**Nucleotide sequence of RNA 1, the largest genomic segment of rice stripe virus, the prototype of the tenuiviruses**

Shigemitsu Toriyama,1* Mami Takahashi,1 Yoshitaka Sano,2 Takumi Shimizu1 and Akira Ishihama3

1 National Institute of Agro-Environmental Sciences, Kannondai 3, Tsukuba, Ibaraki 305, 2 Graduate School of Science and Technology, Niigata University, Ikarashi, Niigata 950-21 and 3 National Institute of Genetics, Mishima, Shizuoka 411, Japan

The complete nucleotide sequence of RNA 1, the largest genomic segment of rice stripe virus (RSV), was determined using two sets of overlapping cDNA clones. RNA segment 1 comprises 8970 nucleotides and on the viral complementary sequence has a single long open reading frame coding for a protein of 2919 amino acids with an estimated $M_r$ of 336860. Amino acid sequence comparisons of the putative protein indicated strong homology (30% amino acid identity over about 1500 residues) with the L protein of the genus *Phlebovirus* of the *Bunyaviridae*, but no detectable similarity with other members of the *Bunyaviridae*. However, weak similarity was detected with the L protein of Tacaribe arenavirus. The highly homologous sequence domain includes the conserved motifs of the putative RNA-dependent RNA polymerase. The data presented here, along with previous work clearly show significant similarities in genome organization, structure and expression between RSV and members of the genus *Phlebovirus* of the *Bunyaviridae*. Taken together, we propose that tenuiviruses should be included in the *Bunyaviridae* under the genus *Tenuivirus*.

**Introduction**

Rice stripe virus (RSV), the prototype of the genus *Tenuivirus*, has a broad host range in the *Gramineae* and causes serious damage to rice, particularly Japonica-type rice varieties (Toriyama, 1983; Francki et al., 1991). RSV is transmitted by the small brown planthopper *Laodelphax striatellus* Fallen, and planthoppers of three other species. In planthoppers, RSV replicates and is transovarially transmitted to a high percentage of the progeny (reviewed in Toriyama, 1986b). The genome of RSV comprises four ssRNA segments; as well as low levels of four dsRNAs, duplexes of vRNA and its complementary RNA can also be detected (Toriyama & Watanabe, 1989; Ishikawa et al., 1989). The dsRNAs found in tenuiviruses seem to be artifacts generated by annealing of complementary strands (Falk & Tsai, 1984). The complete nucleotide sequences have been determined for RNAs 3 and 4 from two different isolates (Kakutani et al., 1990, 1991; Zhu et al., 1991, 1992) and for RNA 2 from one isolate (Takahashi et al., 1993). The results suggest that all three RNA segments have ambisense coding strategies. This was also experimentally shown by *in vitro* translation of RNA transcribed from the cDNA sequences (Hamamatsu et al., 1993).

The 3'- and 5'-terminal sequences of approximately 18 nucleotides are conserved among all four RNA segments and are complementary to each other, except for one base change (U to A) at the sixth position from the 3' end of ssRSV RNA 1 (Takahashi et al., 1990). Moreover, eight terminal nucleotides out of ten conserved nucleotides are identical to those present in the terminal consensus sequences of the genus *Phlebovirus* of the family *Bunyaviridae* (Elliott, 1990; Elliott et al., 1991; Kakutani et al., 1990; Takahashi et al., 1990). Weak but significant amino acid sequence similarity exists between the nucleocapsid proteins from RSV and Punta Toro phlebovirus (Kakutani et al., 1990). Likewise, similarity exists between the putative $M_r$ 94K protein of RSV RNA segment 2 and the membrane glycoproteins of Punta Toro and Uukuniemi phleboviruses (Ihara et al., 1985; Rönnholm & Petterson, 1987; Takahashi et al., 1993). These observations suggest an evolutionary relationship between RSV and the phleboviruses.

The tenuiviruses include maize stripe virus (MStV), rice hoja blanca virus (RHBV), rice grassy stunt virus (RGSV) and three other possible members (Francki et al., 1991). The DDBJ accession number for the sequence of RSV RNA 1 is D31879.
Recent nucleotide sequencing studies of RNAs 3 and 4 of MStV (Huiet et al., 1991, 1992) and RNA 4 of RHBV (Ramirez et al., 1993) showed strong homology in the RNA 3 and 4 sequences between MStV, RHBV and RSV. The 18 nucleotide terminal sequences are conserved in RNAs 3 and 4 of MStV, RHBV and RSV. All of these RNAs have an ambisense coding strategy.

Filamentous particles of RSV and RGSV are associated with a high level of RNA-dependent RNA polymerase activity. A minor polypeptide, M, 230K, constituting purified filamentous virus particles of RSV and RGSV, is considered to be the RNA polymerase protein (Toriyama, 1986a, 1987). The largest genome segment, RNA 1, is presumed to encode this 230K RNA-dependent RNA polymerase, because this RNA segment alone is large enough to encode the 230K protein. In this paper, we present the nucleotide sequence of the RNA segment 1 of RSV. Analysis of the amino acid sequence of the predicted open reading frame reveals that RNA segment 1 does indeed encode the RNA polymerase. A high degree of homology was found between RSV RNA 1 and the L RNAs of phleboviruses. This homology was even greater than that detected for RNAs 2 and 3 of RSV.

Methods

**Virus and plant.** RSV isolate T was propagated in wheat plants with transmission by the viruliferous small brown planthopper *L. striatellus*, and purified as described previously (Toriyama, 1986a; Toriyama & Watanabe, 1989). The nB component, which contains RNA segment 1, was further purified at least twice by centrifugation on linear 5 to 35% sucrose gradients. RSV RNA was prepared as described previously (Toriyama, 1986a). ssRNA 1 was separated from RNAs 2, 3 and 4 by electrophoresis in 1% low-melting-point agarose gel (GTG agarose; Nakarai Chemicals) (Toriyama & Watanabe, 1989).

**cDNA synthesis and cloning.** vRNA-dependent cDNA synthesis was done by the method of Gubler & Hoffman (1983) using M-MLV reverse transcriptase lacking RNase H activity (BRL) and a synthetic oligonucleotide, primer A, with the sequence 5'-AGAGGAAAAAATAATTTTGA-3', which is complementary to the unique nucleotide sequence of RSV RNA segment 1, expressed as viral complementary sequence (cRNA). Other short overlapping clones, covering the 3' half of RNA 1 is composed of 8970 bases, with a base composition of 26.83% A, 34.33% U, 22.02% C and 16.82% G. The sequence was determined using two independent overlapping clones, except for the region between nucleotides 5233 and 5584 (this region was analysed only for clone pRS1S161). The sequences of both termini were obtained from the data determined by direct sequencing of viral RNA 1 (Takahashi et al., 1990). The complete nucleotide sequence of RSV RNA segment 1, expressed as viral complementary sense, is shown in Fig. 2. RNA segment 1 is composed of 8970 bases, with a base composition of 26.83% A, 34.33% U, 22.02% C and 16.82% G. The sequence was scanned for AUG-initiated open reading frames (ORFs). A single large ORF was detected in the viral complementary sequence (cRNA). Other short ORFs were identified on viral sense RNA (vRNA), which may encode M, 93K, 81K and 63K products. The large ORF present in cRNA extends from the 5'-proximal AUG codon at positions 58 to 60 to the UGA stop codon at position 8815 to 8817 (Fig. 2). The non-coding sequences are therefore 57 nucleotides at the 5' end and 153 nucleotides at the 3' end.

The amino acid sequence derived from the long ORF is shown in Fig. 2. The predicted gene product is 2919 amino acids long and has an estimated Mr of 336860.
Among the L proteins of the phleboviruses, the percentage of identical amino acids was 36.9% between which was based on the relative migration in an SDS-protein reveals 31.1% identical residues and 71.2%

This region contains the sequence of the putative RNA motifs proposed by Poch (1991) and Toscana virus (TOSV) (Accardi et al., 1992). At the amino acid level, the similarity was maximal between the RSV Pol protein and the L proteins of the phleboviruses (UUKV, RVFV and TOSV), as shown in the dot-plot analysis of protein homology (Fig. 3a, b, c). Among the L proteins of the phleboviruses, the percentage of identical amino acids was 36.9% between UUKV and RVFV (Fig. 3d), 35.8% between UUKV and TOSV, and 51.5% between RVFV and TOSV over the entire amino acid sequences. An optimal sequence alignment of the RSV Pol protein with the UUKV L protein reveals 31.1% identical residues and 71.2% overall similarity (including conserved amino acids). The similarity was maximal between residues 493 to 2026 (RSV) with only a few minor gaps, except for a 26 amino acid gap in the sequence of UUKV L protein between residues 1333 to 1362. The greatest similarity was between residues 1362 to 1931 (569 amino acids) where there is 39.3% identity and 78.3% similarity (Fig. 4). This region contains the sequence of the putative RNA polymerase domain, including the four polymerase motifs proposed by Poch et al. (1989) in the RNA-dependent RNA polymerase and identified in L proteins of UUK and RVF phleboviruses (Elliott et al., 1992). In addition, this region contains one distinct homologous stretch of 21 amino acid residues at 1408 to 1428 and a leucine zipper motif, L-X₆-L-X₆-L-X₆-L at residues 1531 to 1552, located roughly in the central portion of the RNA polymerase region. No significant homology was, however, found between RSV Pol and the L proteins of Bunyamwera, Hantaan and tomato spotted wilt viruses (Elliott, 1989; Schmaljohn, 1990; De Haan et al., 1991) (Fig. 3e, f). A weak similarity was found with the L protein of Tacaribe arenavirus: 19.2% identity over 449 amino acids in the region containing the polymerase motifs (Iapalucci et al., 1989) (Fig. 3g, h).

**Homologies of the RSV RNA segment 1 and the L RNA of phleboviruses**

The similarity search using the GenBank/EMBL nucleotide and NBRF/PIR protein databases showed clearly that RSV RNA segment 1 is homologous to the L RNA of phleboviruses, i.e. Uukuniemi virus (UUKV) (Elliott et al., 1992), Rift Valley fever virus (RVFV) (Muller et al., 1991) and Toscana virus (TOSV) (Accardi et al., 1993). At the amino acid level, the similarity was maximal between the RSV Pol protein and the L proteins of the phleboviruses (UUKV, RVFV and TOSV), as shown in the dot-plot analysis of protein homology (Fig. 3a, b, c). Among the L proteins of the phleboviruses, the percentage of identical amino acids was 36.9% between UUKV and RVFV (Fig. 3d), 35.8% between UUKV and TOSV, and 51.5% between RVFV and TOSV over the entire amino acid sequences. An optimal sequence alignment of the RSV Pol protein with the UUKV L protein reveals 31.1% identical residues and 71.2% overall similarity (including conserved amino acids). The similarity was maximal between residues 493 to 2026 (RSV) with only a few minor gaps, except for a 26 amino acid gap in the sequence of UUKV L protein between residues 1333 to 1362. The greatest similarity was between residues 1362 to 1931 (569 amino acids) where there is 39.3% identity and 78.3% similarity (Fig. 4). This region contains the sequence of the putative RNA polymerase domain, including the four polymerase motifs proposed by Poch et al. (1989) in the RNA-dependent RNA polymerase and identified in L proteins of UUK and RVF phleboviruses (Elliott et al., 1992). In addition, this region contains one distinct homologous stretch of 21 amino acid residues at 1408 to 1428 and a leucine zipper motif, L-X₆-L-X₆-L-X₆-L at residues 1531 to 1552, located roughly in the central portion of the RNA polymerase region. No significant homology was, however, found between RSV Pol and the L proteins of Bunyamwera, Hantaan and tomato spotted wilt viruses (Elliott, 1989; Schmaljohn, 1990; De Haan et al., 1991) (Fig. 3e, f). A weak similarity was found with the L protein of Tacaribe arenavirus: 19.2% identity over 449 amino acids in the region containing the polymerase motifs (Iapalucci et al., 1989) (Fig. 3g, h).

**Discussion**

The nucleotide sequences of the RNAs 2, 3 and 4 of RSV isolate T have been reported previously (Zhu et al., 1991, 1992; Takahashi et al., 1993). The determination of the sequence of RNA 1 completes the genome sequence of RSV. As summarized in Fig. 5, the complete genome comprises 17145 nucleotides, of which 86.8% code for seven ORFs. Each of RNAs 2, 3 and 4 has an ambisense coding strategy, and RNA segment 1 is a negative strand RNA. MStV, another member of the tenuiviruses, contains five RNA segments, of which the smallest RNA (RNA 5) is a negative strand and encodes a highly basic protein (Huiet et al., 1993). Although the existence of a small RNA has been reported for a different isolate of RSV (Ishikawa et al., 1989), we have been unable to find such a small distinct RNA in RSV (Toriyama, 1982; Toriyama & Watanabe, 1989). Furthermore, a purified preparation of RSV containing the four RNA species alone reproduced the original chlorotic stripe symptoms on rice seedlings, when inoculated through the plant-vector (Toriyama, 1982). Among the seven putative viral-coded proteins, the nucleocapsid protein and a non-structural protein (S-protein) were shown to be encoded by cRNA 3 and vRNA 4, respectively (Hamamatsu et al., 1993). We have now shown that the predicted Pol protein (3368K) is encoded by cRNA 1. This Pol protein is most probably the previously designated 230K protein which is associated with RSV nucleoproteins and was considered to be a putative RNA
Complete sequence of the RSV genome

Fig. 2. For legend see page 3575.
polymerase (Toriyama, 1986a). A discrepancy in $M_i$ values has also been reported for the L (RNA polymerase) proteins of Bunyamwera and tomato spotted wilt viruses (Elliott, 1989; De Haan et al., 1991). The other putative viral proteins predicted from the sequence have not yet been identified, although other four putative viral proteins predicted from the genome products (Hamamatsu et al., 1992), and the second SDD motif is present in the L protein of segmented negative strand viruses (Poch et al., 1989). Other prominent conserved sequences are found in the extreme 5'- and 3'-terminal nucleotide sequences of RSV and phleboviruses (Kakutani et al., 1990; Takahashi et al., 1990). The terminal base-paired, pan-handle structure (Takahashi et al., 1990) is presumed to have an important role in the initiation of transcription of influenza virus (Hsu et al., 1987). Thus, it is likely that these highly conserved sequences are essential for replication and transcription of these viruses.

An additional distinguishing similarity of RSV and phleboviruses is that they have an ambisense genome: RNAs 2, 3 and 4 of RSV, and the S RNA of phleboviruses. The intergenic sequence of the S RNA centrally located between two ORFs is G-rich in three phleboviruses, TOSV, Sicilian sandfly fever and RVFV, and AU-rich in the other two phleboviruses, UUKV and Punta Toro virus (Giorgi et al., 1991). The intergenic sequences of the ambisense genome RNA 3 of the tenuiviruses RSV and MStV are all AU-rich (Kakutani et al., 1991; Zhu et al., 1991; Huiet et al., 1991).
suggesting that tenuiviruses, in this respect, are similar to UUK or Punta Toro phleboviruses.

Given the strong similarity between RSV and phleboviruses, we propose that RSV and the other tenuiviruses, MSTV, RHBV and RGSV (Francki et al., 1991), should be classified in the family Bunyaviridae, but in the genus Tenuivirus not Phlebovirus. This is because the genome of tenuiviruses comprises four segments (RSV and RHBV) or five segments (MSTV), while all phleboviruses have three RNA segments, although the genetic organization, expression strategies, amino acid and nucleotide sequence similarities strongly suggest that these viruses have evolved from a common ancestor.

No significant homology was observed between the Pol protein and the L protein of other members of the Bunyaviridae, including tomato spotted wilt virus, previously the sole plant-infecting member of the Bunyaviridae (De Haan et al., 1991). A weak homology was found with the L protein of Tacaribe virus of the Arenaviridae (Iapalucci et al., 1989). It is probable that Tenuivirus is in a unique position in the evolution of these ambisense genome viruses.

One of the differences between tenuiviruses and other members of the Bunyaviridae is viral particle morphology. Virions of tenuiviruses are thin filamentous particles which are pleomorphic: partially or completely unfolded coiled filaments, branched configurations, or circular filaments (Koganezawa et al., 1975; Toriyama, 1982; Ishikawa et al., 1989). So far, enveloped spherical particles (the morphology of virions of other Bunyaviridae) have not been observed for tenuiviruses, despite extensive examination by electron microscopy of infected plant and insect tissues. Immunogold labelling with anti-IgG to nucleoprotein of RSV resulted in labelling of amorphous or membranous structures in the cytoplasm of the small brown planthopper L. striatellus (Suzuki et al., 1992). Observations of thin filamentous particles or circular filamentous particles of tenuiviruses seem to
Fig. 4. Amino acid sequence homology between the predicted protein Pol of RSV and the L protein of UUKV. Identical residues are indicated by two dots and are shaded; consensus amino acid similarities are indicated by one dot. Gaps inserted in the sequences to maximize homology are indicated by dashes. Sequence data were obtained from Elliott et al. (1992), in which the RNA polymerase motifs of UUKV were indicated by underlining in the amino acid sequence.

Fig. 5. Genome structure and coding arrangement of RSV. Black lines are genomic RNAs, with the nucleotide numbers on both ends. The open reading frame and its direction are indicated with arrows on vRNA and cRNA. Shaded arrows show that the corresponding proteins have been found: Pol, the putative RNA polymerase protein (probably the 230K protein); N, nucleocapsid; Ns, non-structural protein (S protein).
suggest that these particles might correspond to the nucleocapsids of enveloped viruses of Bunyaviridae or Arenaviridae (von Bondorf et al., 1969; Palmer et al., 1977).

Grateful acknowledgement is made to Professor M. Kojima of Niigata University for his continuing interest, encouragement and discussions throughout this work. We also thank Dr N. Ogasawara of Plant Biological Defence System Lab. Co., Ltd for providing DNA sequencing facilities, and Dr T. Ogasawara of Hitachi Software Engineering Co., Ltd for the homology search of nucleotide and protein sequences. This work was supported in part by Grants-in-Aid for Scientific Research (C) (No. 16653005) from the Ministry of Education, Science, Sports and Culture of Japan.

References


TORIYAMA, S. (1986b). Rice stripe virus: prototype of a new group of


(Received 7 June 1994; Accepted 19 August 1994)