Completion of the impatiens necrotic spot virus genome sequence and genetic comparison of the L proteins within the family *Bunyaviridae*

Frank van Poelwijk,† Marcel Prins and Rob Goldbach

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

The nucleotide sequence of the large (L) genome segment of impatiens necrotic spot virus (INSV) has been determined, and herewith the complete nucleotide sequence of the whole, tripartite genome of this important tospovirus has been elucidated. The L RNA is 8776 nucleotides long and of negative polarity, containing one large ORF on the viral complementary strand. Comparison of the deduced amino acid sequence of the INSV L RNA primary translation product (330·3 kDa) with those of the L RNAs of other members of the family *Bunyaviridae* reveals that this protein represents the putative viral RNA-dependent RNA polymerase. A cluster dendrogram of the (putative) RNA polymerases indicates that the genera *Tospovirus* and *Tenuivirus*, though both encompassing ambisense plant-infecting viruses, have different affinities to the animal-infecting *Bunyaviridae*, tospoviruses being most closely related to the genus *Bunyavirus*, and tenuiviruses to the genus *Phlebovirus*.

Most members of the family *Bunyaviridae*, a large family of enveloped, arthropod borne RNA viruses (Elliott, 1990; Murphy *et al.*, 1995), infect animals, but some are able to infect plants. These latter bunyaviruses are classified into a separate genus, *Tospovirus*, named after the type species tomato spotted wilt virus (TSWV). Tospoviruses are exclusively transmitted by thrips in a propagative manner (Wijkamp *et al.*, 1993; Ullman *et al.*, 1993). Owing to the recent worldwide spread of one of the most efficient vectors, the western flower thrips (*Franklinellia occidentalis* Pergande), not only TSWV but also a second tospovirus, impatiens necrotic spot virus (INSV), has recently emerged (Goldbach & Peters, 1994) and is gaining in economic impact. INSV has become a serious threat for the cultivation of ornamental plants both in Northern America and in Europe (Law & Moyer, 1990; DeAngelis *et al.*, 1994; Vaira *et al.*, 1993). Thus far, molecular studies have mainly focussed on TSWV, whereas information on INSV is limited. Both the small (S) and the middle (M) segments of the tripartite RNA genome of INSV have been sequenced (de Haan *et al.*, 1992; Law *et al.*, 1992), but molecular data on the largest (L) genomic segment are lacking. Here we report the nucleotide sequence of the INSV L RNA, thereby providing the sequence of the complete genome of this tospovirus.

Author for correspondence: Rob Goldbach.
Fax +31 317 484 820. e-mail Rob.Goldbach@medew.viro.wau.nl

† **Present address**: Institute for Animal Science and Health (ID-DLO), Research Branch Houtribweg, Houtribweg 39, Postbus 65, 8200 AB Lelystad, The Netherlands.

The EmbI/GenBank/DDBJ accession number for the sequence of INSV L RNA is X93218.

Nucleocapsid preparations were purified from systemically infected *Nicotiana benthamiana* and RNA was extracted as previously described (de Haan *et al.*, 1989). Reverse transcription was carried out at 37 °C with viral (v) or viral-complementary (vc) specific primers and Moloney murine leukaemia virus reverse transcriptase (M-MLV, Gibco BRL). Primer pairs were added to the first strand reaction for both v and vc sense followed by amplification using Taq polymerase (Supertaq, SphaeroQ). Sequence information for the PCR primers was initially obtained from a number of cDNA clones that were obtained by random priming of nucleocapsid RNA. Seven fragments of 421, 593, 1609, 564, 1702, 2071 and 200 nucleotides (nt), respectively, were amplified and cloned into a pGEM-T vector (Promega). The first nine terminal nucleotides of all tospovirus genome segments sequenced to date are conserved (vRNA: 3′ UCUCGUUAG 5′). This feature allowed us to design a primer that could be used to amplify fragments from both the 5′ and 3′ ends. The cloned PCR fragments were selected by their size, subjected to dideoxynucleotide sequencing and run on an automatic (ABI 373A) DNA sequencing system.

The complete nucleotide sequence of the L RNA has been determined from either cDNA- or PCR-derived clones. Both strands were sequenced and the sequence is available from the EMBL, GenBank and DDBJ databases under the accession number X93218. The INSV L segment is 8776 nt long and has 37% A, 30% U, 14% C and 19% G content. The segment

0001-4287

	MOTIF A	MOTIF B
INSV	**K**SKLAFL**SAD**QS**KWS**	STNSYPVSM**NWLQGNLNY**L**SS**VY**H**
TSWV	**K**SRLAFL**SAD**QS**KWS**	TTNTYPVSM**NWLQGNLNY**L**SS**VY**H**
La Crosse	**K**GLKMEIN**AD**MS**KWS**	A**Q**.**D**VFY**K**Y**NWLQGN**F**NY**T**SS**Y**V**H**
Bunyamwera	**K**ALKLEIN**AD**MS**KWS**	A**Q**.**D**VFY**K**Y**NWLQGN**F**NY**I**SS**Y**V**H**
RVFV	PVWTCAT**SDDAR**K**W**N	.**Q**G**H**FV**T**K**F**GMM**QG**IL**H**Y**T**SS**LLH**
Toscana	SVWTCAT**SDDAR**K**W**N	.**Q**G**H**YV**T**K**F**GMM**QG**IL**H**F**T**SS**LLH**
Uukuniemi	HHETVAT**SDDAA**K**W**N	.**Q**CHHV**T**K**F**GMM**QG**IL**H**Y**T**SS**LLH**
Hantaan	**K**RKLMYV**SAD**AT**KWS**	P.**GD**NSA**K**F**NWLQGNLN**KC**SSL**FG
Seoul 80-39	**K**RKLMYV**SAD**AT**KWS**	P.**GD**NSA**K**F**NWLQGNLN**KC**SSL**FG
Puumala	**K**RKLMYV**SAD**AT**KWS**	P.**GD**NSA**K**F**NWLQGNLN**KC**SSL**FG

	MOTIF C	MOTIF D	MOTIF E
INSV	IV**HSDD**NATSLI	SHFKSFCITLNP**KK**SYAS	SSEV.**EFIS**
TSWV	IV**HSDD**NATSLI	AHFKSFCITLNP**KK**SYAS	SSEV.**EFIS**
La Crosse	LV**HSDD**NQTSIT	**L**TFG**C**.QA..NM**KKT**YVT	NCIK.**EFVS**
Bunyamwera	MV**HSDD**NQTSLA	**L**TFG**C**.QA..NM**KKT**YIT	HTCK.**EFVS**
RVFV	MQG**SDD**SSMLIS	KEL**G**VYLAIYPSE**KST**AN	TDFVM**EYNS**
Toscana	MQG**SDD**SSMIIS	KSL**G**TYIGIYPSE**KST**PN	TDFVM**EYNS**
Uukuniemi	LQS**SDD**SGMMIS	KVI**G**KYLGIYSSV**KST**NN	TLHLL**EFNS**
Hantaan	AH**HSDD**ALFIYG	**LLLG**SIKI..SP**KKTT**VS	PTNA.**EFLS**
Seoul 80-39	AH**HSDD**ALFIYG	**LLLG**SIKI..SP**KKTT**LS	PTNA.**EFLS**
Puumala	AH**HSDD**ALFIYG	**LLMG**SIKI..SP**KKTT**VS	PTNA.**EFLS**

Fig. 1. Amino acid identity between *Bunyaviridae* L proteins. Conserved amino acids were identified using the UW–GAP and PILEUP options of the GCG package from the University of Wisconsin. Conserved residues are in bold (see Fig. 3 for references).

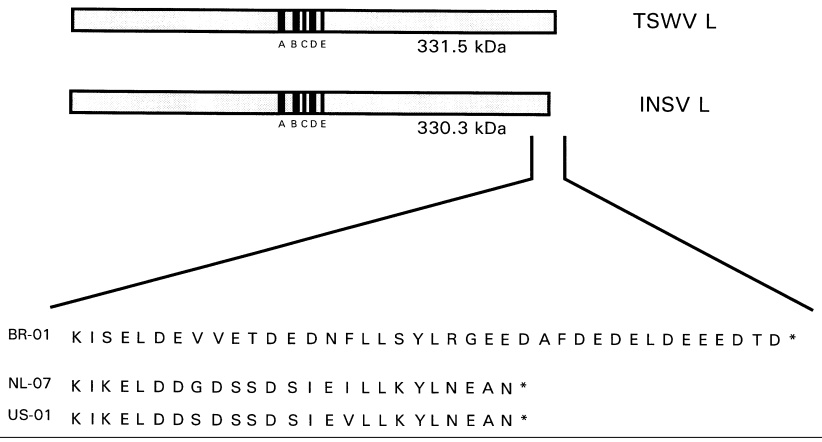

Fig. 2. Comparison of the L ORFs of TSWV (BR-01), INSV (NL-07) and INSV (US-01). Polymerase motifs (A–E) are depicted as bars.

contains one large ORF in the vc strand, starting with an AUG codon at position 8744 (numbered from the 5′ end of the viral RNA) and extends to an UAA stop codon at position 139. This results in a primary translation product of 2865 amino acids with a predicted molecular mass of 330·3 kDa. It contains the sequence motifs which are diagnostic for RNA-dependent RNA polymerases of all negative-stranded RNA viruses (Poch *et al.*, 1989) (Fig. 1).

The 5′ and 3′ ends of INSV L RNA are complementary and can be folded into a panhandle structure, a typical feature of all segmented negative-stranded viruses, which is thought to play a role in transcription/replication. Interestingly, the L RNA sequence of INSV is 124 nt shorter than that of TSWV, resulting in an ORF which is 10 amino acids shorter and lacking the acidic tail (Fig. 2). The significance of this extremely acidic C terminus (five glutamic acid and five aspartic acid residues out of 15, see Fig. 2) in the TSWV L protein (the presence of which was reconfirmed from independent clones), and its absence in that of INSV is not clear yet. It is, however, very likely that the exposure and folding of the two tospoviral polymerases will be rather dissimilar in this region. The acidic tail might be involved in interactions with the (basic) nucleocapsid protein, although the lacking acidic residues do not seem to have an effect on *in vitro* transcription/replication activity (unpublished results). Furthermore, the 5′ non-translated region of INSV L RNA (v strand) is 140 nt long

whereas for TSWV this region comprises 242 nt (a duplication of the AUUU sequence at position 55 was found in the TSWV L sequence). The INSV L sequence was verified by cloning and sequencing several PCR fragments in this region derived from independent amplification experiments. Primer extension analysis of nucleocapsid RNA confirmed that the L RNA of INSV is indeed shorter (data not shown). Sequence analysis of US-01, an American isolate of INSV (Law & Moyer, 1990) revealed the same sequence in this region, including the shortened ORF (Fig. 2) and 5′ end (data not shown).

Comparison of the INSV L segment with that of TSWV revealed 68·9% identity in the nucleotide sequence, whereas at the amino acid level 69·5% identity and 83·6% similarity were found using the GAP function of the UW–GCG package. The alignment of TSWV and INSV L proteins revealed a frameshift in the L ORF of TSWV. Upon resequencing of the original TSWV clone 806 (de Haan *et al.*, 1991), an insertion of a U residue at position 4206 and a deletion of a U residue at position 4129 were found. After restoring these errors the reading frame showed 100% identity in this region. Analysis of all three RNA segments of INSV and TSWV in a dot plot revealed that the L RNA is the most conserved RNA segment, whereas the S RNA is least conserved.

Classification of tospoviruses has been based on serological differences of the S RNA-encoded nucleocapsid protein. The L sequence reported here is determined from a Dutch isolate (NL-07) obtained from infected *Impatiens*, from which the S RNA has also been described (de Haan *et al.*, 1992). The M RNA sequence, however, has been determined from an American isolate (US-01; Law *et al.*, 1992). Parts of the M RNA of NL-07 have been amplified by PCR, cloned and sequenced (results not shown). Out of more than 700 nt sequenced, only four nucleotide changes with the US-01 M RNA sequence were found and these changes did not lead to differences on the amino acid level. This result indicates that the isolates NL-07 and US-01 are almost completely identical.

Comparison of the L proteins of tospoviruses INSV and TSWV with those of animal-infecting members of the *Bunyaviridae* indicates that tospoviruses are most closely related to the genus *Bunyavirus* (Fig. 3). This is an interesting observation as tospoviruses have an ambisense S RNA, a feature they only share with members of the genus *Phlebovirus*. This may suggest that the generation of an ambisense genome segment is a relatively late event during bunyaviral evolution. The evidence that TSWV has been adapted to plant hosts by inclusion of the NS_M gene, in an ambisense arrangement within the M RNA segment, supports this hypothesis (Kormelink *et al.*, 1994; Storms *et al.*, 1995). Comparison of the bunyaviral L proteins with that of rice stripe virus (Toriyama *et al.*, 1994), belonging to the floating genus *Tenuivirus* (Murphy *et al.*, 1995), indicates that tenuiviruses are most closely related to the genus *Phlebovirus* (Fig. 3) and are only distantly related to the tospoviruses. Hence, it seems that these two genera of ambisense plant-infecting viruses have descended from the animal-infecting *Bunyaviridae* by two independent evolutionary pathways. This observation, together with the fact that tenuiviruses have four and sometimes even five genomic segments, indicates that tenui- and tospoviruses, though both representing ambisense plant-infecting RNA viruses, are not to be placed easily in a single virus family.

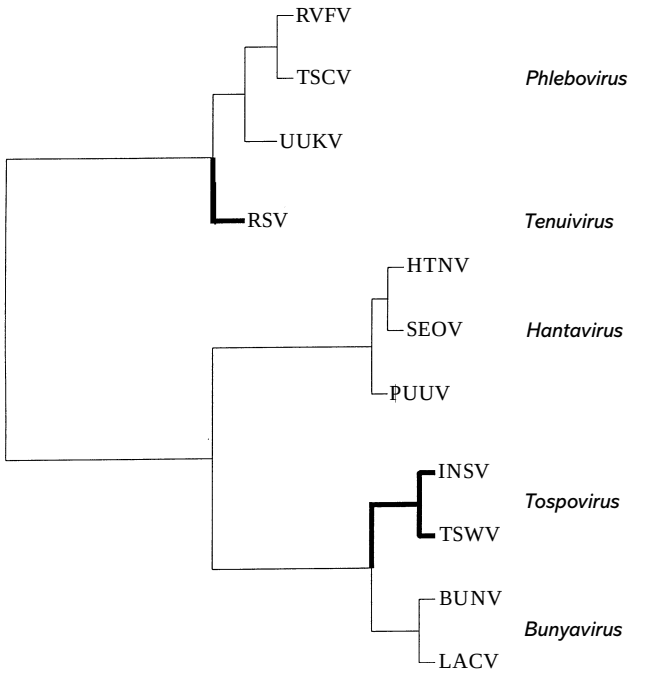

Fig. 3. Phylogenetic dendrogram of bunyaviral L proteins. PILEUP and GROWTREE options of the UW–GCG package were used to construct the phylogenetic tree. RVFV, Rift Valley fever virus; TSCV, Toscana virus; UUKV, Uukuniemi virus; RSV, rice stripe virus; HTNV, Hantaan virus; SEOV, Seoul-89 virus; PUUV, Puumala virus; INSV, impatiens necrotic spot virus; TSWV, tomato spotted wilt virus; BUNV, Bunyamwera virus; LACV, La Crosse virus. Sequence data were obtained from Elliott (1989); Schmaljohn (1990); Antic *et al.* (1991); Stohwasser *et al.* (1991); Müller *et al.* (1991); Elliott *et al.* (1992); Accardi *et al.* (1993); Toriyama *et al.* (1994) and Roberts *et al.* (1995). The most parsimonious tree of 9191 steps was obtained with the Branch & Bound option giving a consistency index of 0·95 and a retention index of 0·88. The rescaled consistency index was 0·84.

The authors would like to thank Peter de Haan, Richard Kormelink and Dick Peters for their continuous interest and stimulating discussions. This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (N.W.O.)

References

Accardi, L., Gro, M. C., Di Bonito, P. & Giorgi, C. (1993). Toscana virus genomic L segment: molecular cloning, coding strategy and amino acid sequence in comparison with other negative strand RNA viruses. *Virus Research* **27**, 119–131.

Antic, D., Lim, B. & Kang, C. Y. (1991). Nucleotide sequence and coding capacity of the large (L) genomic RNA segment of Seoul 80-39 virus, a member of the hantavirus genus. *Virus Research* **19**, 59–66.

Biswas, S. K. & Nayak, D. P. (1994). Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. *Journal of Virology* **68**, 1819–1826.

DeAngelis, J. D., Sether, D. M. & Rossignol, P. A. (1994). Transmission of impatiens necrotic spot virus in peppermint by western flower thrips (Thysanoptera: Thripidae). *Journal of Economical Entomology* **87**, 197–201.

de Haan, P., Wagemakers, L., Peters, D. & Goldbach, R. (1989). Molecular cloning and terminal sequence determination of the S and the M RNAs of tomato spotted wilt virus. *Journal of General Virology* **70**, 3469–3473.

de Haan, P., Kormelink, R., de Oliveira Resende, R., van Poelwijk, F., Peters, D. & Goldbach, R. (1991). Tomato spotted wilt virus L RNA encodes a putative RNA polymerase. *Journal of General Virology* **72**, 2207–2216.

de Haan, P., de Avila, A. C., Kormelink, R., Westerbroek, A., Gielen, J. J. L., Peters, D. & Goldbach, R. (1992). The nucleotide sequence of the S RNA of *Impatiens* necrotic spot virus, a novel tospovirus. *FEBS Letters* **306**, 27–32.

Elliott, R. M. (1989). Nucleotide sequence analysis of the large genomic RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae. *Virology* **173**, 426–436.

Elliott, R. M. (1990). Molecular biology of the Bunyaviridae. *Journal of General Virology* **72**, 501–522.

Elliott, R. M., Dunn, E., Simons, J. F. & Pettersson, R. F. (1992). Nucleotide sequence and coding strategy of the Uukuniemi virus L RNA segment. *Journal of General Virology* **73**, 1745–1752.

Goldbach, R. & Peters, D. (1994). Possible causes of the emergence of tospovirus diseases. *Seminars in Virology* **5**, 113–120.

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

Law, M. D. & Moyer, J. W. (1990). A tomato spotted wilt-like virus with a serologically distinct N protein. *Journal of General Virology* **71**, 933–938.

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

Müller, R., Argentini, C., Bouloy, M., Prehaud, C. & Bishop, D. H. L. (1991). Completion of the genome sequence of Rift Valley fever phlebovirus indicates that the L RNA is negative sense or ambisense and codes for a putative transcriptase-replicase. *Nucleic Acids Research* **19**, 5433.

Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A. & Summers, M. D. (editors), (1995). *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses*. Vienna & New York: Springer-Verlag.

Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerases encoding elements. *EMBO Journal* **8**, 3867–3874.

Roberts, A., Rossier, C., Kolakofsky, D., Nathanson, N. & Gonzalez-Scarano, F. (1995). Completion of the La Crosse virus genome sequence and genetic comparison of the L proteins of the Bunyaviridae. *Virology* **206**, 742–745.

Schmaljohn, C. S. (1990). Nucleotide sequence of the L genome segment of Hantaan virus. *Nucleic Acids Research* **18**, 6728.

Stohwasser, R., Raab, K., Darai, G. & Bautz, E. K. F. (1991). Primary structure of the large (L) RNA segment of nephropathia epidemica virus strain Hällnäs B1 coding for the viral RNA polymerase. *Virology* **183**, 386–391.

Storms, M. M. H., Kormelink, R., Peters, D., van Lent, J. W. M. & Goldbach, R. (1995). The nonstructural NS_M protein of tomato spotted wilt virus induces tubular structures in plant and insect cells. *Virology* **214**, 485–493.

Toriyama, S., Takahashi, M., Sano, Y., Shimizu, T. & Ishihama, A. (1994). Nucleotide sequence of RNA **1**, the largest genomic segment of rice stripe virus, the prototype of the tenuiviruses. *Journal of General Virology* **75**, 3569–3579.

Ullman, D. E., German, T. L., Sherwood, J. L., Westcot, D. M. & 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.

Vaira, A. M., Roggero, P., Luisoni. E., Masenga, V., Milne, R. G. & Lisa V. (1993). Characterization of two tospoviruses in Italy: tomato spotted wilt virus and impatiens necrotic spot virus. *Plant Pathology* **42**, 530–542.

Wijkamp, I., van Lent, J., Kormelink, R., Goldbach, R. & Peters, D. (1993). Multiplication of tomato spotted wilt virus in its insect vector, *Franklinellia occidentalis*. *Journal of General Virology* **74**, 341–349.

Received 22 July 1996; Accepted 5 October 1996