Volume 306, number 1, 27–32 FEBS 11265 July 1992

The nucleotide sequence of the S RNA of *Impatiens* necrotic spot virus, a novel tospovirus

Peter de Haan[a], Antonio C. de Avila[a], Richard Kormelink[a], Annemarie Westerbroek[b], Jan J.L. Gielen[b], Dick Peters[a] and Rob Goldbach[a]

[a]*Department of Virology, Agricultural University Wageningen, Binnenhaven 11, PO Box 8045, 6700 EM Wageningen, The Netherlands* and [b]*Zaadunie B.V., Department of Biotechnology, Westeinde 62, PO Box 26, 1600 AA Enkhuizen, The Netherlands*

Received 30 March 1992; revised version received 19 May 1992

Impatiens necrotic spot virus (INSV) shares a number of properties with tomato spotted wilt virus (TSWV), the type species of the genus tospovirus within the family Bunyaviridae. INSV, however, differs from TSWV in plant host range and serology. In order to define the genomic structure and the taxonomic status of this TSWV-like virus, the nucleotide sequence of its genomic S RNA segment has been determined. The molecular data obtained demonstrate that, like TSWV, INSV has an ambisense S RNA molecule, encoding a non-structural protein in viral sense and the nucleocapsid protein in viral complementary sense. The level of nucleotide sequence homology between their S RNAs, as well as the divergence in amino acid sequence homology of their gene products, confirm previous conclusions from serological studies that INSV and TSWV represent distinct virus species within the newly created genus, tospovirus.

Impatiens necrotic spot virus; cDNA cloning; S RNA sequence; Ambisense gene arrangement; Nucleocapsid protein; Non-structural protein

1. INTRODUCTION

On the basis of its unique properties tomato spotted wilt virus (TSWV) has been classified as the single representative of the genus, tospovirus, within the family, Bunyaviridae [1]. TSWV occurs over the whole world, infecting many important crops and ornamental plants [2,3]. The virus is persistently transmitted by thrips. So far, eight species have been reported to serve as natural vectors [4,5].

Virus particles (80–110 nm in diameter) consist of nucleocapsids surrounded by lipid envelopes, which are covered with membrane glycoprotein spikes [6,7]. Purified virus preparations contain four proteins, the nucleocapsid (N) protein, two membrane glycoproteins (G1 and G2) and a large (L) protein, present in minor amounts and which most likely represents the viral transcriptase [8,9].

In cells of infected plants, TSWV particles accumulate in clusters within the rough endoplasmatic reticulum. In addition to mature virus particles, free nucleocapsid aggregates can be found in the cytoplasm [10], together with fibrous structures consisting of aggregates of NSs protein, a non-structural protein encoded by the viral genome [11].

The genome consists of three linear RNA species, denoted S, M and L RNA [12]. The L RNA (8.9 kb) is of negative polarity and encodes the putative RNA polymerase from a genome-sized mRNA molecule [13]. The S RNA segment has an ambisense gene arrangement and codes for the N protein in viral complementary (vc) sense and the non-structural (NSs) protein in viral (v) sense [14]. The proteins encoded by the ambisense RNA molecule are expressed via subgenomic mRNAs, which are terminated in the U-A rich intergenic regions [15].

Recently, a second TSWV-like virus has been reported which possessed a serologically distinct nucleocapsid protein. The membrane glycoproteins of this new isolate, however, share many epitopes with those of TSWV [16,17]. Initially, this virus was considered to represent a distinct isolate of TSWV and was tentatively denoted TSWV-I [16]. On the basis of serology and its symptoms induced on *Impatiens* sp. however, it was suggested to represent a distinct tospovirus and was later renamed *Impatiens* necrotic spot virus (INSV) [17,18]. The occurrence of INSV is restricted so far to Northern America, Western Europe and the Mediterranean countries, where infections of exclusively ornamental plant species have been reported.

To verify whether INSV indeed represents a distinct tospovirus species and not another strain of TSWV, i.e. for proper classification of both viruses, nucleotide sequence information will be required. In this regard, determination of the nucleotide sequence of INSV S RNA is of most interest, since this genomic RNA segment encodes the major determinants for bunyavirus classifi-

Correspondence address: P. de Haan, Department of Virology, Agricultural University Wageningen, Binnenhaven 11, PO Box 8045, 6700 EM Wageningen, The Netherlands.

cation, i.e. the structural N protein, as well as a non-structural protein, possibly involved in virulence [11].

In this paper the nucleotide sequence of the INSV S RNA is reported and both its sequence and genetic organization compared with those of TSWV S RNA.

2. MATERIALS AND METHODS

2.1. *Viruses and plants*

Impatiens necrotic spot virus, isolate NL-07, a Dutch isolate from *Impatiens* spp. [18], was propagated in *Nicotiana benthamiana* (L.) plants by mechanical inoculation. Viral nucleocapsids were purified according to Avila et al. [19]. RNA was extracted from purified nucleocapsids as described previously [14].

2.2. *Synthesis and cloning of cDNA*

Purified INSV RNA was polyadenylated using poly-(A) polymerase (Gibco-BRL) according to Devos et al. [20] and cDNA was synthesized according to Gubler and Hoffman [21] primed by a synthetic oligonucleotide complementary to the first eight nucleotides at the 3′ ends of the genomic RNAs (5′-GGGATCCTTTTTTTTTTAGAG-CAAT-3′), which are conserved among all isolates studied so far. Double-stranded cDNA was made blunt-ended, digested with *Bam*HI and cloned in *Bam*HI- and *Sma*I-digested pUC19 plasmid vectors [22]. A second set of cDNA clones was obtained using random hexamers to prime cDNA synthesis. Clones specific for INSV S RNA were selected by Northern blot hybridization. Inserts of identified recombinant plasmids were subsequently used as probes to select more S RNA-specific cDNA clones in colony hybridization experiments [23].

2.3. *Nucleotide sequence analysis*

The nucleotide sequences of S RNA-specific cDNA clones were determined according to Zhang et al. [24] with alkaline-denaturated plasmid DNA as template, using either the standard M13 forward and reverse sequencing primers or synthetic oligonucleotides complementary to previously determined sequences. All data were compiled and edited using the GCG Wisconsin software package [25].

3. RESULTS

3.1. *Molecular cloning of the INSV S RNA*

Two libraries of cDNA clones were obtained using either random primers or a specific primer complementary to the 3′ ends of the genomic RNAs. One cDNA clone was selected with an insert of approximately 860 nucleotides, which contained the complete 3′ terminus of INSV S RNA. Using the insert of this and other clones as probes, additional S RNA-specific cDNA clones were identified and a set of clones could be aligned which almost completely covered the INSV S RNA.

3.2. *Nucleotide sequence analysis*

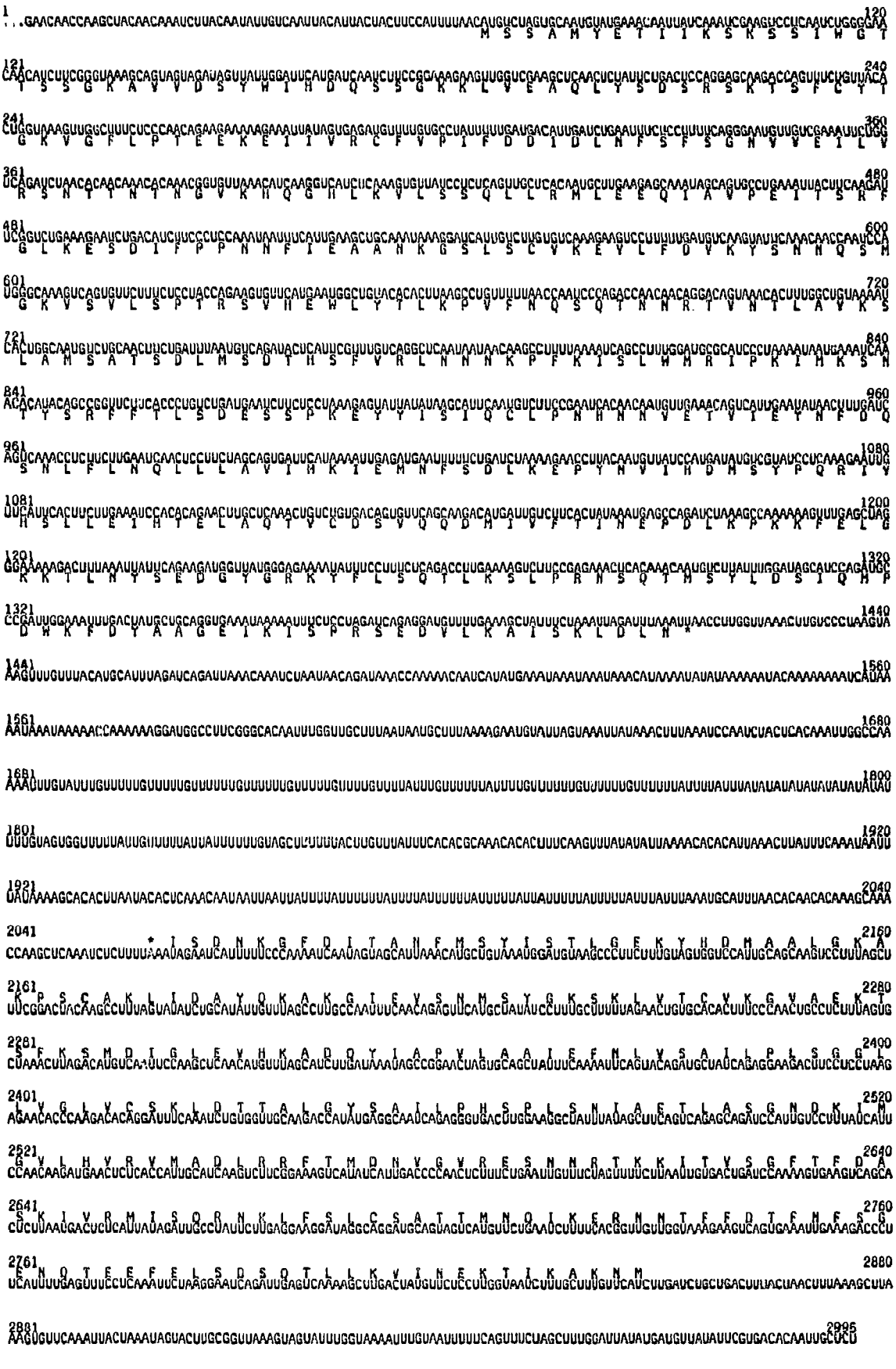
Sequence analysis of the aligned cDNA clones revealed a contiguous nucleotide sequence of 2,992 residues. Analysis of the six possible reading frames in the v and vc RNA strand demonstrated that, as for TSWV S RNA, INSV S RNA has an ambisense character (Fig. 1). Alignment of the INSV and TSWV S RNA sequences showed 55% overall nucleotide sequence homology (identity) (Fig. 2). This alignment furthermore indicated that the nucleotide sequence homologies in the 5′ and 3′ non-translated regions are relatively low (50.7 and 38.7%, respectively). A characteristic property of RNA molecules of negative-strand viruses is the presence of complementary terminal sequences. The termini can be folded in panhandle structures, which are involved in the formation of circular nucleocapsids [2,26,27]. For TSWV these panhandle structures are approximately 60–70 nucleotides in length. Folding of the sequenced 3′- and 5′-terminal regions of INSV S RNA revealed that this terminal structure would be much less extensive, comprising only 6 basepairs with 3′-terminal overhang of 24 nucleotides. This finding suggests that 24 nucleotides at the 5′ end of INSV S RNA were missing in the selected cDNA clones and that the actual length of the panhandle sequence in INSV S RNA is approximately 30 nucleotides long. The S RNA of INSV is most likely approximately 100 nucleotides longer than the S RNA of TSWV isolate BR-01, but only 20 nucleotides longer than that of a Bulgarian TSWV isolate [28]. Inspection of the various S RNAs revealed that differences in lengths of the A-U rich intergenic regions are responsible for the variable sizes of the S RNA molecules of different tospoviruses. Folding of the inter-cistronic region of INSV S RNA revealed a relatively complex secondary structure, compared with the stable simple 'hairpin' structure in the intergenic region of the S RNA of TSWV isolate, BR-01. Between the U- and A-rich tracks, in the centre of the intergenic region that is at the top of the secondary structure, a conserved nucleotide sequence, CAAUUUGG, can be designated. This conserved sequence might be involved in termination of transcription as suggested previously [14,15].

3.3. *The nucleocapsid (N) protein gene*

The gene located on the vc RNA strand of the INSV S RNA, ranging from position 2,058 to 2,845 (numbered from the 5′ end of the v RNA), is 787 nucleotides in length and codes for a protein of 262 amino acids with a predicted molecular mass of 28,734 (Fig. 1). The amino acid sequence shares 55.4% identity to the TSWV N sequence (Fig. 2), with several stretches of homologous residues spread along the entire sequence. On the

→

Fig. 1. The nucleotide sequence of the S RNA of *Impatiens* necrotic spot virus (numbered from the 5′ end) and the predicted gene products. The deduced amino acid sequence of the NSs protein encoded by the viral RNA is written below the RNA sequence, whereas that of the N protein encoded by the viral complementary RNA is written above the RNA sequence. The three dots at the 5′ end of the RNA represent the missing 5′-terminal nucleotides. The asterisks (*) indicate termination codons.



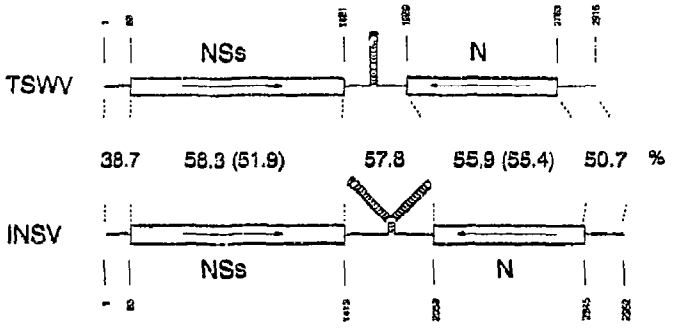


Fig. 2. Comparison of the genomic S RNA molecules of tomato spotted wilt virus and *Impatiens* necrotic spot virus (numbered from the 5′ end of the v RNAs). The drawing is approximately to scale, except for both intergenic regions, which are slightly enlarged to emphasize the interviral differences. The arrows in both ORFs indicate the polarity of these genes on the v and vc strands of both S RNAs. The percentages refer to nucleotide sequence homology between the various non-coding and coding domains. The amino acid sequence homology of the NSs and N proteins of both tospoviruses are indicated between brackets. The black markers at the top of the hairpin loops indicate the conserved sequence motifs (CAAUUUGG), which might be involved in termination of transcription.

basis of the high sequence homology this gene was identified as that encoding the viral N protein. The predicted size is in agreement with its faster migration in SDS-PAA gels compared to the TSWV N protein [17]. The N proteins of INSV isolates, NL-07, from the Netherlands and 'I type' from the USA [18] are almost completely identical (except for one leucine/proline exchange at position 170), which indicates that both isolates indeed belong to the same distinct virus species (Fig. 3).

3.4. *The non-structural (NSs) protein gene*

A second gene located on the v strand of the S RNA segment, ranging from position 65 to 1,413 and with a length of 1,347 nucleotides, encodes a protein of 449 amino acids with a predicted molecular mass of 51,197 (Fig. 1). The amino acid sequence shares 51.9% identity with the NSs sequence of TSWV (Fig. 2), which is significantly lower than the amino acid sequence homology between both N proteins. Also, between two isolates of TSWV, e.g. BR-01 and L3, the variability in the NSs sequence is higher than in the N proteins, indicating that among tospoviruses the NSs protein is less conserved than the N protein (Fig. 3).

4. DISCUSSION

Overlapping cDNA clones have been aligned to almost completely cover the smallest (S) genomic RNA segment of INSV. Nucleotide sequence determination revealed that the 3′ end was completely present in one of the cDNA clones, while most likely only 24 nucleotides at the 5′ end were missing. Since we were mainly interested in the coding capacity of the INSV S RNA, no further attempts were undertaken to determine the missing 5′-terminal nucleotides. Assembly of the sequences revealed 2,992 nucleotides, which is comparable with the sizes of S RNA molecules of two TSWV isolates. Similar to TSWV S RNA, the small genomic RNA of INSV has an ambisense character, with the NSs gene on the v strand and the N gene on the vc strand.

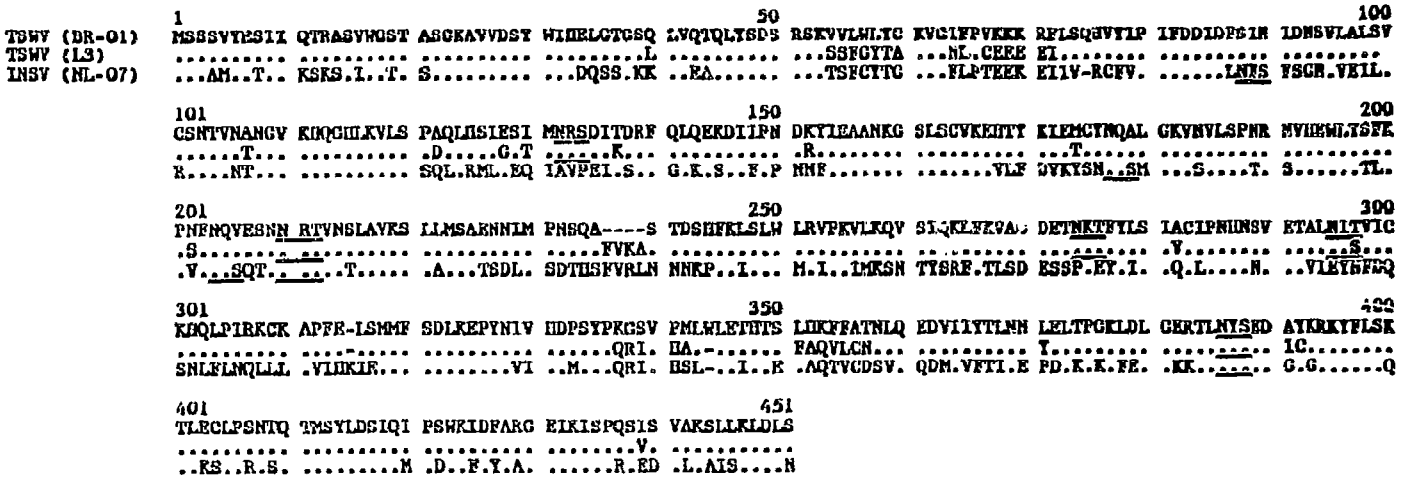
For TSWV S RNA it has been suggested that, based on the lengths of both mRNAs, their 3′ ends map in the centre of the intergenic region [14]. The A-U rich intercistronic domains of both TSWV and INSV S RNA can be folded into extensive and stable secondary structures. Close inspection of the central parts of these structures revealed the presence of a conserved sequence motif, CAAUUUGG (Fig. 2). It can thus be assumed that either the secondary structures (of the template RNA or mRNA) or this conserved sequence represents the signal for termination of transcription. This signal is somehow ignored when replication takes place and genome-sized RNA molecules are synthesized. In this regard, the S RNA molecules of TSWV and INSV resemble those of Punta toro virus, Uukuniemi virus and segments 3 and 4 of tenuiviruses [29–32]. In contrast, the ambisense S RNAs of Rift Valley fever virus, Toscana, Sicilian sandfly fever virus and the ambisense S and L RNAs of arena viruses have G-C rich intergenic domains which cannot be folded into stable secondary structures [33–36]. Mapping of the 3′ ends of both S RNA-directed mRNAs is needed to precisely characterize the transcriptional termination signals and to gain more insight into the regulation of transcription and replication of ambisense RNA molecules.

It has been demonstrated for other negative-strand viruses that the terminal nucleotides are conserved and contain the essential signals for replication and encapsidation with N protein [37–39]. The overall nucleotide sequence homology between the S RNAs of TSWV and INSV is 55%. The sequence homology is lower in the 5′-and 3′-untranslated regions (38.7 and 50.7%, respectively) where a complete match is only found between the 8 terminal nucleotides.

The N protein of INSV is slightly smaller than that of TSWV, as reported previously [18]. Both proteins share several stretches of homologous amino acid sequences. The INSV N protein, however, does not cross-react with antisera to the N protein of TSWV, which might suggest that the homologous parts of the proteins are located internally and that the variable domains, containing the epitopes, are located at the surface of the nucleocapsids.

The NSs protein is less conserved among both viruses (Fig. 3). This protein aggregates as paracrystalline flexuous or rigid rods in the cytoplasm of infected cells. It has been suggested that the amounts and/or structure of these aggregates play a role in determining the severity of the symptoms induced by the corresponding to-

PANEL A



PANEL B

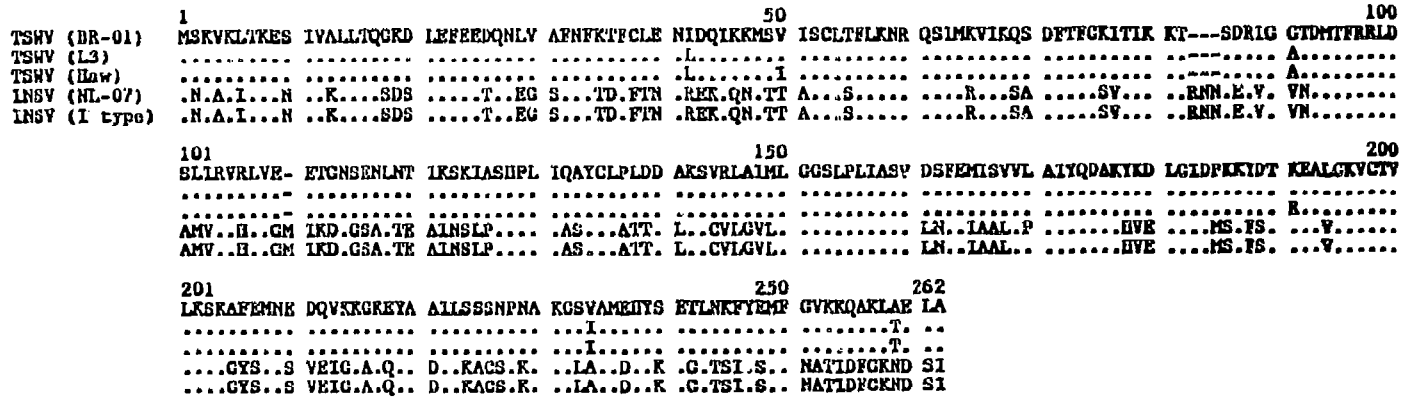


Fig. 3. Alignment of the NSs (Panel A) and N (Panel B) proteins of several isolates of tomato spotted wilt virus and *Impatiens* necrotic spot virus. Data were obtained from [14,18,28] and the EMBL database acc. X61799. The putative *N*-glycosylation sites are underlined. The dots represent identical amino acids. Gaps (–) were introduced to reach an optimal alignment.

spovirus isolate [10,11]. In immunogold labelling experiments, INSV NSs aggregates are decorated using an antiserum to the NSs protein of TSWV. This might indicate that the homologous parts in both TSWV and INSV NSs contain most of the epitopes and are exposed at the surface. In the case of INSV (isolates NL-07 and I-type) large amounts of crystalline NSs aggregates can be found in infected plant cells [17,40]. In contrast, in TSWV (isolate BR-01)-infected cells, relatively small amounts of flexuous fibers are present [10]. More studies are required to predict tertiary structures from other data, including primary amino acid sequences.

The obtained sequence data demonstrate that the amino acid sequence homology between the N proteins of TSWV and INSV is only 55.4%. This rate of divergence parallels that found among the structural proteins of different potyviruses, which are considered to represent distinct species within this plant virus family [41]. In addition, distinct comovirus species share approximately 55% amino acid sequence homology in their coat proteins [42], a level of homology again comparable to that of the N proteins of TSWV and INSV. In view of this, together with the differences in host range and serology [16,18], the conclusion can be drawn that INSV represents a second distinct virus species within the genus, tospovirus, of the family, Bunyaviridae [1].

REFERENCES

[1] Francki, R.I.B., Fauquet, C.M., Knudson, D.D. and Brown, F. (1991) Arch. Virol. Suppl. 2, 1–450.

[2] Peters, D., Avila, A.C. de, Kitajima, E.W., Resende, R. de O., De Haan, P. and Goldbach, R. (1991) In: Virus–Thrips–Plant Interactions of TSWV (H.-T. Hsu and R.H. Lawson, eds.) pp. 1–14, Proc. USDA Workshop, Beltsville, MD, USA.

[3] De Haan, P. (1992) In: Encyclopedia of Virology (R.G. Webster and Granoff eds.) (in press) Saunders Scientific Publications, London.
[4] Sakimura, K. (1962) In: Biological Transmission of Disease Agents, (K. Maramorosch ed.) pp. 33–40, Academic Press, NY.
[5] Paliwal, Y.C. (1974) Can. J. Bot. 52, 1177–1182.
[6] Kitajima, E.W. (1965) Virology 26, 89–99.
[7] Milne, R.G. (1970) J. Gen. Virol. 6, 267–276.
[8] Mohamed, N.A., Randles, J.W. and Francki, R.I.B. (1973) Virology 56, 12–21.
[9] Tas, P.W.L., Boerjan, M.L. and Peters, D. (1977) J. Gen. Virol. 36, 267–279.
[10] Kitajima, E.W., Avila, A.C. de, Resende, R. de O., Goldbach, R.W. and Peters, D. (1992) J. Submicrosc. Cytol. Pathol. 24, 1–14.
[11] Kormelink, R., Kitajima, E.W., De Haan, P., Zuidema, D., Peters, D. and Goldbach, R. (1991) Virology 181, 459–468.
[12] De Haan, P., Wagemakers, L., Goldbach, R. and Peters, D. (1989) In: Genetics and Pathogenicity of Negative-Strand Viruses, (D. Kolakofski and B.W.J. Mahy eds.) pp. 287–290, Elsevier, Amsterdam.
[13] De Haan, P., Kormelink, R., Resende, R. de O., Van Poelwijk, F., Peters, D. and Goldbach, R. (1991) J. Gen. Virol. 72, 2207–2216.
[14] De Haan, P., Wagemakers, L., Peters, D. and Goldbach, R. (1990) J. Gen. Virol. 71, 1001–1007.
[15] Kormelink, R., De Haan, P., D. and Goldbach, R. (1992) J. Gen. Virol. 73, 687–693.
[16] Law, M.D. and Moyer, J.W. (1990) J. Gen. Virol. 71, 933–938.
[17] Law, M.D., Speck, J. and Moyer, J.W. (1991) J. Gen. Virol. 72, 2597–2601.
[18] Avila, A.C. de, De Haan, P., Kitajima, E.W., Kormelink, R., Resende, R. de O., Goldbach, R.W. and Peters, D. (1992) J. Phytopathol. 134, 133–151.
[19] Avila, A.C. de, Huguenot, C., Resende, R. de O., Kitajima, E.W., Goldbach, R.W. and Peters, D. (1990) J. Gen. Virol. 71, 2801–2807.
[20] Devos, R.G.E., Gillis, E. and Fiers, W. (1976) Eur. J. Biochem. 62, 401–410.
[21] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
[22] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119
[23] Grunstein, J.M. and Hogness, D.S. (1975) Proc. Nat. Acad. Sci. USA 72, 3961–3965.
[24] Zhang, H., Scholl, R., Browse, J. and Somerville, C. (1988) Nucleic Acids Res. 16, 1220.
[25] Deveraux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[26] Elliott, R.M. (1990) J. Gen. Virol. 71, 501–522.
[27] Bouloy, M. (1991) Adv. Virus Res. 40, 235–275.
[28] Maiss, E., Ivanova, L., Breyel, E. and Adam, G. (1991) J. Gen. Virol. 72, 461–464.
[29] Emery, V.C. and Bishop, D.H.L. (1987) Virology 156, 1–11.
[30] Simons, J.F. and Petterson, R.F. (1991) J. Virol. 65, 4741–4748.
[31] Kakutani, T., Hayano, Y., Hayashi, T. and Minobe, Y. (1990) J. Gen. Virol. 71, 1427–1432.
[32] Kakutani, T., Hayano, Y., Hayashi, T. and Minobe, Y. (1991) J. Gen. Virol. 72, 465–468.
[33] Auperin, D.D., Romanowski, V., Valinski, M. and Bishop, D.H.L. (1984) J. Virol. 52, 897–904.
[34] Salvato, M.and Shimomaye, E.M. (1989) Virology 173, 1–10.
[35] Iapalucci, S., Lopez, N. and Franze-Fernandez, M.T. (1991) Virology 182, 269–278.
[36] Giorgi, C., Accardi, L., Nicoletti, L., Gro, M.C., Takehara, K., Hilditch, C., Morikawa, S. and Bishop, D.H.L. (1991) Virology 180, 738–753.
[37] Parvin, J.D., Palese, P., Honda, A., Ishihama, A. and Krystal, M. (1989) J. Virol. 63, 5142–5152.
[38] Vidal, S. and Kolakofski, D. (1989) J. Virol. 63, 1951–1958.
[39] Seong, B.J., and Brownlee, G.G. (1992) Virology 186, 247–260.
[40] Urban, L.A., Huang, P.-Y. and Moyer, J.W. (1991) Phytopathology 81, 525–529.
[41] Ward, C.W. and Shukla, D.D. (1991) Intervirology 32, 269–296.
[42] Francki, R.I.B., Fauquet, C.M., Knudson, D.D. and Brown, F. (1991) Arch. Virol. Suppl. 2, 1–450.