Journal of General Virology (1991), **71**, 2207–2216. *Printed in Great Britain*

Tomato spotted wilt virus L RNA encodes a putative RNA polymerase

Peter de Haan,* Richard Kormelink, Renato de Oliveira Resende, Frank van Poelwijk, Dick Peters and Rob Goldbach

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

The complete nucleotide sequence of the large (L) genome segment of tomato spotted wilt virus (TSWV) has been determined. The RNA is 8897 nucleotides long and contains complementary 3′ and 5′ ends, comprising 62 nucleotides at the 5′ end and 66 nucleotides at the 3′ end. The RNA is of negative polarity, with one large open reading frame (ORF) located on the viral complementary strand. This ORF corresponds to a primary translation product of 2875 amino acids in length, with a predicted M_r of 331 500. Comparison with the polymerase proteins of other negative-strand viruses indicates that this protein most likely represents the viral polymerase. The genetic organization of TSWV L RNA is similar to that of the L RNA segments of Bunyamwera and Hantaan viruses, animal-infecting representatives of the Bunyaviridae.

Introduction

Based on its unique properties among other plant viruses, tomato spotted wilt virus (TSWV) has previously been classified as the single representative of a distinct virus group (Ie, 1970; Matthews, 1982). Recently, molecular data have provided evidence that TSWV should be considered as a member of the arthropod-borne Bunyaviridae, although unique in being able to infect plants (de Haan *et al.*, 1989*a*, *b*, 1990).

Like the established members of the Bunyaviridae (Elliott, 1990), TSWV is characterized by spherical enveloped particles of approximately 80 to 110 nm in diameter. Two virus-encoded glycoproteins, denoted G1 (M_r 78K) and G2 (M_r 58K) are associated with the virus envelope (Tas *et al.*, 1977). The internal pseudo-circular nucleocapsids consist of three species of ssRNA, denoted S RNA (2916 nucleotides), M RNA (approximately 5000 nucleotides) or L RNA (approximately 8000 nucleotides), which are tightly encapsidated with the nucleocapsid (N) protein (M_r 28·8K) (de Haan *et al.*, 1989*b*). In addition a few copies of a large (L) protein (approximately 200K) are present in the virus particle, and may represent the viral polymerase (Mohamed *et al.*, 1973; Mohamed, 1981; Tas *et al.*, 1977).

Recently, the genomic RNA segments have been cloned (de Haan *et al.*, 1989*b*) and the complete nucleotide sequence of the S RNA has been determined from a set of overlapping cDNA clones (de Haan *et al.*, 1990). TSWV S RNA encodes two proteins, the N protein and a non-structural (NSs) protein, in an ambisense gene arrangement. The N protein is expressed from a subgenomic mRNA species of approximately 1·2 kb, transcribed from the viral RNA strand, and the NSs protein (M_r 52·4K) is expressed from an mRNA of approximately 1·7 kb, transcribed from the viral complementary RNA strand. The structure of TSWV S RNA conforms with that of the phleboviruses and uukuviruses, two genera of the family Bunyaviridae (Giorgi *et al.*, 1991).

Here we report the complete nucleotide sequence of TSWV L RNA. It contains a single large open reading frame (ORF) in the viral complementary sense, which most probably corresponds to the viral polymerase gene. The genetic organization of the TSWV L RNA segment further strengthens our previous conclusion that this virus represents a plant-infecting member of the Bunyaviridae.

Methods

Virus and plants. TSWV CNPH1 (now BR-01), a Brazilian isolate from tomato, was maintained in tomato by grafting and infected leaf tissue was stored in liquid nitrogen. *Nicotiana rustica* plants were either mechanically inoculated from this original virus stock, or from previously inoculated, systemically infected *N. rustica*. Virus was purified from infected *N. rustica* leaves according to Tas *et al.* (1977) and RNA was extracted as described previously (de Haan *et al.*, 1989*b*).

Synthesis, cloning and sequence determination of cDNA. cDNA to TSWV RNA was synthesized and cloned as previously described (de Haan *et al.*, 1989*b*). To obtain cDNA clones containing the 3′ end of the L RNA, a 5 μg portion of genomic RNA was polyadenylated at the 3′ end, using 1 unit of poly(A) polymerase (Bethesda Research Laboratories), according to Devos *et al.* (1976). First-strand cDNA synthesis was primed with oligo(dT), followed by second-strand synthesis according

to Gubler & Hoffman (1983). Double-stranded cDNA was made blunt-ended using T4 DNA polymerase and subsequently cloned into the *Sma*I site of plasmid pUC19 (Maniatis *et al.*, 1982).

DNA sequencing was performed by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977), on dsDNA templates (Zhang *et al.*, 1988), or after subcloning of restriction fragments in M13mp18 or -mp19 vectors (Yanisch-Perron *et al.*, 1985). Nucleotide and amino acid sequences were compiled and analysed using programs developed by the University of Wisconsin Genetics Computer Group (UWGCG).

Results

Cloning and sequence determination of the TSWV L RNA

Northern blot analysis of genomic RNA, purified from the original BR-01 virus stock, revealed that the previously reported restriction map of TSWV M RNA (de Haan *et al.*, 1989*b*) actually represented that of a defective L RNA molecule of 4·7 kb in length. This defective RNA molecule was abundantly present in the TSWV BR-01 line used in this study and masked the authentic M RNA segment (5·0 kb). This TSWV line had been maintained by mechanical passage of the virus for many years.

In order to obtain cDNA clones corresponding to the full-length genomic RNA sequence, the original cDNA library (de Haan *et al.*, 1989*b*) was screened again, and additional cDNA clones to TSWV L RNA could be aligned, yielding a restriction map covering approximately 8900 nucleotides (Fig. 1). The cDNA clones denoted 70, 266, 280, 299, 329, 420, 662, 669, 803, 808 and 810 were selected for sequence analysis. Since clones 280, 803, 806 and 808 hybridized only to the full-length L RNA and not to the defective L RNA molecule (results not shown), it can be assumed that the latter molecule is the result of an internal deletion in TSWV L RNA. The nucleotide sequences and origin of defective L RNA species in TSWV isolates will be discussed in a separate paper.

Direct dideoxynucleotide sequencing, using L RNA as a template and four different synthetic oligonucleotides as primers, was used to obtain the 5′-terminal sequence and to verify internal sequences (Fig. 1). To obtain cDNA clones containing the 3′-terminal sequences of the L RNA, genomic RNA was polyadenylated and cDNA was synthesized by priming first-strand cDNA synthesis with oligo(dT). Clones were subsequently selected, using a 830 bp *Eco*RI/*Sph*I restriction fragment of cDNA clone 662 as a probe in a colony hybridization experiment. One of the selected clones, denoted 669, contained the sequence 5′ ..ACCTGATTGCTCT$(A)_{22}$ 3′, which is complementary to the sequence at the 5′ end of the TSWV L RNA (5′ AGAGCAAUC.. 3′), as determined by primer extension sequencing (Fig. 1). These terminal sequences are also identical for the first eight nucleotides to the 3′ and 5′ termini of the S RNA (de Haan *et al.*, 1990), indicating that the entire L RNA sequence was indeed included. The identification of clone 669 as an L RNA-specific cDNA clone was further confirmed by Northern blot hybridization (results not shown).

Characteristics of the TSWV L RNA

The complete nucleotide sequence of the TSWV L RNA is shown in Fig. 2. The RNA is 8897 nucleotides long,

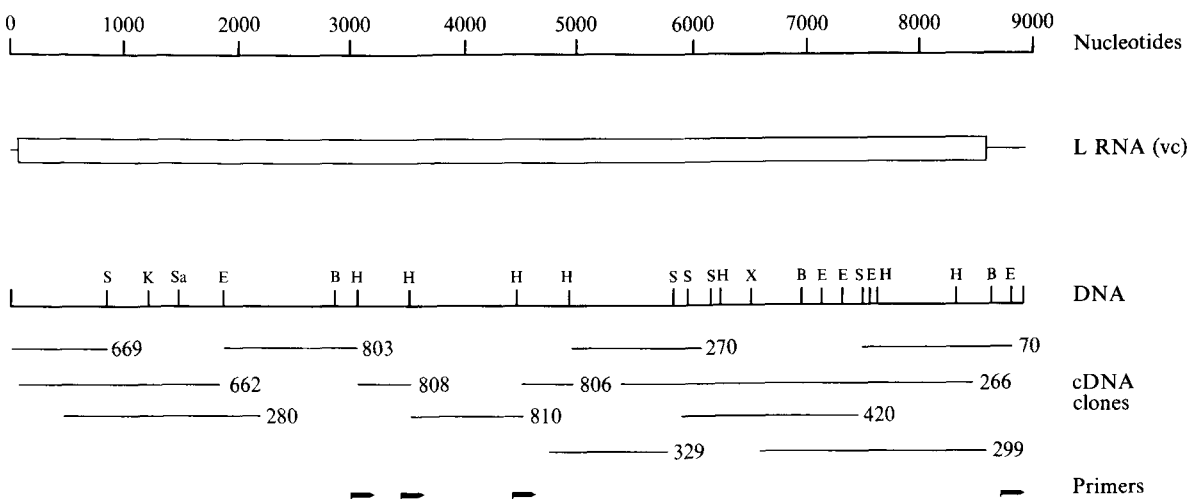

Fig. 1. Cloning strategy for the TSWV L RNA segment. The viral complementary (vc) RNA strand is represented. The box corresponds to the large ORF. The arrows represent the synthetic oligonucleotides used for primer extension sequencing on the L RNA as a template. The numbers correspond with the cDNA clones used. Restriction enzymes are abbreviated as follows: Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sph*I; Ss, *Sst*I; X, *Xba*I.

```
                                                M  N  I  Q  K  I  Q  K  L  I  E  N  G  T  T  L  L  L  S  I  E  D
1    AGAGCAAUCA GGUAACAACG AUUUUAAGCA AACAUGAACA UCCAGAAAAU ACAAAAAUUA AUAGAAAAUG GAACCACUUU ACUGUUGUCU AUUGAGGAUU

     C  V  G  S  N  H  D  L  A  L  D  L  H  K  R  N  S  D  E  I  P  E  D  V  I  I  N  N  N  A  K  N  Y  E
101  GUGUAGGUUC UAACCACGAU CUAGCUUUGG AUUUACAUAA GAGAAAUAGU GAUGAGAUCC CAGAAGAUGU GAUUAUAAAU AAUAAUGCAA AAAAUUAUGA

       T  M  R  E  L  I  V  K  I  T  A  D  G  E  G  L  N  K  G  M  A  T  V  D  V  K  K  L  S  E  M  V  S
201  GACAAUGAGA GAGUUAAUUG UCAAAAUCAC UGCUGAUGGU GAAGGACUAA ACAAAGGGAU GGCAACUGUG GAUGUCAAAA AGCUAAGUGA GAUGGUCUCU

     L  F  E  Q  K  Y  L  E  T  E  L  A  R  H  D  I  F  G  E  L  I  S  R  H  L  R  I  K  P  K  Q  R  N
301  CUGUUUGAGC AAAAAUACCU AGAAACAGAG UUAGCAAGGC AUGACAUUUU UGGAGAGCUG AUCUCCAGGC ACCUGAGAAU AAAGCCCAAA CAAAGAAAUG

     E  V  E  I  E  H  A  L  R  E  Y  L  D  E  L  N  K  K  S  C  I  N  K  L  S  D  D  E  F  E  R  I  N  K
401  AAGUGGAGAU AGAGCAUGCA CUAAGAGAAU AUCUGGAUGA ACUCAACAAA AAGUCCUGCA UUAACAAGCU CUCUGAUGAU GAGUUUGAGA GAAUAAAUAA

       E  Y  V  A  T  N  A  T  P  D  N  Y  V  I  Y  K  E  S  K  N  S  E  L  C  L  I  I  Y  D  W  K  I  S
501  AGAAUAUGUA GCAACUAAUG CCACCCCUGA UAACUAUGUG AUAUAUAAAG AAUCAAAAAA CAGUGAGCUU UGUUUAAUCA UUUAUGAUUG GAAAAUAUCU

     V  D  A  R  T  E  T  K  Q  W  R  N  T  Y  K  N  I  W  K  S  F  K  D  I  K  V  N  G  K  P  F  L  E
601  GUCGAUGCCA GGACUGAAAC CAAACAAUGG AGAAAUACCU ACAAGAAUAU UUGGAAAUCU UUCAAAGAUA UAAAAGUGAA UGGAAAGCCA UUCCUGGAAG

     E  H  P  V  F  V  S  I  V  I  L  K  P  I  A  G  M  P  I  T  V  T  S  S  R  V  L  E  K  F  E  D  S  P
701  AGCAUCCUGU UUUCGUUUCU AUAGUUAUAU UGAAACCUAU UGCUGGGAUG CCAAUCACUG UUACUAGUAG CAGGGUUUUG GAGAAAUUCG AAGAUUCUCC

       S  A  L  H  G  E  R  I  K  H  A  K  N  A  K  L  L  N  I  S  Y  V  G  Q  I  V  G  T  T  P  T  V  V
801  AUCAGCAUUG CACGGAGAAA GAAUAAAGCA UGCUAAAAAU GCCAAAUUGC UAAAUAUUUC UUAUGUUGGG CAAAUAGUUG GAACCACACC CACAGUGGUG

     R  N  Y  Y  A  N  T  Q  R  I  K  S  E  V  R  G  I  L  G  D  D  F  G  S  K  D  V  F  F  S  H  W  T
901  AGAAACUAUU AUGCAAACAC UCAAAGAAUC AAAUCUGAAG UCAGAGGAAU CUUAGGUGAU GAUUUUGGAU CUAAAGAUGU GUUUUUCAGU CACUGGACCA

     S  K  Y  K  E  R  N  P  T  E  I  A  Y  S  E  D  I  E  R  I  I  D  S  L  V  T  D  E  I  P  R  E  E  I
1001 GCAAAUACAA AGAAAGAAAU CCUACUGAGA UAGCCUAUUC CGAAGAUAUU GAAAGAAUAA UUGAUUCACU UGUUACAGAU GAAAUCCCUA GAGAGGAAAU

       I  H  F  L  F  G  N  F  C  F  H  I  E  T  M  N  D  Q  H  I  A  D  K  F  K  G  Y  Q  N  S  C  I  N
1101 AAUACAUUUU UUGUUUGGAA AUUUCUGUUU CCACAUUGAA ACAAUGAAUG ACCAGCAUAU AGCUGACAAA UUUAAAGGGU ACCAAAACUC UUGUAUCAAU

     L  K  I  E  P  K  A  D  L  A  D  L  K  D  H  L  I  Q  K  Q  Q  I  W  E  S  L  Y  G  K  H  L  E  K
1201 UUAAAAAUAG AGCCAAAAGC UGAUUUAGCU GAUUUGAAAG ACCACUUAAU CCAAAAGCAG CAAAUAUGGG AAUCUCUGUA UGGAAAACAC CUUGAGAAGA

     I  M  L  R  I  R  E  K  K  R  K  E  K  E  I  P  D  I  T  T  A  F  N  Q  N  A  A  E  Y  E  E  R  Y  P
1301 UCAUGCUUAG AAUUAGAGAA AAAAAGAGAA AAGAAAAAGA GAUACCUGAC AUAACCACAG CUUUUAACCA GAAUGCUGCU GAAUAUGAAG AAAGGUAUCC

     N  C  F  N  D  L  S  E  L  K  L  T  F  H  D  L  V  P  S  L  K  I  E  L  S  S  E  V  D  Y  N  N  A
1401 UAACUGUUUC AAUGAUCUCU CUGAACUAAA ACUAACUUUC CAUGACUUGG UCCCCAGUUU GAAGAUAGAA UUGAGCUCAG AGGUAGAUUA CAACAACGCA

     I  I  N  K  F  R  E  S  F  K  S  S  S  R  V  I  Y  N  S  P  Y  S  S  I  N  N  Q  T  N  K  A  R  D
1501 AUUAUUAACA AGUUUCGGGA GAGCUUCAAA AGUUCUUCAA GGGUUAUUUA UAAUAGCCCA UAUAGUAGCA UAAAUAACCA AACAAAUAAA GCAAGAGAUA

     I  T  N  L  V  R  L  C  L  A  E  L  S  C  D  T  T  K  M  E  K  Q  E  L  E  D  E  I  D  I  N  T  G  S
1601 UAACAAACUU AGUUAGACUG UGUUUAGCAG AGCUAAGUUG UGAUACAACG AAAAUGGAAA AGCAGGAACU UGAAGAUGAA AUAGAUAUAA ACACCGGGAG

       I  K  V  E  R  T  K  K  S  K  E  W  N  K  Q  G  S  C  L  T  R  N  K  N  E  F  C  M  K  D  T  G  R
1701 UAUCAAAGUU GAGAGAACAA AAAAGUCUAA AGAAUGGAAU AAGCAAGGUU CGUGUUUAAC CAGAAACAAA AAUGAAUUUU GCAUGAAAGA UACAGGCAGG

     E  N  K  T  T  Y  F  K  G  L  A  V  M  N  I  G  M  S  S  K  K  R  I  L  K  K  E  E  I  K  E  R  I
1801 GAGAACAAAA CUACCUAUUU UAAAGGCUUA GCAGUAAUGA AUAUAGGAAU GAGUUCUAAG AAAAGAAUUC UAAAAAAAGA AGAAAUAAAA GAAAGGAUCU

     S  K  G  L  E  Y  D  T  S  E  R  Q  A  D  P  N  D  D  Y  S  S  I  D  M  S  S  L  T  H  M  K  K  L  I
1901 CUAAAGGCUU GGAAUAUGAU ACCUCUGAAA GGCAGGCUGA CCCAAAUGAU GAUUACUCAA GUAUAGACAU GUCUUCUCUG ACUCAUAUGA AAAAACUGAU

       R  H  D  N  D  D  S  L  S  G  K  R  F  K  G  S  F  F  L  L  H  N  F  N  I  I  E  D  G  K  I  T  S
2001 AAGGCAUGAC AAUGAUGAUA GCUUAAGUGG UAAAAGAUUU AAGGGCUCUU UUUUUCUACU UCAUAAUUUU AAUAUAAUAG AGGAUGGUAA GAUCACAUCU

     V  F  N  N  Y  A  K  N  P  E  C  L  Y  I  Q  D  S  V  L  K  T  E  L  E  T  C  K  K  I  N  K  L  C
2101 GUUUUCAAUA AUUAUGCUAA AAAUCCUGAA UGCUUGUACA UUCAAGAUUC AGUACUGAAG ACUGAAUUAG AGACUUGCAA AAAGAUAAAC AAAUUAUGCA

     N  D  L  A  I  Y  H  Y  S  E  D  M  M  Q  F  S  K  G  L  M  V  A  D  R  Y  M  T  K  E  S  F  K  I  L
2201 AUGAUCUAGC CAUUUACCAU UACUCUGAAG ACAUGAUGCA AUUCUCCAAA GGUUUAAUGG UGGCUGACAG GUACAUGACU AAAGAAAGUU UCAAGAUAUU

       T  T  A  N  T  S  M  M  L  L  A  F  K  G  D  G  M  N  T  G  G  S  G  V  P  Y  I  A  L  H  I  V  D
2301 AACCACAGCA AAUACUAGCA UGAUGCUAUU AGCAUUCAAA GGGGAUGGAA UGAACACCGG AGGAUCGGGA GUUCCUUACA UAGCAUUGCA UAUAGUGGAU

     E  D  M  S  D  H  F  N  I  C  Y  T  K  E  I  Y  S  Y  F  R  S  G  S  N  Y  I  Y  I  M  R  P  Q  R
2401 GAAGACAUGU CAGAUCAUUU UAACAUAUGU UAUACUAAAG AAAUUUAUAG CUAUUUCCGA AGUGGUAGUA AUUACAUUUA UAUAAUGAGG CCGCAGAGAC

     L  N  Q  V  R  L  L  R  L  F  K  T  P  S  K  V  P  V  C  F  P  Q  F  S  K  K  A  N  E  I  G  K  S  L
2501 UAAACCAGGU GAGGCUGCUG AGGCUUUUCA AAACGCCUAG UAAAGUUCCU GUAUGUUUUC CACAAUUUUC AAAGAAAGCU AAUGAAAUCG GAAAAUCGCU

     K  N  K  D  I  E  K  V  N  L  F  S  M  T  M  T  V  K  Q  I  L  I  N  I  V  F  S  S  V  M  I  G  T
2601 GAAAAAUAAA GAUAUAGAAA AAGUAAAUCU CUUUUCUAUG ACAAUGACUG UAAAACAGAU AUUAAUAAAU AUUGUGUUUU CAUCUGUCAU GAUAGGAACU

     V  T  K  L  S  R  M  G  I  F  D  F  M  R  Y  A  G  F  L  R  L  S  D  Y  S  N  I  K  E  Y  I  R  D
2701 GUGACAAAGC UCAGUAGAAU GGGAAUUUUU GAUUUCAUGC GGUAUGCAGG UUUUUUGCGA CUAUCCGAUU AUUCUAACAU AAAAGAAUAC AUUAGAGACA

     K  S  D  P  D  I  T  N  C  G  R  Y  L  F  R  N  G  I  K  K  L  L  F  R  M  E  D  L  N  L  S  T  N  A
2801 AAUCUGAUCC UGAUAUAACU AACUGUGGCA GAUAUCUAUU UCGUAAUGGA AUCAAAAAAC UAUUGUUCAG AAUGGAAGAU CUCAAUUUAA GCACAAAUGC

     K  P  V  V  V  D  H  E  N  D  I  I  G  G  I  T  N  L  N  I  K  C  P  I  T  G  S  T  L  L  T  L  E
2901 CAAGCCUGUU GUUGUGGACC ACGAAAAUGA UAUUAUAGGA GGGAUAACAA ACUUGAAUAU AAAAUGUCCU AUAACAGGAU CAACUCUACU GACACUUGAG

     D  L  Y  N  N  V  Y  L  A  I  Y  M  M  P  K  S  L  H  N  H  V  H  N  L  T  S  L  L  N  L  P  A  E
3001 GACCUGUACA AUAAUGUUUA UUUGGCUAUU UAUAUGAUGC CUAAAUCACU GCACAAUCAU GUUCACAAUC UAACAAGCUU AUUAAAAUCUC CCUGCUGAGU
```

```
     W  E  L  K     F  R  K     E  L  G     F  N  I  F     E  D  I     Y  P  K     K  A  M  F     D  D  K     D  L  F     S  I  N  G
3101 GGGAGCUAAA GUUCAGAAAA GAAUUAGGUU UCAACAUAUU UGAAGACAUA UACCCUAAGA AAGCAAUGUU UGAUGACAAA GACCUAUUCU CCAUAAAUGG

     A  L  N     V  K  A     L  S  D  Y     Y  L  G     N  I  E     N  V  G  L     M  R  S     E  I  E     N  K  E  D     F  L  S
3201 AGCUUUGAAC GUGAAAGCAU UAUCUGAUUA CUAUCUAGGA AAUAUAGAAA AUGUUGGUUU AAUGAGAUCA GAAAUAGAAA AUAAAGAAGA UUUCCUAAGC

     P  C  Y     K  I  S  T     L  K  S     S  K  K     C  S  Q  S     N  I  I     S  T  D     E  I  I  E     C  L  Q     D  A  K
3301 CCUUGUUAUA AAAUAUCUAC UUUAAAAUCU UCAAAAAAAU GCUCACAAUC AAACAUUAUA AGUACUGAUG AGAUAAUAGA GUGUCUUCAG GAUGCAAAGA

     I  Q  D  I     E  N  W     K  G  N     N  L  A  I     I  K  G     L  I  R     T  Y  N  E     E  K  N     R  L  V     E  F  F  E
3401 UUCAAGAUAU AGAAAAUUGG AAAGGAAAUA ACCUAGCUAU UAUAAAAGGG CUUAUAAGAA CCUACAAUGA GGAGAAAAAU AGAUUGGUGG AAUUUUUUGA

     D  N  C     V  N  S     L  Y  L  I     E  K  L     K  E  I     I  S  S  G     S  I  T     V  G  K     S  V  T  S     K  F  I
3501 AGAUAAUUGU GUCAAUUCAU UAUAUCUUAU AGAAAAGCUU AAAGAGAUAA UUAGUAGUGG AUCAAUAACU GUAGGGAAAU CUGUAACAUC UAAAUUCAUA

     R  N  N     H  P  L  T     V  E  T     Y  L  K     T  K  L  Y     Y  R  N     N  V  T     V  L  K  S     K  K  V     S  E  E
3601 AGAAACAAUC AUCCUUUAAC AGUAGAAACA UAUCUCAAAA CAAAACUAUA UUAUAGAAAU AAUGUAACAG UUUUAAAGUC UAAAAAAGUG UCAGAGGAGC

     L  Y  D  L     V  K  Q     F  H  D     M  M  E  I     D  L  D     S  V  M     N  L  G  K     G  T  E     G  K  K     L  T  F  L
3701 UCUAUGACCU UGUAAAACAG UUCCAUGACA UGAUGGAAAU AGACCUAGAU UCUGUUAUGA ACCUCGGGAA AGGUACAGAA GGAAAAAAAC UCACAUUCUU

     Q  M  L     E  F  V     M  S  K  A     K  N  V     T  G  S     V  D  F  L     V  S  V     F  E  K     M  Q  R  T     K  T  D
3801 GCAGAUGCUU GAAUUUGUCA UGUCCAAGGC UAAAAAUGUC ACCGGGUCUG UAGAUUUUCU AGUUUCUGUU UUUGAAAAAA UGCAGAGAAC CAAAACAGAC

     R  E  I     Y  L  M  S     M  K  V     K  M  M     L  Y  F  I     E  H  T     F  K  H     V  A  Q  S     D  P  S     E  A  I
3901 AGAGAAAUAU ACUUGAUGAG CAUGAAAGUG AAAAUGAUGC UUUAUUUUAU AGAGCAUACA UUCAAACAUG UAGCGCAGAG UGAUCCAUCG GAAGCCAUAU

     S  I  S  G     D  N  K     I  R  A     L  S  T  L     S  L  D     T  I  T     S  Y  N  D     I  L  N     K  N  S     K  K  S  R
4001 CUAUAAGUGG AGACAAUAAA AUAAGAGCAC UUUCUACGUU AUCUUUGGAC ACAAUCACGU CUUACAAUGA UAUUUUAAAC AAAAAUUCAA AGAAGUCAAG

     L  A  F     L  S  A     D  Q  S  K     W  S  A     S  G  L     T  T  Y  K     Y  V  L     A  I  I     L  N  P  I     L  T  T
4101 AUUGGCUUUC CUAUCUGCAG AUCAGUCGAA AUGGUCGGCA UCAGGCCUUA CCACCUAUAA AUAUGUUUUA GCUAUCAUAU UAAAUCCAAU UUUAACUACU

     G  E  A     S  L  M  I     E  C  I     L  M  Y     V  K  L  K     K  V  C     I  P  T     D  I  F  L     N  L  R     K  A  Q
4201 GGUGAAGCUA GCUUAAUGAU AGAAUGCAUC UUAAUGUAUG UUAAAUUGAA GAAGGUUUGU AUACCAACAG AUAUUUUUUU GAAUCUAAGA AAAGCUCAAC

     Q  T  F  G     E  N  E     T  A  I     G  L  L  T     K  G  L     T  T  N     T  Y  P  V     S  M  N     W  L  Q     G  N  L  N
4301 AAACUUUUGG GGAAAAUGAA ACUGCCAUAG GACUUUUGAC CAAAGGCUUG ACGACAAACA CAUACCCUGU UAGCAUGAAU UGGUUGCAAG GCAAUUUAAA

     Y  L  S     S  V  Y     H  S  C  A     M  K  A     Y  H  N     T  L  E  C     Y  K  N     C  D  F     Q  T  R  W     I  V  H
4401 UUAUCUGUCU UCUGUUUAUC ACUCUUGUGC AAUGAAAGCU UAUCACAACA CUUUGGAAUG UUACAAAAAC UGUGAUUUCC AAACUAGAUG GAUUGUGCAC

     S  D  D     N  A  T  S     L  I  A     S  G  E     V  D  K  M     L  T  D     F  S  S     S  S  L  P     E  M  L     F  R  S
4501 UCUGAUGAUA AUGCAACAUC AUUAAUAGCC AGUGGAGAGG UUGAUAAAAU GCUGACAGAC UUUUCAAGCU CAUCUCUGCC AGAAAUGUUG UUUAGAAGCA

     I  E  A  H     F  K  S     F  C  I     T  L  N  P     K  K  S     Y  A  S     S  S  E  V     E  F  I     S  E  R     I  S  K  W
4601 UUGAAGCUCA UUUCAAAAGU UUUUGCAUAA CUUUGAACCC AAAAAAGAGU UAUGCUUCUU CAUCAGAAGU AGAGUUUAUA UCUGAAAGAA UUAGUAAAUG

     S  D  Y     S  S  L     L  Q  A  F     S  K  L     L  H  R     I  F  A  Y     K  L  F     D  D  L     M  S  L  S     I  H  V
4701 GAGCGAUUAU UCCUCUCUAU UGCAGGCAUU UAGCAAACUG UUGCACAGAA UCUUCGCAUA UAAGUUAUUU GAUGAUCUAA UGUCACUCAG UAUACAUGUU

     T  M  L     L  R  K  G     C  P  N     E  V  I     P  F  A  Y     G  A  V     Q  V  Q     A  L  S  I     Y  S  M     L  P  G
4801 ACAAUGCUUC UGAGAAAAGG CUGUCCUAAU GAAGUUAUAC CUUUUGCUUA UGGGGCUGUG CAGGUACAAG CGUUAAGCAU CUAUUCAAUG CUUCCUGGUG

     E  V  N  D     S  I  R     I  F  N     K  L  G  V     S  L  K     S  N  E     I  P  T  N     M  G  G     W  L  T     S  P  I  E
4901 AAGUGAAUGA UAGUAUUAGA AUUUUUAACA AGCUUGGAGU AAGUUUAAAG UCAAACGAGA UUCCCACAAA CAUGGGGGGC UGGUUGACCU CUCCUAUAGA

     P  L  S     I  L  G     P  S  S  N     D  Q  I     I  Y  Y     N  V  I  R     D  F  L     N  K  K     S  L  E  E     V  K  D
4501 GCCGUUGUCU AUAUUAGGUC CAUCAUCAAA UGAUCAAAUC AUCUAUUACA AUGUGAUAAG AGAUUUUUUG AACAAAAAAA GUUUAGAAGA AGUAAAAGAU

     S  V  S     S  S  S  Y     L  Q  M     R  F  R     E  L  K  E     K  Y  E     R  G  T     L  E  E  K     D  K  K     M  I  F
5101 AGUGUCUCUU CUUCCAGUUA UCUACAGAUG AGAUUCAGAG AGCUAAAAGA AAAGUAUGAA AGAGGAACUC UGGAAGAAAA AGAUAAAAAG AUGAUAUUUC

     L  I  N  L     F  E  K     A  S  V     S  E  D  S     D  V  L     T  I  G     M  K  F  Q     T  M  L     T  Q  I     I  K  L  P
5201 UUAUCAAUCU GUUUGAGAAA GCAUCAGUGU CUGAAGAUUC AGACGUUCUA ACAAUUGGGA UGAAAUUUCA AACUAUGUUA ACUCAGAUUA UAAAAUUACC

     N  F  I     N  E  N     A  L  N  K     M  S  S     Y  K  D     F  S  K  L     Y  P  N     L  K  K     N  E  D  L     Y  K  S
5301 UAAUUUUAUA AAUGAGAAUG CUUUAAACAA GAUGUCAAGU UAUAAAGAUU UUUCAAAACU UUAUCCUAAU UUAAAAAAGA AUGAAGAUUU AUAUAAAAGC

     T  K  N     L  K  I  D     E  D  A     V  L  E     E  D  E  L     Y  K  K     I  A  S     S  L  E  M     E  S  V     H  D  I
5401 ACUAAGAACU UAAAGAUAGA CGAGGAUGCU GUUUUAGAGG AAGAUGAGUU AUAUAAGAAG AUUGCAUCUA GCUUAGAAAU GGAAUCUGUC CAUGACAUAA

     M  I  K  N     P  E  T     I  L  I     A  P  L  N     D  R  D     F  L  L     S  Q  L  F     M  Y  T     S  P  S     K  R  N  Q
5501 UGAUAAAAAA UCCUGAAACA AUUCUGAUAG CACCAUUGAA UGAUAGAGAU UUUUUACUUA GUCAGCUGUU CAUGUACACA AGCCCUUCUA AAAGAAACCA

     L  S  N     Q  S  T     E  K  L  A     L  D  R     V  L  R     S  K  A  R     T  F  V     N  I  S     S  T  V  K     M  T  Y
5601 GUUAUCGAAC CAAUCUACAG AGAAACUUGC UUUAGAUAGA GUGUUAAGGU CAAAAGCUAG AACAUUUGUA AACAUUUCUU CCACUGUGAA GAUGACUUAU

     E  E  N     M  E  K  K     I  L  E     M  L  K     F  D  L  D     S  Y  C     S  F  K     T  C  V  N     L  V  I     K  D  V
5701 GAAGAAAACA UGGAAAAGAA AAUCUUAGAA AUGCUAAAAU UUGAUUUAGA UUCAUAUUGU UCAUUUAAAA CAUGUGUAAA UCUAGUUAUC AAGGAUGUUA

     N  F  S  M     L  I  P     I  L  D     S  A  Y  P     C  E  S     R  K  R     D  N  Y  N     F  R  W     F  Q  T     E  R  W  I
5801 AUUUCAGCAU GCUGAUUCCA AUAUUAGAUU CUGCAUACCC UUGUGAAUCU AGGAAAAGAG AUAACUACAA UUUCAGGUGG UUUCAGACUG AGAGAUGGAU

     P  V  V     E  G  S     P  G  L  V     V  M  H     A  V  Y     G  S  N  Y     I  E  N     L  G  L     K  N  I  P     L  T  D
5901 ACCUGUUGUU GAAGGCUCUC CGGGACUAGU AGUAAUGCAU GCUGUCUAUG GAUCAAAUUA UAUAGAGAAC UUAGGUUUAA AAAACAUCCC UCUAACAGAC

     D  S  I     N  V  L  T     S  T  F     G  T  G     L  I  M  E     D  V  K     S  L  V     N  G  K  D     S  F  E     T  E  A
6001 GAUAGUAUUA AUGUUUUAAC AAGCACGUUU GGAACAGGUU UAAUCAUGGA AGAUGUAAAA UCCCUAGUUA AUGGCAAAGA CAGCUUCGAA ACAGAGGCUU

     F  S  N  S     N  E  C     Q  R  L     V  K  A  C     N  Y  M     I  A  A     Q  N  R  L     L  A  I     N  T  C     F  T  R  K
6101 UUAGCAAUUC UAAUGAAUGU CAAAGAUUGG UGAAAGCAUG CAAUUAUAUG AUAGCAGCAC AAAACAGGCU UUUAGCAAUU AACACAUGCU UUACUAGGAA

     S  F  P     F  Y  S     K  F  N  L     G  R  G     F  I  S     N  T  L  A     L  L  S     T  I  Y     S  K  E  E     S  Y  H
6201 AAGCUUUCCC UUCUAUUCUA AGUUCAAUCU AGGGAGAGGG UUUAUCUCAA ACACAUUAGC UCUCCUAUCC ACCAUCUACA GUAAAGAAGA AUCCUAUCAU
```

```
     F  V  S  T  A  S  Y  K  L  D  K  T  I  R  T  V  V  S  A  Q  Q  D  M  N  L  E  K  I  L  D  T  A  V
6301 UUUGUUUCUA CAGCUAGUUA UAAAUUAGAC AAAACUAUCA GAACUGUGGU AAGUGCUCAG CAAGAUAUGA ACUUAGAGAA AAUACUGGAC ACUGCUGUAU

     Y  I  S  D  K  L  Q  S  L  F  P  T  I  T  R  E  D  I  V  L  I  L  Q  N  V  C  L  D  S  K  P  I  W  Q
6401 ACAUAUCAGA UAAAUUGCAG UCACUUUUCC CAACAAUUAC AAGAGAGGAU AUAGUUUUAA UAUUGCAAAA UGUUUGCCUU GACAGUAAAC CUAUAUGGCA

     S  L  E  D  K  M  K  K  I  N  N  S  T  A  S  G  F  T  V  S  N  V  I  L  S  H  N  S  E  L  N  T  I
6501 GAGUCUAGAA GACAAAAUGA AAAAGAUUAA CAAUUCAACA GCAAGUGGCU UCACAGUGUC AAAUGUGAUU CUAUCACAUA ACAGUGAAUU GAACACAAUC

     Q  K  Q  I  V  W  M  W  N  M  G  L  C  S  H  R  T  L  D  F  V  I  R  Y  I  R  R  R  D  V  R  Y  V
6601 CAGAAACAAA UUGUCUGGAU GUGGAACAUG GGUUUGUGUU CUCACAGAAC AUUAGAUUUU GUUAUCAGGU AUAUUAGAAG AAGGGAUGUA AGAUAUGUAA

     K  T  E  E  Q  D  E  S  G  N  Y  V  S  G  T  M  Y  K  I  G  I  M  T  R  S  C  Y  V  E  L  I  A  S  D
6701 AAACUGAAGA ACAAGAUGAA UCAGGAAAUU AUGUCUCUGG AACUAUGUAC AAAAUAGGGA UCAUGACAAG AAGCUGCUAU GUGGAAUUGA UAGCAUCUGA

     Q  D  V  A  V  S  L  R  T  P  F  E  I  L  N  E  R  E  Y  L  F  D  T  Y  R  E  S  I  E  K  L  L  A
6801 UCAAGAUGUA GCAGUUUCUU UGAGAACACC AUUUGAGAUA UUGAAUGAAA GAGAGUAUCU UUUUGACACA UACAGAGAAA GUAUAGAGAA AUUAUUGGCA

     E  I  M  F  D  K  V  N  I  I  N  Q  T  T  T  D  C  F  L  R  T  R  R  S  C  I  R  M  T  T  D  N  K
6901 GAAAUUAUGU UUGAUAAAGU GAACAUAAUA AAUCAAACAA CCACAGAUUG UUUUCUUAGA ACCAGGAGAU CUUGCAUCAG AAUGACCACA GACAACAAAA

     M  I  V  K  V  N  A  T  S  R  Q  I  R  L  E  N  V  K  L  V  V  K  I  K  Y  E  N  V  N  S  D  V  W  D
7001 UGAUUGUAAA GGUUAAUGCU ACAUCAAGAC AAAUAAGACU AGAGAAUGUA AAAUUAGUUG UAAAGAUAAA AUAUGAAAAU GUGAAUUCCG AUGUAUGGGA

     I  I  E  S  Q  K  S  L  V  L  R  L  P  E  V  G  E  F  F  S  D  M  Y  K  T  A  D  S  E  T  E  T  I
7101 UAUUAUAGAA AGCCAAAAAU CUCUAGUCUU AAGGCUCCCU GAAGUAGGGG AAUUUUUCUC UGAUAUGUAU AAAACUGCAG ACUCUGAAAC UGAAACAAUC

     K  T  I  K  N  R  L  M  T  S  L  T  F  I  E  A  F  G  N  L  S  Q  Q  I  K  E  I  V  D  D  D  I  R
7201 AAAACCAUAA AAAACAGGCU UAUGACUUCU UUAACUUUCA UAGAAGCCUU UGGAAACUUA UCACAGCAGA UCAAAGAGAU UGUAGAUGAU GAUAUCAGAG

     E  T  M  D  E  F  L  M  N  I  R  D  T  C  L  E  G  L  E  N  C  K  S  V  E  E  Y  D  S  Y  L  D  E  N
7301 AAACGAUGGA UGAAUUCUUA AUGAACAUCC GGGAUACCUG CUUAGAAGGU UUGGAAAACU GCAAAAGUGU GGAAGAAUAU GAUAGCUAUC UUGAUGAAAA

     G  F  N  D  T  V  E  L  F  E  N  L  L  R  T  H  D  N  F  E  N  E  Y  S  P  L  F  S  E  I  V  D  K
7401 UGGAUUUAAU GACACAGUAG AACUAUUCGA AAACUUGCUA AGAACACAUG ACAACUUUGA AAAUGAGUAU AGUCCUCUUU UUUCAGAGAU UGUUGACAAA

     A  K  Q  Y  T  R  D  L  E  G  F  K  E  I  L  L  M  L  K  Y  S  L  I  N  D  A  S  G  F  K  S  Y  R
7501 GCAAAACAGU AUACUAGAGA UUUAGAAGGU UUCAAAGAAA UACUGCUCAU GCUUAAAUAU UCUCUAAUAA AUGAUGCAUC AGGAUUUAAA AGCUAUAGAG

     A  T  G  M  H  A  V  E  L  M  A  K  K  H  I  E  I  G  E  F  N  L  L  G  M  I  Q  L  I  K  A  C  E  T
7601 CCACUGGAAU GCAUGCUGUU GAGCUAAUGG CAAAAAAGCA CAUAGAGAUA GGGGAAUUCA ACUUGUUAGG AAUGAUCCAA UUGAUUAAAG CUUGUGAAAC

     C  H  N  N  D  S  I  L  N  L  A  S  L  R  N  V  L  S  R  T  Y  A  T  F  G  R  R  I  R  L  D  H  D
7701 AUGCCACAAC AAUGACUCUA UAUUAAACUU AGCAAGUUUA AGGAAUGUUC UUAGCAGGAC AUAUGCCACA UUUGGGAGGA GAAUAAGAUU GGAUCAUGAU

     L  D  L  Q  N  N  L  M  E  K  S  Y  D  F  K  T  L  V  L  P  E  I  K  L  S  E  L  S  R  E  I  L  K
7801 CUGGACUUGC AAAACAACUU AAUGGAAAAA AGUUAUGAUU UCAAGACGCU GGUUUUACCA GAAAUAAAAU UAUCAGAACU AUCUAGGGAA AUACUGAAAG

     E  N  G  F  V  I  S  G  E  N  L  K  M  D  R  S  D  E  E  F  V  G  L  A  S  F  N  V  L  R  L  D  E  E
7901 AAAAUGGGUU UGUUAUAUCU GGAGAGAAUC UAAAAAUGGA UAGGUCUGAU GAAGAAUUUG UGGGUCUUGC CAGUUUUAAU GUGUUGAGGC UAGAUGAGGA

     E  M  Y  E  G  L  I  K  E  M  K  I  K  R  K  K  K  G  F  L  F  P  A  N  T  L  L  L  S  E  L  I  K
8001 AGAAAUGUAU GAAGGUUUGA UCAAAGAAAU GAAAAUUAAA AGGAAAAAGA AAGGGUUUUU AUUUCCAGCA AACACACUUC UACUAAGUGA GUUGAUAAAG

     F  L  I  G  G  I  K  G  T  S  F  D  I  E  T  L  L  R  N  S  F  R  P  D  I  F  S  T  D  R  L  G  R
8101 UUCUUGAUUG GAGGAAUAAA GGGAACCAGC UUUGAUAUAG AGACAUUGUU ACGGAACAGU UUUAGACCAG ACAUAUUUUC AACUGACAGA UUGGGAAGAU

     L  S  S  S  V  P  A  L  K  V  Y  A  T  V  Y  M  E  Y  K  N  V  N  C  P  L  N  E  I  A  D  S  L  E  G
8201 UAAGUUCCAG UGUACCUGCA CUCAAAGUUU AUGCAACUGU UUAUAUGGAA UAUAAGAAUG UCAAUUGUCC UUUAAAUGAG AUAGCUGACA GCUUAGAAGG

     Y  L  K  L  T  K  S  R  S  K  E  H  F  L  S  G  R  V  K  K  A  L  I  Q  L  R  D  E  Q  S  R  T  K
8301 UUAUCUAAAA CUGACAAAAA GCAGGUCCAA GGAACAUUUC UUGUCUGGAA GAGUUAAAAA AGCUUUGAUA CAAUUAAGAG AUGAACAAUC GCGAACUAAA

     K  L  E  V  Y  K  D  I  A  N  F  L  A  R  H  P  L  C  L  S  E  K  T  L  Y  G  R  Y  T  Y  S  D  I
8401 AAACUAGAGG UCUAUAAGGA UAUCGCAAAU UUCCUUGCUA GGCACCCACU AUGUUUAUCA GAAAAAACAU UGUAUGGAAG AUAUACCUAC UCUGAUAUCA

     N  D  Y  I  M  Q  T  R  E  I  I  L  S  K  I  S  E  L  D  E  V  V  E  T  D  E  D  N  F  L  L  S  Y  L
8501 AUGAUUAUAU CAUGCAAACA AGAGAGAUUA UUUUGAGUAA AAUAAGUGAG UUGCACGAGG UUGUUGAAAC AGAUGAAGAC AAUUUCUUGC UUAGUUAUCU

     R  G  E  E  D  A  F  D  E  D  E  L  D  E  E  E  D  T  D  *
8601 AAGAGGGGAA GAAGAUGCCU UUGAUGAAGA UGAGCUUGAU GAAGAAGAAG ACACAGAUUA AAUUGAAAGU AAUGACUAAC AAUCCAUGAA UAACAGAUUA

8701 GAUAUAACUU AGAAUAUAAA UUUAUUGCUA UUUUAGAAUU AGAUUAGAUC UACUUAGCCU AAAACAAUUU GGUGAACCAA AUCUAUAGUG UAUAUAAAUG

8801 UAGAGUCCCG GUAUAGUUUC ACUGGAGGGA AUUCUUAUGU AAUUUGUAAA GUCUGGCUGU GGAGAGGUUA UAUGUUUUAG UUGUACCUGA UUGCUCU
```

Fig. 2. The complete nucleotide sequence of TSWV L RNA (numbered from the 5′ end of the viral complementary RNA strand) and its predicted gene product. The deduced amino acid sequence of the protein encoded by the viral complementary RNA is written above the RNA sequence. The asterisk (*) indicates the UAA termination codon.

with a base composition of 28·7% A, 37·8% U, 19·0% C and 14·5% G. The length is in rather good agreement with the previously estimated size, deduced from electrophoretic mobility (Van den Hurk *et al.*, 1977; de Haan *et al.*, 1989*a*). The L RNA exhibits complementarity between its 3′ and 5′ ends for 62 nucleotides at the 5′ end to 66 nucleotides at the 3′ end, similar in range to the complementary termini of the S RNA (de Haan *et al.*, 1990). The resulting 'panhandle' structure (Fig. 3) has a free energy of $\Delta G = -217{\cdot}1$ kJ/mol. Moreover, the 10 3′- and 5′-terminal nucleotides show a remarkable homology to that of RNA segment 3 of Thogoto virus, a tick-borne member of the Orthomyxoviridae (Clerx *et al.*, 1983; Staunton *et al.*, 1989) (Fig. 4).

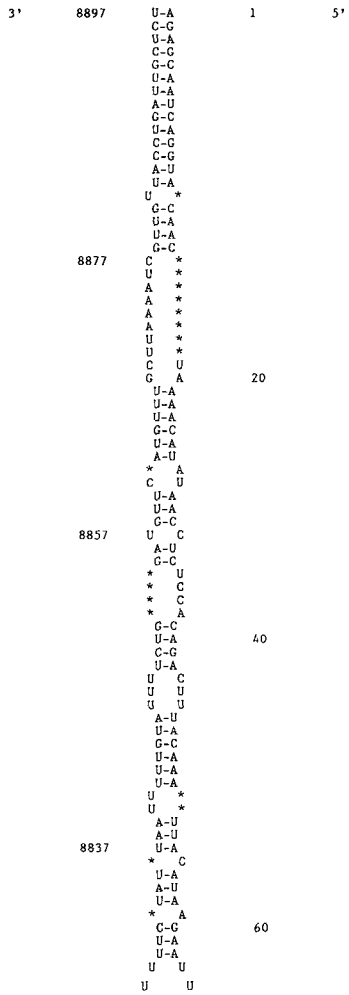

Fig. 3. The complementary sequence at the 5′ and 3′ end of the TSWV L RNA. The nucleotide positions are numbered from the 5′ end. Asterisks (*) represent gaps corresponding to unpaired nucleotides in the sequence.

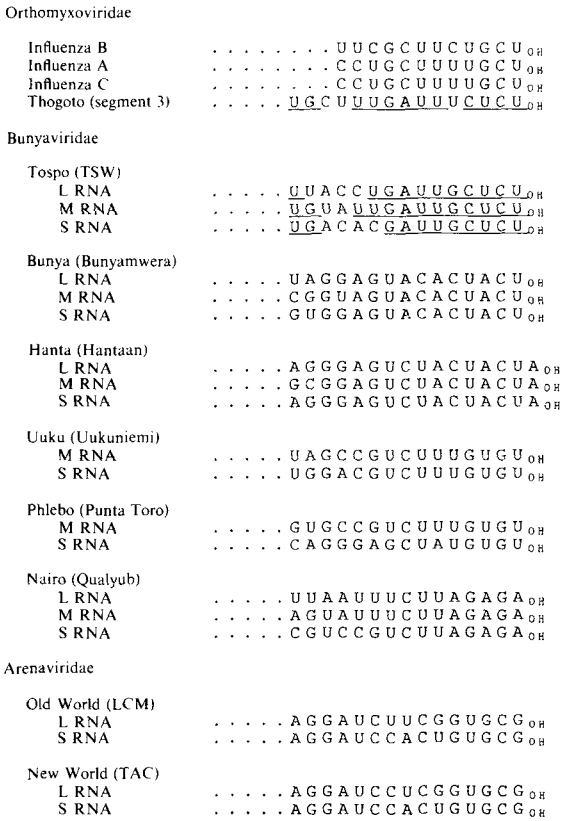

Fig. 4. Comparison of the 3′-terminal sequences of the genomic RNA molecules of TSWV to those of members of the Arenaviridae, Bunyaviridae and Orthomyxoviridae. Nucleotides conserved between TSWV and Thogoto virus are underlined.

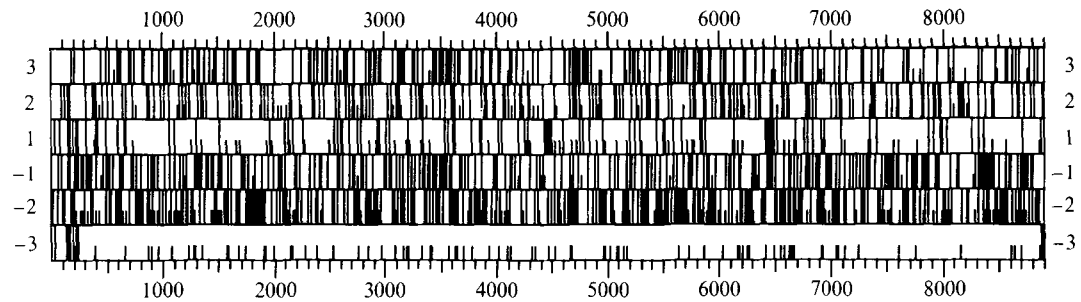

Fig. 5. Distribution of translation initiation (short vertical bars) and termination (long vertical bars) codons in the three possible reading frames of the viral (1, 2 and 3) and viral complementary (−1, −2 and −3) L RNA strands.

Predicted gene product encoded by TSWV L RNA

Analysis of the six reading frames of the viral and viral complementary RNA strand revealed only one large ORF, located on the viral complementary RNA strand (Fig. 5). This ORF starts with an AUG codon at position 34 and extends to a UAA stop codon at position 8659, hence the non-coding regions of the plus-sense RNA are 33 bases long at the 5′ end and 235 bases at the 3′ end.

The amino acid sequence derived from this ORF is shown in Fig. 2. The sequence of the predicted gene product is 2875 amino acids long and has an estimated

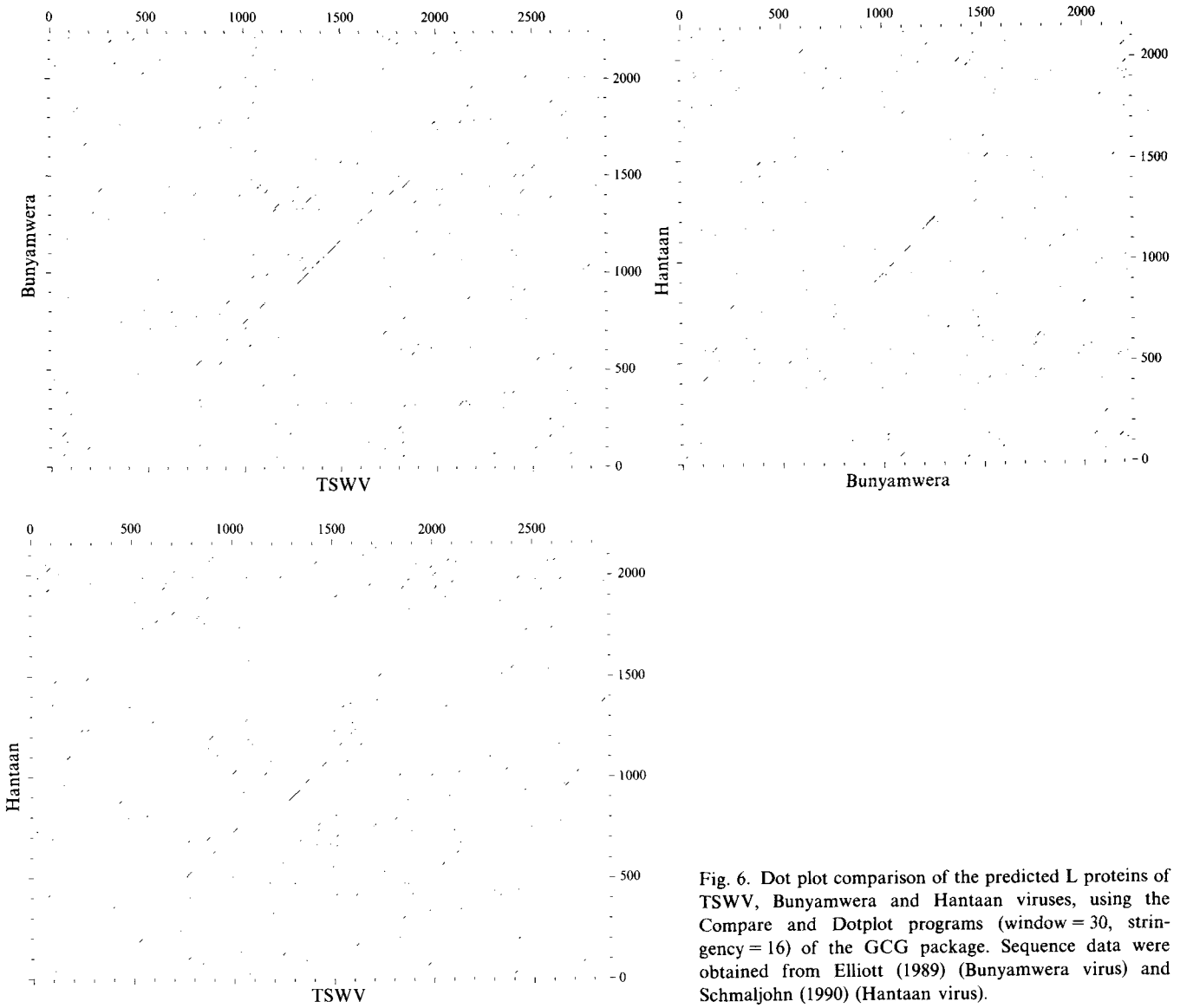

Fig. 6. Dot plot comparison of the predicted L proteins of TSWV, Bunyamwera and Hantaan viruses, using the Compare and Dotplot programs (window = 30, stringency = 16) of the GCG package. Sequence data were obtained from Elliott (1989) (Bunyamwera virus) and Schmaljohn (1990) (Hantaan virus).

M_r of 331·5K. Analysis of the amino acid sequence of the predicted protein reveals several short hydrophobic regions (Kyte & Doolittle, 1982) and a very acidic carboxy terminus, as can be seen by the large number of aspartic acid (D) and glutamic acid (E) residues (Fig. 2).

A search in the EMBL protein and nucleotide sequence database revealed that the predicted protein encoded by TSWV L RNA is homologous to the L proteins of the animal-infecting Bunyaviridae. Hence, it can be deduced that the L RNA segment of TSWV encodes the L protein. The discrepancy between the size reported here (331·5K) and the previously estimated size (200K) may be due to the gel systems used in those experiments (Mohamed *et al.*, 1973; Tas *et al.*, 1977). Computer-assisted alignment of the predicted L protein of TSWV with that of Bunyamwera virus (Elliott, 1989) reveals one internal region (approximately 1000 amino acids long) with significant (27% identity) amino acid sequence homology (Fig. 6). Homology between TSWV and Hantaan virus L proteins, and between those of Bunyamwera and Hantaan virus, however, is lower and restricted to a shorter internal stretch of approximately 200 to 250 residues (Fig. 6).

For the animal-infecting Bunyaviridae it has been proposed that the L proteins represent the viral RNA polymerases. Proteins involved in transcription and replication of RNA viruses contain conserved signature sequences, such as putative polymerase, helicase or methyltransferase motifs (Kamer & Argos, 1984; Goldbach, 1987; Hodgman, 1988; Gorbalenya *et al.*, 1989). The presence or absence of these motifs, together with other molecular characteristics such as genome structure

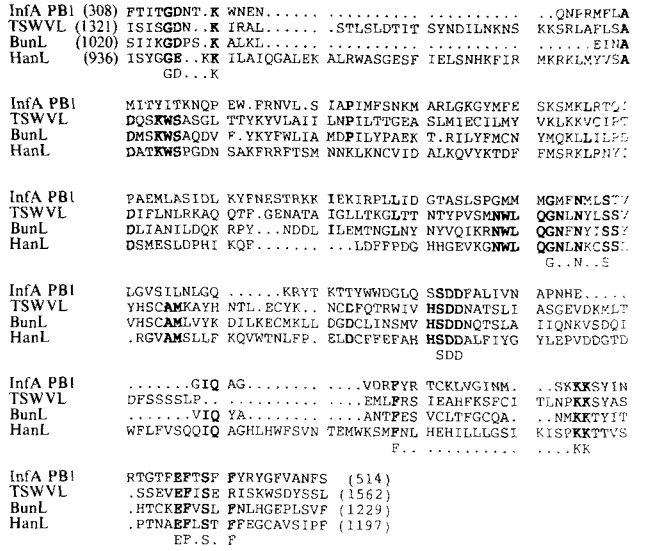

Fig. 7. Amino acid sequence homology between the RNA polymerases of members of the Bunyaviridae and protein PB1, the core polymerase of the influenza viruses. Residues conserved in at least three sequences are indicated in bold. Sequence data were obtained from Yamashita *et al.* (1989) (influenza A virus), Elliott (1989) (Bunyamwera virus) and Schmaljohn (1990) (Hantaan virus).

and expression, are important determinants for evolutionary relationships between viruses and virus families (Goldbach, 1986; Strauss & Strauss, 1988; Poch *et al.*, 1989; Candresse *et al.*, 1990). A search for such conserved sequences in the (putative) polymerase proteins of members of the Bunyaviridae and Orthomyxoviridae reveals five types of short consensus sequences: $GDX_{1-3}K$, GXXNXXS, SDD, $FX_{10-17}KK$ and EFXSXF (Fig. 7). These amino acid motifs are present in the region where the predicted L protein of TSWV shows sequence homology to Bunyamwera and Hantaan L proteins and to influenza A virus protein PB1, the core polymerase of this virus (Braam *et al.*, 1983; Krug *et al.*, 1989). Hence, it is anticipated that the major ORF in TSWV L RNA represents the polymerase gene.

Discussion

Determination of the complete nucleotide sequence of the TSWV L RNA demonstrates that TSWV is a negative-strand RNA virus. The presented nucleotide sequence data confirm the previous conclusion, derived from the S RNA sequence, that TSWV should be considered as a member of the Bunyaviridae. Indeed, at the ICTV meeting during the Eighth International Congress of Virology in Berlin (1990), TSWV was accepted as the first member of a newly created genus, tospovirus, within the Bunyaviridae.

The TSWV L RNA segment is 8897 nucleotides long, which is significantly longer than the L RNAs of Bunyamwera (6875 nucleotides) and Hantaan viruses (6530 nucleotides) (Elliott, 1989; Schmaljohn, 1990). Additional domains may be present in the gene product of TSWV L RNA, which may reflect adaptation of this bunyavirus to plants.

TSWV L RNA contains complementary ends of 62 to 66 nucleotides in length. Hence, the RNA can be folded into a stable panhandle structure (Fig. 3), which may be involved in the appearance of circular nucleocapsids in virus particles (Peters *et al.*, 1991), as also found for the Bunyaviridae (Raju & Kolakofsky, 1989). Moreover, these terminal sequences will play an important role in genome transcription and replication, since they contain the initiation signals for encapsidation and RNA synthesis (Krug *et al.*, 1989; Parvin *et al.*, 1989). An alignment of the 3′-terminal sequences of the RNAs from segmented negative-strand viruses is shown in Fig. 4. On the basis of terminal nucleotide sequence homology, the animal bunyaviruses can be clustered into three groups, the nairoviruses (Clerx-van Haaster *et al.*, 1982), the uuku-/phleboviruses (Ihara *et al.*, 1984, 1985; Ronnholm & Pettersson, 1987; Simons *et al.*, 1990) and the hanta-/bunyaviruses (Schmaljohn *et al.*, 1986, 1987; Clerx-van Haaster *et al.*, 1982). Members of the Arenaviridae and Orthomyxoviridae have their own distinct terminal sequences (Fig. 4; Desselberger *et al.*, 1980; Auperin *et al.*, 1982). Strikingly, the termini of the TSWV RNAs show considerable sequence homology to that of RNA segment 3 of Thogoto virus, a member of the Orthomyxoviridae, which might reflect ancestral relationships between both virus families.

TSWV L RNA contains a single ORF in the viral complementary sense, corresponding to a protein with a predicted M_r of 331·5K. Analysis of viral RNA species in infected plant cells indicates that this ORF is expressed by the formation of an mRNA of approximately genome length. No subgenomic RNA species derived from the L RNA could be detected (unpublished results). Remarkably, in several TSWV isolates, defective L RNA species appear when maintained under laboratory conditions. In line BR-01, which has been used for sequence determination of the L RNA, a deleted form of this RNA segment accumulates, which is approximately the size of the M RNA. The genesis and implications of these defective RNA molecules for virus multiplication are currently under investigation.

The predicted 331·5K protein encoded by TSWV L RNA most probably corresponds to the viral polymerase. Comparisons of (putative) RNA polymerases from TSWV, Bunyamwera, Hantaan and influenza A viruses reveal the presence of amino acid sequence motifs that are present in all polymerases showing RNA template

specificity and most likely form the active sites for RNA synthesis (Poch *et al.*, 1989). The region in the predicted TSWV L protein, surrounding these 'polymerase' motifs, shows considerable sequence homology (approximately 27% identity) to the putative polymerase of Bunyamwera virus, but to a much lesser extent to that of Hantaan virus. All three L proteins in their turn share conserved amino acid motifs, in a stretch of 200 to 250 residues, with the PB1 polymerase subunit of influenza viruses (Fig. 7). These findings further underline the importance of these common signature sequences and, moreover, justify the assumption that TSWV L RNA indeed encodes the viral polymerase. Strikingly, on the basis of amino acid homology, TSWV is more closely related to Bunyamwera virus than Hantaan virus is to this prototype bunyavirus. It may be anticipated that the amino acid homology between TSWV L protein and those of phlebo- and uukuviruses is even higher, since these viruses are even more closely related to TSWV, sharing similarly organized ambisense S RNA segments (de Haan *et al.*, 1990; Giorgi *et al.*, 1991).

The data presented furthermore imply that, based on molecular properties, such as terminal sequences, and based on the exclusive host range and mode of transmission, TSWV is indeed a member of a new distinct genus (tospovirus) within the Bunyaviridae.

The authors wish to thank Jan Gielen and Mart van Grinsven of Zaadunie Research, Enkhuizen, The Netherlands, for providing the synthetic oligonucleotides and for helpful discussions.

References

AUPERIN, D. D., COMPANS, R. W. & BISHOP, D. H. L. (1982). Nucleotide sequence conservation at the 3′ termini of the virion RNA species of New world and Old world arenaviruses. *Virology* **121**, 200–203.

BRAAM, J., ULMANEN, I. & KRUG, R. M. (1983). Molecular model of a eukaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell* **34**, 609–618.

CANDRESSE, T., MORCH, M. D. & DUNEZ, J. (1990). Multiple alignment and hierarchical clustering of conserved amino acid sequences in the replication-associated proteins of plant RNA viruses. *Research in Virology* **141**, 315–329.

CLERX, J. P. M., FULLER, F. & BISHOP, D. H. L. (1983). Tick-borne viruses structurally similar to Orthomyxoviruses. *Virology* **127**, 205–219.

CLERX-VAN HAASTER, C. M., CLERX, J. P. M., USHIJIMA, H., AKASHI, H., FULLER, F. & BISHOP, D. H. L. (1982). The 3′ terminal RNA sequences of bunyaviruses and nairoviruses (Bunyaviridae): evidence of end sequence generic differences within the virus family. *Journal of General Virology* **61**, 289–292.

DE HAAN, P., WAGEMAKERS, L., GOLDBACH, R. & PETERS, D. (1989*a*). Tomato spotted wilt virus, a new member of the Bunyaviridae? In *Genetics and Pathogenicity of Negative Strand Viruses*, pp. 287–290. Edited by D. Kolakofsky & B. W. J. Mahy. Amsterdam: Elsevier.

DE HAAN, P., WAGEMAKERS, L., PETERS, D. & GOLDBACH, R. (1989*b*). Molecular cloning and terminal sequence determination of the S and M RNA of tomato spotted wilt virus. *Journal of General Virology* **70**, 3469–3473.

DE HAAN, P., WAGEMAKERS, L., PETERS, D. & GOLDBACH, R. (1990). The S RNA segment of tomato spotted wilt virus has an ambisense characcharacter. *Journal of General Virology* **71**, 1001–1007.

DESSELBERGER, U., RACANIELLO, V. R., ZAZRA, J. J. & PALESE, P. (1980). The 3′- and 5′-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene* **8**, 315–328.

DEVOS, R. G. E., GILLIS, E. & FIERS, W. (1976). The enzymatic addition of poly(A) to the 3′ end of RNA using bacteriophage MS2 RNA as a model system. *European Journal of Biochemistry* **62**, 401–410.

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* **71**, 501–522.

GIORGI, C., ACCARDI, L., NICOLETTI, L., GRO, M. C., TAKEHARA, K., HILDITCH, C., MORIKAWA, S. & BISHOP, D. H. L. (1991). Sequences and coding strategies of the S RNAs of Toscana and Rift Valley fever viruses compared to those of Punta Toro, Sicilian sandfly fever, and Uukuniemi viruses. *Virology* **180**, 738–753.

GOLDBACH, R. (1986). Molecular evolution of plant RNA viruses. *Annual Review of Phytopathology* **24**, 289–310.

GOLDBACH, R. W. (1987). Genome similarities between plant and animal RNA viruses. *Microbiological Sciences* **4**, 197–202.

GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P. & BLINOV, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Research* **17**, 4713–4730.

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

HODGMAN, T. C. (1988). A new superfamily of replicative proteins. *Nature, London* **335**, 22–23.

IE, T. S. (1970). Tomato spotted wilt virus. *CMI/AAB Descriptions of Plant Viruses*, no. 39.

IHARA, T., AKASHI, H. & BISHOP, D. H. L. (1984). Novel coding strategy (ambisense genomic RNA) revealed by sequence analysis of Punta Toro phlebovirus S RNA. *Virology* **136**, 293–306.

IHARA, T., SMITH, J., DALRYMPLE, J. M. & BISHOP, D. H. L. (1985). Complete sequences of the glycoproteins and M RNA of Punta Toro phlebovirus compared to those of Rift Valley fever virus. *Virology* **144**, 246–259.

KAMER, G. & ARGOS, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Research* **12**, 7269–7282.

KRUG, R. M., ALONSO-CAPLEN, F. V., JULKUNEN, I. & KATZE, M. G. (1989). Expression and replication of the influenza virus genome. In *The Influenza Viruses*, pp. 89–152. Edited by R. M. Krug. New York: Plenum Press.

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

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.

MATTHEWS, R. E. F. (1982). Classification and nomenclature of viruses. *Intervirology* **17**, 1–199.

MOHAMED, N. A. (1981). Isolation and characterization of subviral structures from tomato spotted wilt virus. *Journal of General Virology* **53**, 197–206.

MOHAMED, N. A., RANDLES, J. W. & FRANCKI, R. I. B. (1973). Protein composition of tomato spotted wilt virus. *Virology* **56**, 12–21.

PARVIN, J. D., PALESE, P., HONDA, A., ISHIHAMA, A. & KRYSTAL, M. (1989). Promoter analysis of influenza virus RNA polymerase. *Journal of Virology* **63**, 5142–5152.

PETERS, D., DE AVILA, A. C., KITAJIMA, E. W., DE O. RESENDE, R., DE HAAN, P. & GOLDBACH, R. (1991). An overview of tomato spotted wilt virus. *Proceedings TSWW Workshop, Beltsville, U.S.A.* (in press).

PETERS, D., DE AVILA, A. C., KITAJIMA, E. W., DE O. RESENDE, R., DE HAAN, P. & GOLDBACH, R. (1991). An overview of tomato spotted wilt virus. In *Virus-Thrips-Plant Interactions of TSWV, Proceedings of USDA Workshop, Beltsville, U.S.A.*, pp. 1–14. Edited

by H.-T. Hsu & R. H. Lawson. Springfield: National Technology Information Service.

POCH, O., SAUVAGET, I., DELARUE, M. & TORDO, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO Journal* **8**, 3867–3874.

RAJU, R. & KOLAKOFSKY, D. (1989). The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base-paired. *Journal of Virology* **63**, 122–128.

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.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

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

SCHMALJOHN, C. S., JENNINGS, G. B., HAY, J. & DALRYMPLE, J. M. (1986). Coding strategy of the S genome segment of Hantaan virus. *Virology* **155**, 633–643.

SCHMALJOHN, C. S., SCHMALJOHN, A. L. & DALRYMPLE, J. M. (1987). Hantaan virus M RNA: coding strategy, nucleotide sequence, and gene order. *Virology* **157**, 31–39.

SIMONS, J. F., HELLMAN, U. & PETTERSSON, R. F. (1990). Uukuniemi virus S RNA segment: ambisense coding strategy, packaging of complementary strands into virions, and homology to members of the genus *Phlebovirus*. *Journal of Virology* **64**, 247–255.

STAUNTON, D., NUTTALL, P. A. & BISHOP, D. H. L. (1989). Sequence analysis of Thogoto viral RNA segment 3: evidence for a distant relationship between an arbovirus and members of the Orthomyxoviridae. *Journal of General Virology* **70**, 2811–2817.

STRAUSS, J. H. & STRAUSS, E. G. (1988). Evolution of RNA viruses. *Annual Review of Microbiology* **42**, 657–683.

TAS, P. W. L., BOERJAN, M. L. & PETERS, D. (1977). The structural proteins of tomato spotted wilt virus. *Journal of General Virology* **36**, 276–279.

VAN DEN HURK, J., TAS, P. W. L. & PETERS, D. (1977). The ribonucleic acid of tomato spotted wilt virus. *Journal of General Virology* **36**, 81–91.

YAMASHITA, M., KRYSTAL, M. & PALESE, P. (1989). Comparison of the three large polymerase proteins of influenza A, B and C viruses. *Virology* **171**, 458–466.

YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

ZHANG, H., SCHOLL, R., BROWSE, J. & SOMERVILLE, C. (1988). Double stranded DNA sequencing as a choice for DNA sequencing. *Nucleic Acids Research* **16**, 1220.

(Received 19 March 1991; Accepted 14 May 1991)