity with that of tomato spotted wilt virus (TSWV, serogroup I) and impatiens necrotic spot virus (INSV, serogroup III), and 21–23% identity and 44–47% similarity with those of other members of the genus *Bunyavirus*. The 5′ small ORF (924 nucleotides in the virus-sense strand) encoded a 34·2 kDa protein which was identified as the non-structural (NSm) protein based on 41–43% identity and 60–63% similarity with that of TSWV and INSV. Defective RNA molecules derived from the genomic M RNA were detected during continuous passage of the virus by sap inoculations.

The members of the genus *Tospovirus* of the family *Bunyaviridae* are transmitted by thrips in a propagative mode. The quasi-spherical-shaped membrane-bound virus particles of 80–110 nm in diameter contain three linear genomic RNA species, small (S), medium (M) and large (L) RNA (Elliot, 1990). Tospoviruses have traditionally been classified on the basis of the antigenic properties of their nucleocapsid (N) proteins (de Ávila *et al.*, 1993; Adam *et al.*, 1993). More recently, sequence similarity, particularly of the N gene, has been used for their classification (de Ávila *et al.*, 1993; de Haan *et al.*, 1990; Satyanarayana *et al.*, 1996; Yeh & Chang, 1995). Proteins from members of serogroup I [tomato spotted wilt virus (TSWV) isolates; de Haan *et al.*, 1990] reacted weakly with antibodies to serogroup II members (TSWV-B; tomato chlorotic spot virus; groundnut ring spot virus; de Ávila *et al.*, 1993; Pang *et al.*, 1993) did not react with antibodies raised against members in serogroup III [impatiens necrotic spot virus (INSV); Law *et al.*, 1991]. Recently, serogroup IV [peanut bud necrosis virus (PBNV) and watermelon silver mottle virus] has been recognized on the basis of lack of N protein amino acid sequence homology with the members of serogroups I, II and III (Satyanarayana *et al.*, 1996; Yeh & Chang, 1995).

The S RNA of PBNV was shown to comprise 3057 nucleotides (Satyanarayana *et al.*, 1996). The N and NSs proteins showed 30–34% and 22–28% sequence identity, respectively, with the members of serogroups I, II and III. Here we report the complete nucleotide sequence of the genomic M RNA segment of PBNV, its coding strategy, comparison of amino acid sequences of GP and NSm proteins with those of TSWV and INSV, and occurrence of defective RNAs.

Peanut (*Arachis hypogaea* L.) plants exhibiting typical symptoms of PBNV were collected from field plots at the ICRISAT Asia Center, India. Extraction of virus RNAs from purified nucleocapsids was done as described by Satyanarayana *et al.* (1996). The total RNA of PBNV was fractionated in 1·0% LMP agarose gel under denaturing conditions (Bailey & Davidson, 1976). The M RNA was eluted from the gel (Sambrook *et al.*, 1989) and used as a template for making a cDNA library, RT–PCR and 5′ rapid amplification of cDNA

Author for correspondence: D. V. R. Reddy.
Fax +91 40 241239. e-mail D.Reddy@cgnet.com

† **Present address:** Citrus Research and Education Center, The University of Florida, Lake Alfred, FL 33850-2299, USA.

The nucleotide sequence data presented in this paper have been submitted to the Genbank database and assigned the accession number U42555.

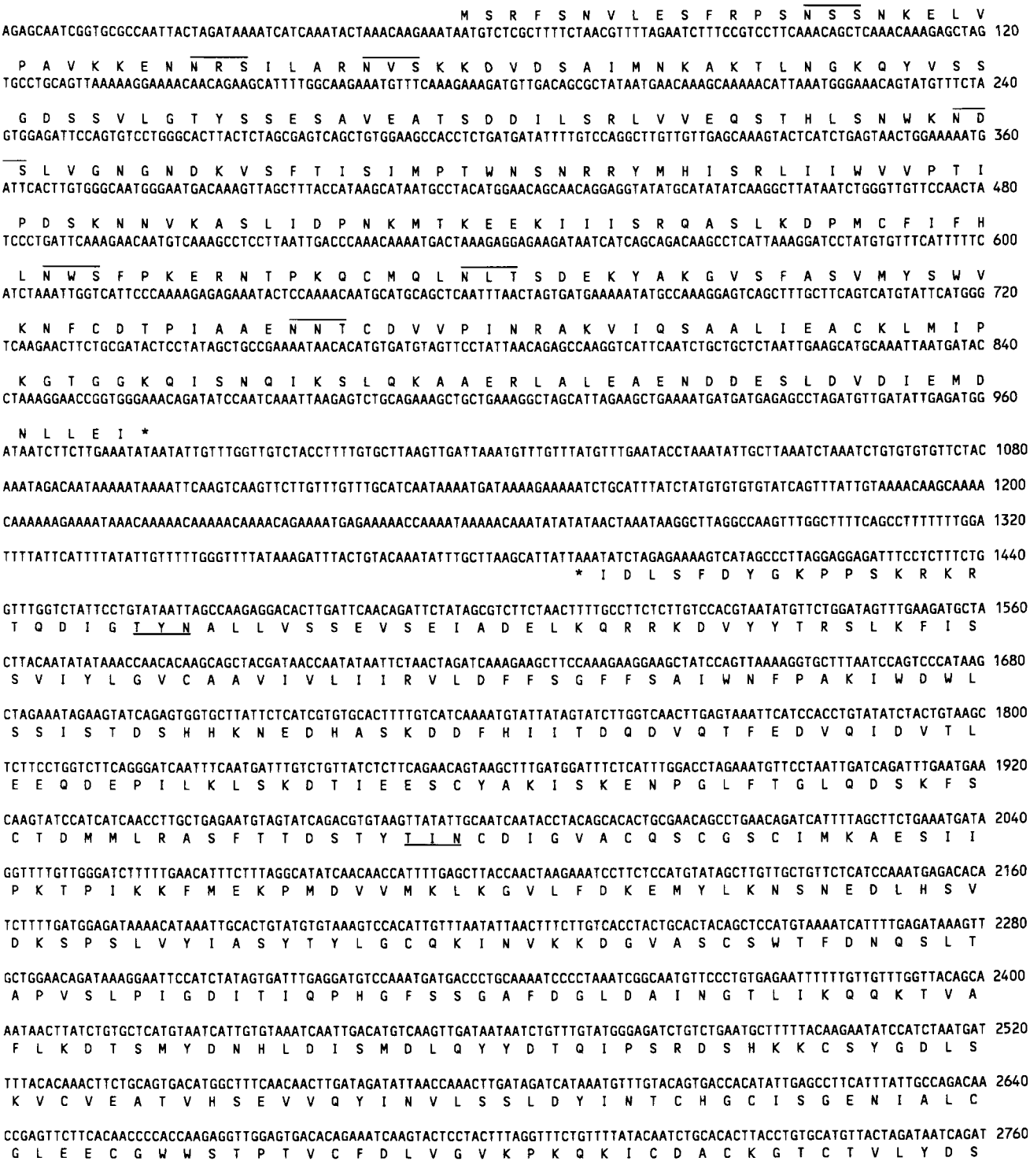

Fig. 1. For legend see opposite.

ends (RACE). Random-primed first- and second-strand synthesis was performed as described by Gubler & Hoffman (1983) using SuperScript Choice system (Gibco BRL). Prior to DNA sequencing, cDNA clones were analysed by dot, Northern and Southern blots (Sambrook *et al.*, 1989) to confirm the presence of inserts specific to M RNA and to select overlapping clones.

The M RNA 3′ terminus was amplified by RT–PCR using

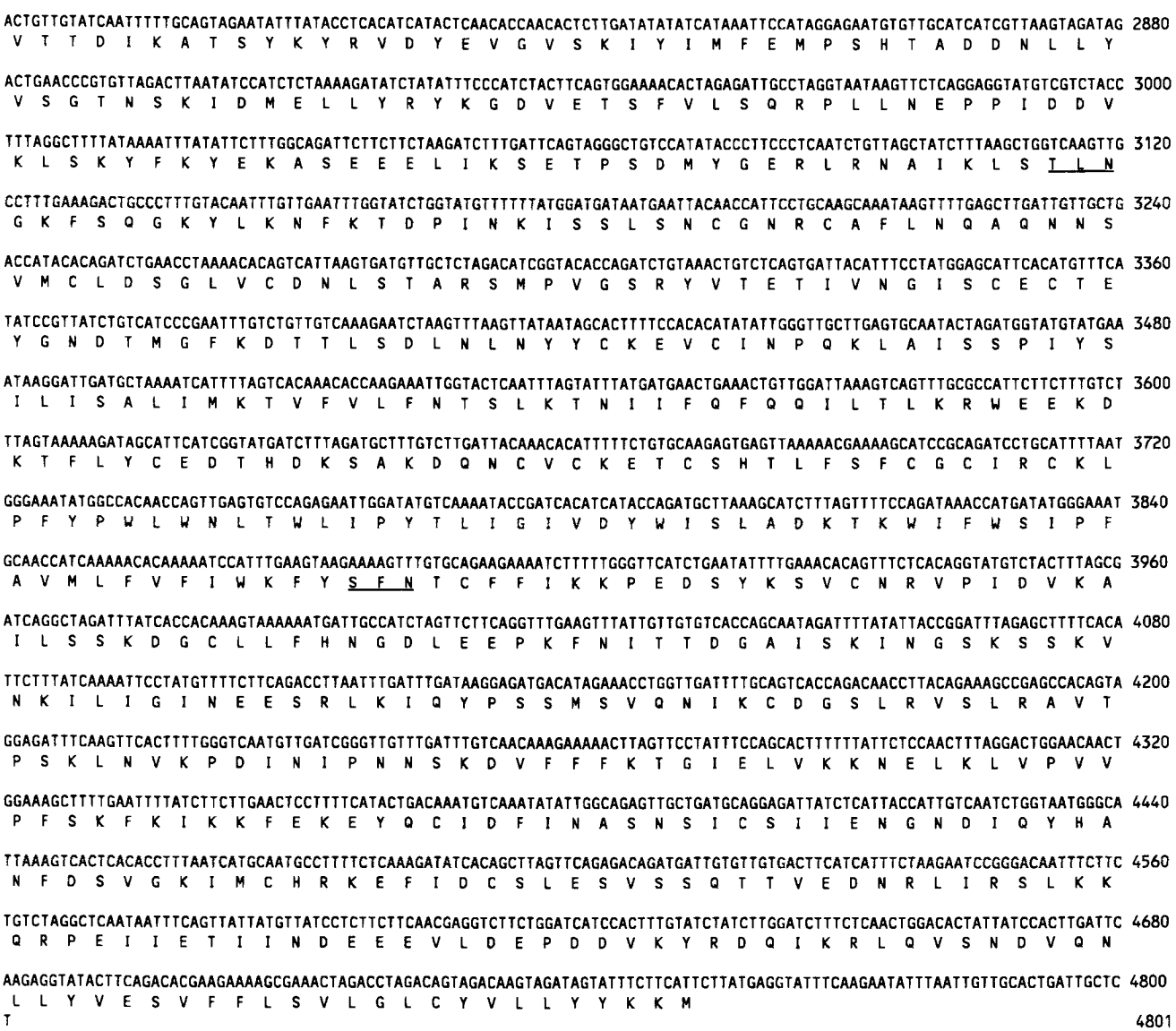

Fig. 1. The complete nucleotide sequence of PBNV M RNA is presented as DNA sequence in 5′ to 3′ virus-sense polarity. The deduced amino acid sequence of the protein encoded by the virus-sense RNA is shown above the sequence. The sequence of the protein encoded by the virus-complementary sense RNA strand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

an oligonucleotide (5′ AGAGCAATC) complementary to the 3′-terminal sequence of M RNA of TSWV and INSV (Kormelink *et al.*, 1992; Law *et al.*, 1992) and an oligonucleotide (5′ TCTTCTGTCTAGGCTCAATAATTTCA) identical to positions 4556–4581 of the PBNV M RNA. The 5′ terminus of PBNV M RNA was obtained by 5′ RACE (Frohman *et al.*, 1988) according to the manufacturer's instructions (Gibco BRL). The PCR products were gel-purified and cloned into a pGEM-T vector (Promega).

Sequencing of cDNA and PCR clones was done with an automated DNA sequencer (Perkin-Elmer/Applied Biosystems model 373 A) using *Taq* cycle sequencing with fluorescence-based chain termination chemistry (Perkin-Elmer/Applied Biosystems). Electrophorograms were edited and cDNA contig assembled with the Sequencher 2.1 program (Gene Code). Nucleotide and deduced amino acid sequences were assembled and analysed using the GCG program package (Devereux *et al.*, 1984).

The complete sequence of the PBNV M RNA (Fig. 1) was 4801 nucleotides long and was consistent with the size estimated from the denaturing agarose gel. It had a base composition of 32·3% A, 32·7% U, 17·8% C and 17·2% G. The 5′ and 3′ termini of the M RNA were identical for 13 of the first 14 nucleotides and could be folded into a stable panhandle structure. The first nine inverted repeats were identical to those of PBNV S RNA and the M RNAs of TSWV and INSV.

Analysis of the sequence revealed two non-overlapping

Table 1. Comparison of the glycoprotein precursor peptide of PBNV with those of other bunyaviruses

Virus*	INSV	TSWV	SSHV	LCV	GSV	BWV
PBNV	37·0†	37·2	21·7	22·9	21·2	20·5
	47·5‡	58·6	46·5	45·3	43·5	45·0
INSV		64·7	22·9	22·2	22·8	21·2
		79·2	44·1	44·0	43·9	43·9
TSWV			22·6	21·3	20·3	20·9
			47·1	45·0	42·4	46·4

* The sequence data were taken from the following papers: TSWV, Kormelink *et al.* (1992); INSV, Law *et al.* (1992); snowshoe hare virus (SSHV), Eshita & Bishop (1984); La Crosse virus (LCV), Grady *et al.* (1987); Germiston virus (GSV), Pardigon *et al.* (1988); Bunyamwera virus (BMV), Lees *et al.* (1986). Amino acid identity and similarity were analysed using the BESTFIT program (gap weight 3·0, length weight 0·1) of the GCG sequence analysis software package.
† Percentage amino acid identity.
‡ Percentage amino acid similarity.

ORFs in an ambisense orientation similar to the M RNAs of TSWV and INSV (Fig. 1). The large ORF (3363 nucleotides) was located on the virus-complementary strand, initiating with an AUG codon at nucleotide 4754 and terminating at an UAA stop codon at nucleotide 1389 (numbered from the 5′ end of the virus-sense strand) (Fig. 1). This ORF could code for a 1121 amino acid polypeptide with a predicted molecular mass of 127·2 kDa. The protein was identified as the glycoprotein precursor (GP) based on the sequence similarity to glycoproteins of TSWV, INSV and other members of the genus *Bunyavirus* (Table 1), and the presence of hydrophobic regions. It was concluded that the ORF encodes the precursor to the G1 and G2 glycoproteins, which is cleaved by an unknown mechanism. Analysis of the amino acid sequence of the predicted translation product of the large ORF demonstrated the presence of four potential *N*-glycosylation sites predicted by the MOTIFS program. Several hydrophobic regions were predicted by the program PEPTIDE STRUCTURE (Kyte & Doolittle, 1982) to occur within the protein sequence at amino acids 5–23, 295–312, 318–338, 409–434, 907–927 and 1035–1070, as expected for a precursor to the G1 and G2 glycoproteins. These hydrophobic regions may be involved in anchoring the individual glycoproteins in the virus envelope (Ronnholm & Pettersson, 1987). Interestingly, the cell attachment site (RGD) present in the TSWV and INSV glycoprotein precursor (Kormelink *et al.*, 1992; Law *et al.*, 1992) was not observed in PBNV.

The GP protein of PBNV was 14 amino acids shorter and 11 amino acids longer than that of TSWV and INSV, respectively (Kormelink *et al.*, 1992; Law *et al.*, 1992). The amino acid sequence of PBNV GP protein showed 37% identity and 58–59% similarity with that of TSWV and INSV (Table 1). These values are in contrast to 65% identity and 79% similarity observed between GP proteins of TSWV and INSV. The GP protein sequence of serogroup II and other serogroup IV members is not available for comparison. Furthermore, a search of the EMBL protein database revealed sequence similarity to the glycoprotein precursor encoded by the M RNA of snowshoe hare (22%), La Crosse (23%), Germiston (21%) and Bunyamwera (21%) viruses of the genus *Bunyavirus* (Eshita & Bishop, 1984; Lees *et al.*, 1986; Grady *et al.*, 1987; Pardigon *et al.*, 1988) (Table 1).

The small ORF was located in the virus-sense strand. It had an AUG codon at nucleotide 57 and terminated with an UAA stop codon at nucleotide 980, encoding a protein of 307 amino acids with a predicted molecular mass of 34·2 kDa (Fig. 1). It had no detectable hydrophobic region that could function as a signal sequence or a transmembrane segment, based on analysis by the PEPTIDE STRUCTURE program. Seven *N*-glycosylation sites were predicted by the MOTIFS program in the polypeptide. However, it is not known whether this protein is glycosylated *in vivo*. The NSm protein of TSWV is reported to be involved in tubule-guided cell-to-cell virus translocation (Kormelink *et al.*, 1994). The protein was five and four amino acids longer and showed 41–43% identity and 60–63% similarity with the NSm proteins of TSWV and INSV, respectively. In contrast, 69% identity and 79% similarity were observed between NSm proteins of TSWV and INSV. Based on sequence homology, it was concluded that the ORF present in the virus-sense strand codes for the NSm protein.

The two M RNA ORFs were separated by a 408 nucleotides A+U-rich intergenic region (Fig. 1). Similar A+U-rich non-coding regions occur in PBNV S RNA and S and M RNAs of TSWV and INSV (de Haan *et al.*, 1990; Kormelink *et al.*, 1992; Law *et al.*, 1991, 1992; Satyanarayana *et al.*, 1996). The computer-generated model predicted this region would form a stem–loop structure between positions 990 and 1368 (numbered from the 5′ end of the RNA) with a calculated free energy of −316·0 kJ/mol (Zucker & Stiegler, 1981). This is comparable to the intergenic regions of TSWV and INSV M RNAs, which have a calculated free energy of −452·8 kJ/mol and −339·8 kJ/mol, respectively (Kormelink *et al.*, 1992; Law *et al.*, 1992). To detect defective RNAs in PBNV, field-collected PBNV was serially passaged for six generations in peanut plants by mechanical inoculation as described by Reddy *et al.* (1992). Virus RNAs were isolated from purified nucleocapsids (Satyanarayana *et al.*, 1996) and analysed by electrophoresis under denaturing conditions (Bailey & Davidson, 1976). After transfer to nylon membranes, the RNA segments were hybridized to ^{32}P-labelled DNA probes corresponding to the GP and NSm genes of M RNA, the N gene of S RNA (Satyanarayana *et al.*, 1996) and the 3′ half of the RNA polymerase gene of PBNV L RNA (S. Gowda, T. Satyanarayana, D. Reddy & W. Dawson, unpublished results).

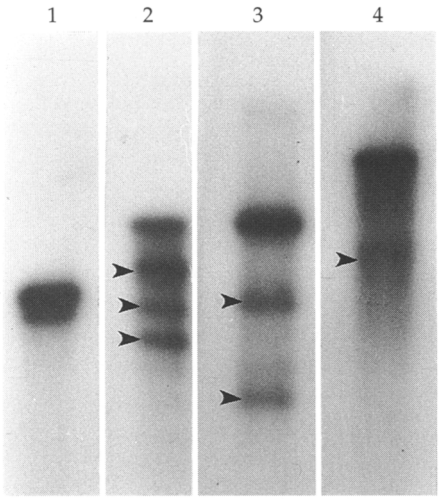

Fig. 2. Northern blot analysis of PBNV M RNA purified from nucleocapsids using double-stranded cDNA probes specific to the nucleocapsid (lane 1), glycoprotein precursor (lane 2), non-structural gene of M RNA (lane 3) and 3′ half of the RNA polymerase (lane 4) genes of PBNV. The positions of the defective RNAs are marked with arrow heads.

The ^{32}P-labelled GP and NSm probes hybridized with the M RNA and RNA species of 3350, 2600 and 2100, and 2600 and 1400 nucleotides, respectively, in plants in which PBNV was serially passaged for six generations (Fig. 2). These were presumed to be defective RNA molecules originating from the M RNA. The L RNA-specific probe detected the L RNA and a 4000 nucleotide truncated L RNA species (Fig. 2). In contrast, the N gene probe detected the S RNA but no smaller RNA molecules. This clearly showed the presence of large molecules of defective M RNAs in PBNV that had been serially passaged for six or more transfers. Mechanically transferred TSWV-BR 01 isolate contained the truncated L RNA species of different sizes in addition to the full-length L RNA (Resende *et al.*, 1991). However, no major abnormalities were detected with S and M RNA-specific probes in the genome of this isolate. Point mutations or very small deletions have been reported to occur in the M RNA from mechanically passaged TSWV in tobacco (Resende *et al.*, 1991).

In conclusion, the observed low sequence identity (37%) between the GP proteins of PBNV and TSWV and INSV, together with the differences in the serology and amino acid sequences of the N protein between PBNV and tospoviruses in serogroups I, II and III (Reddy *et al.*, 1992; Satyanarayana *et al.*, 1996) support the inclusion of PBNV in serogroup IV, albeit as a distinct member. Although we presume close relationships among the glycoprotein precursor and NSm proteins of tospoviruses in serogroup IV, this will become apparent only when the M RNA sequence of other members is available.

The authors wish to thank C. Jester (USDA Plant Genetic Resources Conservation Unit, Griffin, Ga., USA) for help in sequencing. This work was partly supported by a grant from Peanut CRSP (DAN 4048-G-00-0041-00) to the University of Georgia and is submitted as Journal Article Number JA 1902 by ICRISAT.

References

Adam, G., Yeh, S. D., Reddy, D. V. R. & Green, S. K. (1993). The serological comparison of tospovirus isolates from Taiwan and India with *Impatiens* necrotic spot virus and different tomato spotted wilt virus isolates. *Archives of Virology* **130**, 237–250.

Bailey, J. M. & Davidson, N. (1976). Methyl mercury as a reversible denaturing agent for agarose gel electrophoresis. *Analytical Biochemistry* **70**, 75–85.

de Ávila, A. C., de Haan, P., Kormelink, R., Resende, R. de O., Goldbach, R. W. & Peters, D. (1993). Classification of tospoviruses based on phylogeny of nucleoprotein gene sequences. *Journal of General Virology* **74**, 153–159.

de Haan, P., Wagemakers, L., Peters, D. & Goldbach, R. (1990). The S RNA segment of tomato spotted wilt virus has an ambisense character. *Journal of General Virology* **71**, 1001–1007.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

Elliot, R. M. (1990). Molecular biology of the *Bunyaviridae*. *Journal of General Virology* **71**, 501–522.

Eshita, Y. & Bishop, D. H. L. (1984). The complete sequence of the M RNA of Snowshoe hare bunyavirus reveals the presence of internal hydrophobic domains in the viral glycoprotein. *Virology* **137**, 227–240.

Frohman, M. A., Dush, M. K. & Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts by amplification using a single gene specific nucleotide primer. *Proceedings of the National Academy of Sciences, USA* **85**, 8998–9002.

Grady, L. J., Sanders, M. L. & Campbell, W. P. (1987). The sequence of the M RNA of an isolate of La Crosse virus. *Journal of General Virology* **68**, 3057–3071.

Gubler, U. & Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

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. *Journal of General Virology* **73**, 2795–2804.

Kormelink, R., Storms, M., van Lent, J., Peters, D. & Goldbach, R. (1994). Expression and subcellular location of the NSm protein of tomato spotted wilt virus (TSWV), a putative movement protein. *Virology* **200**, 56–65.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology* **157**, 105–132.

Law, M. D., Speck, J. & Moyer, J. W. (1991). Nucleotide sequence of the 3′ non-coding region and N gene of the S RNA of a serologically distinct tospovirus. *Journal of General Virology* **72**, 2597–2601.

Law, M. D., Speck, J. & Moyer, J. W. (1992). The M RNA of *Impatiens* necrotic spot tospovirus (*Bunyaviridae*) has an ambisense genomic organization. *Virology* **188**, 732–741.

Lees, J. F., Pringle, C. R. & Elliot, R. M. (1986). Nucleotide sequence of the Bunyamvera virus M RNA segment: conservation of structural features in the bunyavirus glycoprotein gene product. *Virology* **148**, 1–14.

Pang, S. Z., Slightom, J. L. & Gonsalves, D. (1993). The biological properties of a distinct tospovirus and sequence analysis of its S RNA. *Phytopathology* **83**, 728–733.

Pardigon, N., Vialat, P., Gerbaud, S., Girard, M. & Bouloy, M. (1988). Nucleotide sequence of the M segment of Germiston virus: comparison of the M gene product of several bunyaviruses. *Virus Research* **11**, 73–85.

Reddy, D. V. R., Ratna, A. S., Sudarshana, M. R., Poul, F. & Kirankumar, I. (1992). Serological relationships and purification of bud necrosis virus, a tospovirus occurring in peanut (*Arachis hypogaea* L.) in India. *Annals of Applied Biology* **120**, 279–286.

Resende, R. de O., de Haan, P., de Ávila, A. C., Kitajima, E. W., Kormelink, R., Goldbach, R. & Peters, D. (1991). Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. *Journal of General Virology* **199**, 2375–2383.

Ronnholm, R. & Pettersson, R. F. (1987). Complete nucleotide sequence of the M RNA segment of Uukuniemi virus encoding the membrane glycoproteins G1 and G2. *Virology* **160**, 191–202.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory.

Satyanarayana, T., Mitchell, S. E., Reddy, D. V. R., Brown, S., Kresovich, S., Jarret, R., Naidu, R. A. & Demski, J. W. (1996). Peanut bud necrosis tospovirus S RNA: complete nucleotide sequence, genome organization and homology to other tospoviruses. *Archives of Virology* **141**, 85–98.

Yeh, S. D. & 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.

Zucker, M. & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* **9**, 133–148.

Received 18 January 1996; Accepted 29 April 1996