

# Nucleotide sequence of RNA2 of *Lettuce big-vein virus* and evidence for a possible transcription termination/initiation strategy similar to that of rhabdoviruses

Takahide Sasaya,<sup>1</sup> Shinnosuke Kusaba,<sup>1</sup> Koichi Ishikawa<sup>1</sup>  
and Hiroki Koganezawa<sup>2</sup>

Correspondence  
Takahide Sasaya  
tsasaya@affrc.go.jp

<sup>1</sup>National Agricultural Research Center for Western Region, Shikoku Campus, Zentsuji, Kagawa 765-8508, Japan

<sup>2</sup>National Agricultural Research Center for Western Region, Fukuyama, Hiroshima 721-8514, Japan

*Lettuce big-vein virus* (LBVV) is the type species of the genus *Varicosavirus* and is a two-segmented negative-sense single-stranded RNA virus. The larger LBVV genome segment (RNA1) consists of 6797 nt and encodes an L polymerase that resembles that of rhabdoviruses. Here, the nucleotide sequence of the second LBVV genome segment (RNA2) is reported. LBVV RNA2 consisted of 6081 nt and contained antisense information for five major ORFs: ORF1 (nt 210–1403 on the viral RNA), ORF2 (nt 1493–2494), ORF3 (nt 2617–3489), ORF4 (nt 3843–4337) and ORF5 (nt 4530–5636), which had coding capacities of 44, 36, 32, 19 and 41 kDa, respectively. The gene at the 3' end of the viral RNA encoded a coat protein, while the other four genes encoded proteins of unknown functions. The 3'-terminal 11 nt of LBVV RNA2 were identical to those of LBVV RNA1, and the 5'-terminal regions of LBVV RNA1 and RNA2 contained a long common nucleotide stretch of about 100 nt. Northern blot analysis using probes specific to the individual ORFs revealed that LBVV transcribes monocistronic RNAs. Analysis of the terminal sequences, and primer extension and RNase H digestion analysis of LBVV mRNAs, suggested that LBVV utilizes a transcription termination/initiation strategy comparable with that of rhabdoviruses.

Received 26 February 2004  
Accepted 30 April 2004

## INTRODUCTION

Lettuce big-vein is a fungus-borne disease that is a serious problem in many lettuce-growing areas. *Lettuce big-vein virus* (LBVV), which is almost always isolated from lettuce plants (*Lactuca sativa* L.) showing big-vein symptoms, has been considered the causal agent of the disease for nearly two decades (Kuwata *et al.*, 1983; Vetten *et al.*, 1987; Huijberts *et al.*, 1990). However, a second virus, Mirafiori lettuce virus in the genus *Ophiovirus*, was recently found in lettuce showing big-vein symptoms, so the exact role of LBVV in big-vein disease remains unclear (Roggero *et al.*, 2000; Lot *et al.*, 2002; Natsuaki *et al.*, 2002).

LBVV is the type species of the genus *Varicosavirus*, which also includes Tobacco stunt virus, Camellia yellow mottle virus and Freesia leaf necrosis virus (Mayo, 2000). LBVV virions are rod-shaped with modal lengths of 320–360 nm and diameters of about 18 nm. LBVV virions contain a

single coat protein (CP) with an  $M_r$  of 48 000 and do not have an envelope structure (Kuwata *et al.*, 1983). LBVV was reported to be a two-segmented double-stranded RNA (dsRNA) virus, which contained two dsRNAs of approximately 7.0 and 6.5 kbp (Mayo, 2000). However, we recently demonstrated that LBVV is not a dsRNA virus but a single-stranded negative-sense RNA virus with a bipartite genome. Viral (negative-sense) and virus-complementary (positive-sense) RNAs are separately encapsidated in the virions and the viral RNA is predominant (Sasaya *et al.*, 2001, 2002).

LBVV was found in lettuce plants showing big-vein symptoms in the early 1980s (Kuwata *et al.*, 1983; Vetten *et al.*, 1987; Huijberts *et al.*, 1990), but its molecular characterization has been slow because LBVV particles are refractory to standard methods for virus purification due to their instability and tendency to aggregate. After numerous efforts, we succeeded in obtaining purified LBVV virions from lettuce plants and determining the nucleotide sequence of the CP gene on RNA2 and the L polymerase gene (designated by analogy with the L polymerase of rhabdoviruses, which is a catalytic subunit of the RNA polymerase

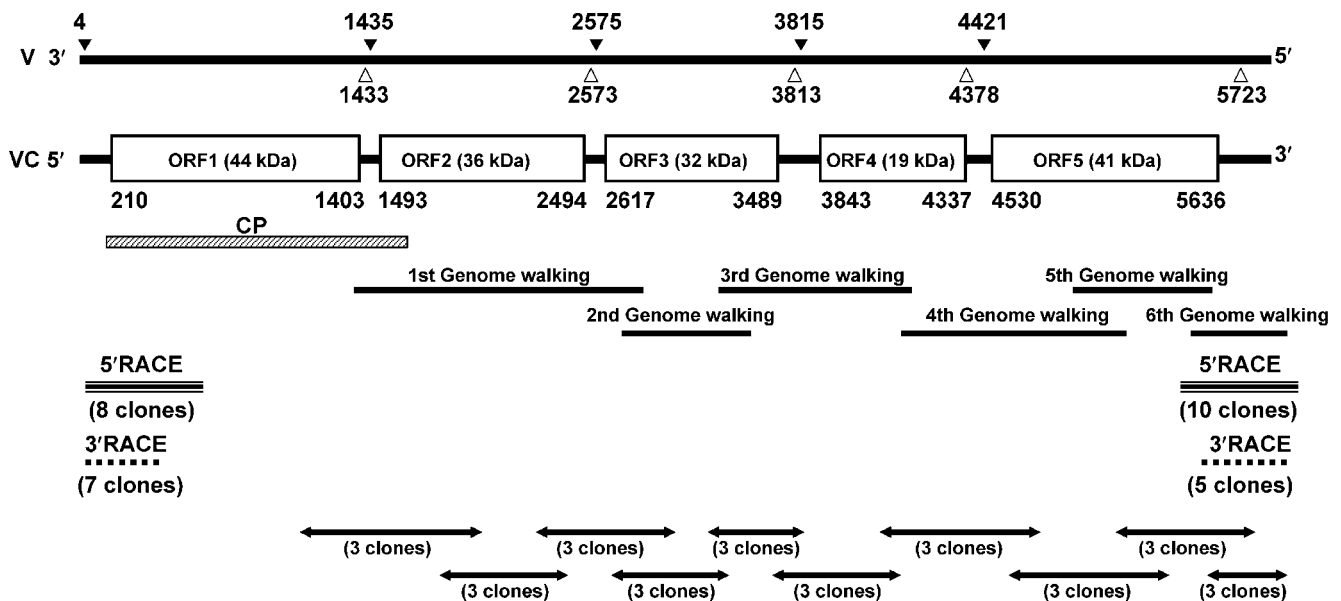
The GenBank/EMBL/DDBJ accession number of the sequence reported in this article is AB114138.

complex) on RNA1 (Sasaya *et al.*, 2001, 2002). Analysis of amino acid sequences of the LBVV CP and L polymerase suggested a close relationship between LBVV and rhabdoviruses in the order *Mononegavirales*, even though LBVV has a two-segment negative-sense RNA genome and its particles are not enveloped. In particular, the LBVV L polymerase is most closely related to that of *Northern cereal mosaic virus* (NCMV) in the genus *Cytorhabdovirus* (Sasaya *et al.*, 2002). In this report, we present the nucleotide sequence of LBVV RNA2 and the predicted non-coding and coding regions. We also determined the 5' and 3' ends of LBVV mRNAs to get an insight into the possible transcription mechanism of LBVV. Our data provide further evidence for a relationship between LBVV and rhabdoviruses, as well as illustrating some unique properties of LBVV.

## METHODS

**Virus and RNA purification.** The source of LBVV, procedure for virus purification and isolation of viral RNAs have been described previously (Sasaya *et al.*, 2001, 2002). Poly(A)<sup>+</sup> RNA was isolated from healthy and LBVV-infected lettuce plants 4 weeks after transplanting to stream-sterilized and LBVV-infested soils, respectively, using a Dynabeads mRNA Direct kit, following the instructions provided by the supplier (Dynal).

**Cloning strategy for LBVV RNA2.** A cDNA library of LBVV RNA2 was constructed and its sequence was determined by a genome-walking method, as described by Fazeli & Rezaian (2000). The cloning strategy for LBVV RNA2 and the synthetic oligonucleotide primers used for the cloning are shown in Fig. 1 and Table 1. Briefly, first-strand cDNA synthesis was done using 20 ng purified LBVV genomic RNAs primed with the LBVV-specific first-strand cDNA synthesis primer 1<sup>st</sup> GW, whose sequence corresponded to nt 1135–1151 of the previously described LBVV CP gene (Sasaya *et al.*, 2001). The second-strand cDNA synthesis primer P1-N<sub>6</sub>, which consisted of a random hexamer linked to a known oligonucleotide at its 5' end, was added to the first-strand reaction mixture and primer extension was carried out at 37 °C for 30 min in a 50 µl reaction volume containing 50 mM Tris/HCl, pH 7.2, 10 mM MgSO<sub>4</sub>, 0.1 mM DTT, 0.3 mM each of the four dNTPs and 5 U Klenow fragment (Toyobo). After excess oligonucleotide had been removed by centrifugal ultrafiltration (Centricon 100 microconcentrators; Amicon), double-strand cDNA was amplified by PCR with the known oligonucleotides P1 and the first LBVV-specific genome-walking primer, GW-1<sup>P</sup>. The PCR product was cloned and sequenced as described previously (Sasaya *et al.*, 2002). The next region of LBVV RNA2 was amplified by PCR using the P1 primer and a new LBVV-specific primer, which was designed on the basis of the sequence information obtained from the previous cloning step. In addition, the sequence of the entire region of LBVV RNA2 was confirmed by recloning the LBVV genomic RNAs using the LBVV-specific primers (the sequence of each primer is not shown) designed on the basis of the sequence information obtained by the genome-walking method.



**Fig. 1.** Schematic representation of LBVV RNA2 and the cloning strategies used in this work. Nucleotide positions of LBVV transcription initiation and termination on the viral RNA (V) are indicated by ▼ and △, respectively. The five major ORFs on the virus-complementary RNA (VC) are represented by boxes, and the nucleotide positions of the ORF start and end are indicated below the boxes. The six cDNA clones derived from the genome-walking method are shown as solid lines. The cDNA CP clone (Sasaya *et al.*, 2001) was the starting point for the genome-walking method (hatched line). cDNA clones of both ends generated by 5'RACE and obtained by 3'RACE after polyadenylation of the LBVV genomic RNAs are shown as triple and dotted lines, respectively. cDNA clones resulting from direct RT-PCR using virus-specific primers in the second round of cloning are shown as arrows. The number of independent cDNA clones used for sequencing is shown in parentheses under the lines.

**Table 1.** Synthetic oligonucleotide primers used in this study

LBVV-specific primers used for the second round of LBVV genome cloning and 5'RACE and 3'RACE for each LBVV mRNA have been omitted.

Primer*	Nucleotide position	Primer sequence (3'→5')	Usage
1 <sup>st</sup> GW	1135–1151	ATACCGTTGCCATCTTG	First-strand cDNA synthesis for genome walking
GW-1 <sup>P</sup>	1342–1358	TAGAGTCCGACGATGAG	First genome walking
GW-2 <sup>P</sup>	2543–2561	TGTACCATGCTGTCGTC	Second genome walking
GW-3 <sup>P</sup>	3090–3108	CTCGTTAGACCCCAATTAC	Third genome walking
GW-4 <sup>P</sup>	4045–4063	AGCTGTTCCGAGAGAGGAGA	Fourth genome walking
GW-5 <sup>P</sup>	5053–5071	CTTCTTATTACTCTGTCCG	Fifth genome walking
GW-6 <sup>P</sup>	5535–5553	GCTGTGATGAGTGTCT	Sixth genome walking
p-5TA-R2-1 <sup>n</sup>	505–519	CATCCACATCCGCTG	First cDNA synthesis for 5'RACE of virus-complementary RNA
5TA-R2-2 <sup>P</sup>	384–401	GGGCTATTTGTGAGTCC	First PCR for 5'RACE of virus-complementary RNA
5TA-R2-3 <sup>n</sup>	351–369	TCACCTCACGAACAGCATC	First PCR for 5'RACE of virus-complementary RNA
5TA-R2-1 <sup>P</sup>	447–464	AGGTCTGTTGATCCCGTG	Second PCR for 5'RACE of virus-complementary RNA
5TA-R2-4 <sup>n</sup>	289–306	TGCTTTCATCATCCCAGG	Second PCR for 5'RACE of virus-complementary RNA and PCR for 3'RACE of viral RNA
p-3TA-R2-1 <sup>P</sup>	5248–5262	CCATAGCCCGAGAAG	First-strand cDNA synthesis for 5'RACE of viral RNA
3TA-R2-1 <sup>P</sup>	5825–5842	GGGACGGATGAAAGAGAG	First PCR for 5'RACE of viral RNA
LB3-6 <sup>n</sup>	5654–5673	TGGGAGATATAGTGTTC	First PCR for 5'RACE of viral RNA
3TA-R2-2 <sup>P</sup>	5899–5916	CTGGAATGAGAAGTATGG	Second PCR for 5'RACE of viral RNA and PCR for 3'RACE of virus-complementary RNA
3TA-R2-3 <sup>n</sup>	5535–5553	AGACAGCACTCATCACAGC	Second PCR for 5'RACE of viral RNA
5'-PrEx <sup>n</sup>	1528–1549	GATGATGCAATGGCCTTCGTC	Primer extension analysis for the 5' end of the ORF2 mRNA
RNase dig1 <sup>n</sup>	1407–1426	ACATATGATAGTAGGATCCC	RNase H analysis for the 3' end of the CP mRNA
RNase dig1 <sup>P</sup>	1389–1408	CCAGTGAAGGAGTGATGAGG	RNase H analysis for the 3' end of the CP mRNA
RNase dig2 <sup>n</sup>	1435–1454	TACCTATAGTTGTCTTTCTC	RNase H analysis for the 3' end of the CP mRNA
P1-N <sub>6</sub>	–	GCCGGAGCTCTGCAGAATTC(N) <sub>6</sub>	Second cDNA synthesis for genome walking
P2-T <sub>17</sub>	–	GGCCACGCGTCGACTAGTAC(T) <sub>17</sub>	First-strand cDNA synthesis for 3'RACE of viral RNA
5'-p-Adapter	–	CAATACCTTCTGACCATGCAGT-GACAGTCAGCATG	Adapter to the 3' end of mRNA
1 <sup>st</sup> -3'-R	–	CATGCTGACTGACACTGCAT	First-strand cDNA synthesis and first PCR for 3'RACE of mRNA
2 <sup>nd</sup> -3'-R	–	TCGATGGTCAGAAGGTATTG	Second PCR for 3'RACE of mRNA

\*P, Positive-sense complementary to viral RNA; <sup>n</sup>, negative-, viral RNA-sense.

Since the purified LBVV RNA preparation contained viral and virus-complementary RNAs (Sasaya *et al.*, 2001), the exact ends of the LBVV RNA2 were verified by a 5'RACE system using the 5'-Full RACE Core Set (TaKaRa Bio) as described previously (Sasaya *et al.*, 2002). To determine the 3' end of the viral RNA of LBVV RNA2, LBVV genomic RNAs were reverse-transcribed using the 5'-end phosphorylated primer p-5TA-R2-1<sup>n</sup>. First-strand cDNA synthesis was followed by circularization of the 5' and 3' ends of the first-strand cDNA using RNA ligase (TaKaRa Bio). Subsequently the ligated junction of the 5' and 3' ends was followed first by PCR amplification with virus-specific primers 5TA-R2-2<sup>P</sup> and 5TA-R2-3<sup>n</sup> and then by PCR amplification with virus-specific primers 5TA-R2-1<sup>P</sup> and 5TA-R2-4<sup>n</sup>. For the 5' end, first-strand cDNA was synthesized using the 5'-end phosphorylated primer p-3TA-R2-1<sup>P</sup>. Two pairs of virus-specific primers, 3TA-R2-1<sup>P</sup> with LB3-6<sup>n</sup>, and 3TA-R2-2<sup>P</sup> with 3TA-R2-3<sup>n</sup>, were used for first- and second-round PCR amplifications, respectively.

The sequences of both ends were also confirmed by 3'RACE after polyadenylation of the LBVV genomic RNAs as described previously (Sasaya *et al.*, 2002). First-strand cDNA was synthesized from the 3'-polyadenylated LBVV RNAs using the first-strand cDNA synthesis

primer P2-T<sub>17</sub>, which consisted of oligo(dT) linked to a known oligonucleotide at its 5' end. cDNA was amplified by PCR with the known oligonucleotides P2 and an LBVV-specific primer, either 5TA-R2-4<sup>n</sup> for the 3' end or 3TA-R2-2<sup>P</sup> for the 5' end of LBVV RNA2.

**Determination of the 5' and 3' ends of LBVV mRNAs.** To determine the 5'-terminal nucleotide sequences of the LBVV mRNAs, including a small ORF (designated gene 6) preceding the L polymerase gene and the L polymerase gene on RNA1 (Sasaya *et al.*, 2002), and to ascertain whether the LBVV mRNA molecules had a cap structure at the 5' end, cDNAs of the corresponding mRNAs were obtained using the FirstChoice RLM-RACE kit (Ambion), which specifically selects capped transcripts from non-capped RNAs using calf intestinal phosphatase (CIP) and tobacco acid pyrophosphatase (TAP) treatment (Liu & Gorovskiy, 1993). Briefly, isolated poly(A)<sup>+</sup> RNA from LBVV-infected lettuce plants was treated with CIP to remove the 5' phosphate from all poly(A)<sup>+</sup> RNA molecules except those that had a cap structure. TAP was then used to remove the cap structure from the poly(A)<sup>+</sup> RNA, leaving a 5'-monophosphate. An RNA adapter oligonucleotide in the kit was ligated in the presence of T4 RNA ligase to the 5' end of the

CIP/TAP-treated poly(A)<sup>+</sup> RNA. The ligated RNA was reverse-transcribed using a random primer, followed by first- and second-round PCR amplification with the 5'RACE outer primer in the kit and the first LBVV-specific 5'RACE mRNA primer (sequence not shown), which was designed on the basis of the LBVV genome sequence specific to the individual gene, and the 5'RACE inner primer in the kit and the second LBVV-specific 5'RACE mRNA primer.

To determine the nucleotide sequence at the 3' ends of all seven LBVV mRNAs and to ascertain the actual size of the poly(A) tails, the 5'-end phosphorylated adapter 5'-p-Adapter was ligated to the 3' terminus of poly(A)<sup>+</sup> RNA using T4 RNA ligase, as described by Liu & Gorovsky (1993). After excess oligonucleotide had been removed by centrifugal ultrafiltration, first-strand cDNA was synthesized using the first 3'RACE primer 1<sup>st</sup>-3'-R, whose sequence is complementary to that of 5'-p-Adapter. cDNA was amplified by PCR with 1<sup>st</sup>-3'-R and the first LBVV-specific 3'RACE mRNA primer (sequence not shown), which was designed on the basis of the LBVV genome sequence specific to the individual gene, followed by PCR amplification with the second 3'RACE primer 2<sup>nd</sup>-3'-R and the LBVV-specific 3'RACE mRNA primer.

**Northern blot hybridization.** Poly(A)<sup>+</sup> RNA samples from healthy and LBVV-infected lettuce plants were analysed by Northern blot hybridization as described previously (Kusaba *et al.*, 1998). After heat denaturation in formaldehyde/formamide, poly(A)<sup>+</sup> RNA was separated on 1.5% agarose horizontal submarine gels in MOPS/EDTA buffer containing 0.22 M formaldehyde and transferred to Hybond<sup>+</sup> nitrocellulose membranes (Amersham-Pharmacia Biotech) by capillary blotting in 20 × SSC. Gel-purified inserts obtained from LBVV cDNA clones harbouring each LBVV gene were labelled with [ $\alpha$ -<sup>32</sup>P]dCTP using a random primer labelling kit (Amersham-Pharmacia Biotech). Prehybridization, hybridization, washings and membrane exposures were carried out according to standard protocols (Sambrook & Russell, 2000).

**Primer extension analysis of the 5' end of LBVV ORF2 mRNA.** The 5' end of the ORF2 mRNA was mapped by primer extension analysis on poly(A)<sup>+</sup> RNA from the LBVV-infected lettuce plants. Two micrograms of poly(A)<sup>+</sup> RNA from healthy and LBVV-infected lettuce plants was reverse-transcribed with SuperScript II RNase H<sup>-</sup> Reverse Transcriptase (Invitrogen) using the LBVV-specific primer 5'-PrEx<sup>n</sup> and [ $\alpha$ -<sup>32</sup>P]dCTP. The radio-labelled cDNA was separated on an 8% polyacrylamide urea gel next to the products of a sequencing reaction of single-stranded cDNA from a plasmid harbouring the gene junction region between the CP and ORF2 using the same oligonucleotide primer.

**RNase H analysis of the 3' end of LBVV CP mRNA.** The 3' end of the CP mRNA was analysed by electrophoretic mobility shifts of Northern blot hybridizations following RNase H digestion, as described by Masters & Samuel (1984). Five micrograms of poly(A)<sup>+</sup> RNA from LBVV-infected lettuce plants was incubated for 20 min at 25 °C in a volume of 30  $\mu$ l reaction mixture containing 20 mM Tris/HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 100 mM KCl and 0.1 mM DTT, with or without 1.0  $\mu$ g of appropriate oligonucleotide primers. RNase H digestions were performed by the addition of 2.0 U RNase H (Gibco-BRL) and incubation for 30 min at 37 °C. Digested RNAs were recovered by phenol/chloroform extraction and ethanol precipitation prior to Northern blot hybridization using the LBVV CP-specific probe. The oligonucleotide primers used were as follows: RNase dig1<sup>n</sup>, which annealed to the virus-complementary RNA from nt 1407 to 1426 of the LBVV RNA2; RNase dig1<sup>p</sup>, which annealed to the viral RNA from nt 1389 to 1408; RNase dig2<sup>n</sup>, which annealed to the virus-complementary RNA from nt 1435 to 1454 and presumably does not exist in the CP mRNA; and oligo(dT)<sub>12-18</sub>, which annealed to the poly(A) sequence.

## RESULTS AND DISCUSSION

### Coding sequence of LBVV RNA2

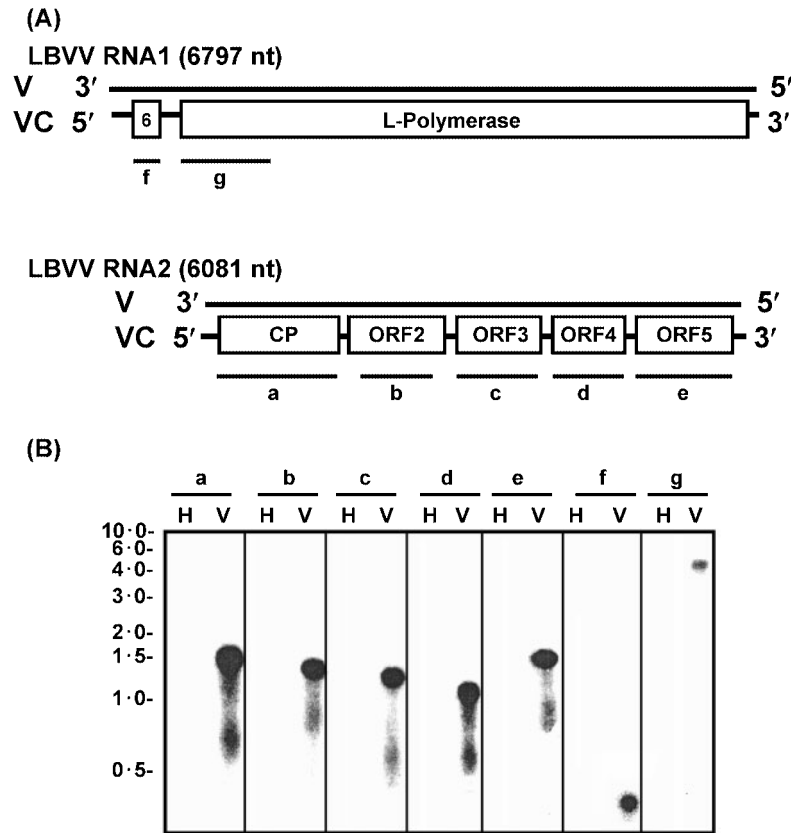
Using the strategy described in Fig. 1, a total of six different PCR amplification steps were carried out to determine the entire nucleotide sequence of LBVV RNA2. LBVV RNA2 consisted of 6081 nt and contained antisense information for five major ORFs, which were designated ORF1 (nt 210–1403), ORF2 (nt 1493–2494), ORF3 (nt 2617–3489), ORF4 (nt 3843–4337) and ORF5 (nt 4530–5636) (Fig. 1). ORF1 corresponded to the previously described LBVV CP, which was distantly similar to the nucleocapsid protein of rhabdoviruses (Sasaya *et al.*, 2001). The remaining four ORFs had coding capacities of 36, 32, 19 and 41 kDa, respectively.

The overall genomic organization of LBVV resembled that of plant rhabdoviruses (genera *Nucleorhabdovirus* and *Cytorhabdovirus*). Plant rhabdoviruses contain at least six genes encoding the nucleocapsid protein, phosphoprotein, non-structural protein, matrix protein, glycoprotein and L polymerase (Scholthof *et al.*, 1994; Wetzel *et al.*, 1994; Chen *et al.*, 1998; Tanno *et al.*, 2000). The matrix protein and glycoprotein form the viral envelope and are involved in the movement of virus into and out of host cells. The nucleocapsid protein, phosphoprotein and L polymerase form the viral nucleocapsids and are involved in expressing and replicating the virus genome (Conzelmann, 1998; Walker *et al.*, 2000; Rose & Whitt, 2001; Barr *et al.*, 2002). In the case of LBVV, its genome appeared to be split between the glycoprotein and L polymerase genes, with the L polymerase gene present on RNA1 and the other five genes present on RNA2 (see Fig. 3A). Although previous analysis of the LBVV CP and L polymerase revealed sequence similarity to the nucleocapsid proteins and the L polymerases of rhabdoviruses, respectively (Sasaya *et al.*, 2001, 2002), FASTA and BLAST analyses of translated sequences from the four other LBVV genes showed no apparent similarity to other known viral sequences. However, this was not surprising since the primary sequences of the phospho-, non-structural, matrix and glycoproteins are poorly conserved among plant rhabdoviruses. The overall amino acid sequence identities of corresponding genes are less than 20% (Heaton *et al.*, 1987; Hillman *et al.*, 1990; Scholthof *et al.*, 1994; Chen *et al.*, 1998; Luo & Fang, 1998; Luo *et al.*, 1998; Tanno *et al.*, 2000).

### 5'- and 3'-terminal sequences of the LBVV genome

The 5' and 3' ends of the genomic RNA of rhabdoviruses are complementary and are thought to represent conserved promoters for RNA synthesis (Conzelmann, 1998). Complementarity between the extreme 5' and 3' ends of the LBVV RNA2 was low (Fig. 2A). Instead, nt 12–24 and 6047–6059 of LBVV RNA2 showed a 13 nt stretch of complementarity with a single mismatch, similar to that in LBVV RNA1 (Sasaya *et al.*, 2002). A close examination





**Fig. 3.** Northern blot analysis of poly(A)<sup>+</sup> RNAs from LBVV-infected lettuce plants. (A) Schematic representation of the genomic organization of LBVV and the location of probes. The six major LBVV ORFs and a small ORF upstream of the L polymerase gene are represented by boxes. V and VC indicate viral and virus-complementary RNAs, respectively. The probes representing each LBVV gene are indicated by a–g. (B) Northern blot analysis. Poly(A)<sup>+</sup> RNAs from healthy or LBVV-infected lettuce plants are indicated by H and V, respectively. Each lane was hybridized with the probes indicated in (A). RNA molecular size markers (kb) (Perfect RNA Marker; Novagen) are shown on the left.

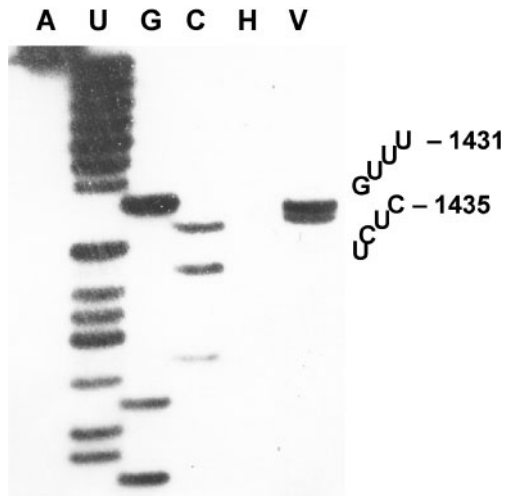
(A)

LBVV RNA2	I	II	III
3'-non/CP	(ac)	G	CUCUGCGUUUU
CP/2	UAUACAUUUUUUU	G	CUCUUUCUGUU
2/3	UAUACAUUUUUUU	G	CUCUUUAUCUC
3/4	UAUACAUUUUUUU	G	CUCUCACAUUC
4/5	AUAAUCUUUUUUU	GN <sub>40</sub> A	CUCUCGUAUA
5/5'-non	AUAAUCUUUUUUU	G	(cuacaggguaa)
LBVV RNA1			
3'-non/6	(ac)	G	CUCUGCGUAUA
6/L	UAUGCAUUUUUUU	G	CUCUUGUUUUUA
L/5'-non	AUAAUCUUUUUUU	G	(cuugggucua)
Consensus Type A	UAUNCAUUUUUUU		
Type B	AUAAUCUUUUUUU	G	CUCUNNNWNU

(B)

Consensus	I	II	III
LNyV consensus	ACUAAAUCUUUUU	GN <sub>3-24</sub> A	CUCUUA
NCMV consensus	AUUCUUUUU	GA	CUCUAGU
SYNV consensus	AUUCUUUUU	GG	UUGW
RYSV consensus	AUUUUUUU	GGG	UUGUG
VSIV consensus	AUACUUUUUUU	GA	UUGUCNNUAG
RABV consensus	GYACUUUUUUU	GN <sub>0-22</sub> R	UUGURRNGA

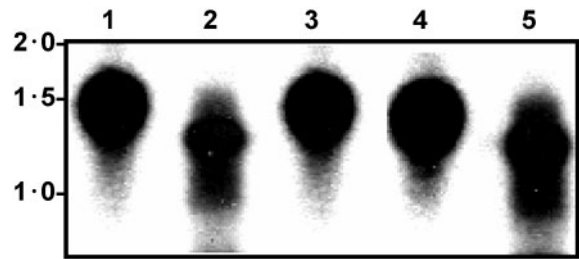
**Fig. 4.** Regulatory sequences of the LBVV genome. (A) LBVV genomic RNA sequences corresponding to the gene junctions. I, Sequences corresponding to the 3' end of each mRNA; II, intergenic regions; III, sequences corresponding to the 5' end of each mRNA. Transcription signals are shown in the viral RNA sense. (B) Consensus transcription termination/polyadenylation, intergenic region and initiation signals of four genera of rhabdoviruses, *Cytorhabdovirus*, *Nucleorhabdovirus*, *Vesiculovirus* and *Lyssavirus*. Consensus sequences were derived as described previously (Tordo *et al.*, 1986; Heaton *et al.*, 1989; Wetzel *et al.*, 1994; Chen *et al.*, 1998; Tanno *et al.*, 2000). LNyV, *Lettuce necrotic yellows virus*; NCMV, *Northern cereal mosaic virus*; SYNV, *Sonchus yellow net virus*; RYSV, *Rice yellow stunt virus*; VSIV, *Vesicular stomatitis Indiana virus*; RABV, *Rabies virus*. R, A or G; Y, C or U; W, A or U; N<sub>m</sub>, number of variable nucleotides.



**Fig. 5.** Identification of the 5' end of LBVV ORF2 mRNA. Poly(A)<sup>+</sup> RNA from healthy (lane H) or LBVV-infected (lane V) lettuce plants was reverse-transcribed and the reaction products were analysed on 8% polyacrylamide urea gels. The same oligonucleotide primer was used to generate a sequencing ladder from the cloned gene junction between the CP and ORF2 (lanes A, U, G and C).

extension up to, but not including, the putative cap structure. These results thus indicated that ORF2 mRNA began at nt 1435 of LBVV RNA2 and that the LBVV mRNA presumably had a cap structure.

The 3'-terminal sequences of LBVV mRNAs were determined by cloning and sequencing of the 3'RACE-generated cDNA clones from poly(A)<sup>+</sup> RNA of LBVV-infected lettuce plants. To obtain the 3'-terminal sequences of these LBVV mRNAs, 6–10 clones derived from each mRNA were analysed. The lengths of their poly(A) tails differed, but all of their clones usually contained more than 100 A residues, with a maximum of 173 A residues. There were two types of 3'-terminal sequence among the LBVV mRNAs: 5'-AUANGUA<sub>n</sub>-3' (genes CP, 2, 3 and 6) and 5'-UAUUAGA<sub>n</sub>-3' (genes 4, 5 and L), which were consistent with the transcription termination/polyadenylation signal sequences 3'-UAUNCAUUUUUUU-5' and 3'-AUAUCUUUUUU-5' in the gene junction regions of the LBVV genome (Fig. 4). Furthermore, to confirm the 3' end of the CP mRNA, an RNase H digestion experiment was carried out (Fig. 6). Treatment of poly(A)<sup>+</sup> RNA from LBVV-infected lettuce plants with RNase H in the presence of the viral RNA sense oligonucleotide primer RNase dig 1<sup>n</sup> and an oligo(dT)<sub>12–18</sub> primer resulted in a shortening of the poly(A)<sup>+</sup> RNA molecules by an approximately 100–200 nt compared with the untreated control. No such change was seen in the size of poly(A)<sup>+</sup> RNA molecules treated with RNase H in the presence of the virus-complementary RNA sense oligonucleotide primer RNase dig 1<sup>p</sup> and viral RNA sense oligonucleotide primer RNase dig 2<sup>n</sup>, which are located on gene 2. These results indicated that CP mRNA possessed a



**Fig. 6.** RNase H digestion analysis of the LBVV CP mRNA. Poly(A)<sup>+</sup> RNA from LBVV-infected lettuce plants was treated with RNase H in the absence (lane 1) or presence of the oligonucleotide primers RNase dig 1<sup>n</sup> (lane 2), RNase dig 1<sup>p</sup> (lane 3), RNase dig 2<sup>n</sup> (lane 4) and oligo(dT)<sub>12–18</sub> (lane 5), respectively. The RNAs were separated on a 1.5% agarose/formamide gel, transferred to a nitrocellulose membrane and hybridized to <sup>32</sup>P-labelled cDNA inserts from the LBVV CP gene. RNA molecular size markers (kb) are indicated on the left.

virus-complementary (positive-sense) RNA polarity and terminated before the initiation site of the ORF2 mRNA, which began at nt 1435 of LBVV RNA2, and that the poly(A) sequences were terminally located at the 3' end of the CP mRNA.

### Sequences of gene junctions

The gene junctions of rhabdoviruses consist of three parts: a gene-end sequence, which is required for transcription termination/polyadenylation; an intergenic region (IGR), which is not represented in the mRNAs; and a gene-start sequence, which plays an important role in transcription initiation. According to the stop-and-start model of transcription, the viral polymerase complex transcribes a gene, polyadenylates and releases the mRNA, and re-initiates transcription of the next gene under the direction of the gene junction sequences (Banerjee *et al.*, 1991; Conzelmann, 1998; Barr *et al.*, 2002). The determination of the 5' and 3' ends of all LBVV mRNAs described above led us to delineate precisely the gene junctions in the LBVV genome (Fig. 4).

A conserved sequence for the initiation of LBVV genes, 3'-CUCU-5', was found. This conserved sequence is also found in the gene junction regions of LNYV and NCMV in the genus *Cytorhabdovirus* (Wetzel *et al.*, 1994; Tanno *et al.*, 2000). Although the transcription initiation consensus signal sequence of most rhabdoviruses, including genus *Nucleorhabdovirus*, is 3'-UUGU-5' (Heaton *et al.*, 1989; Walker *et al.*, 2000), the 3'-CUCU-5' sequence in the gene junction regions was considered to act as a transcription initiation signal in both LBVV and rhabdoviruses in the genus *Cytorhabdovirus*.

In all rhabdoviruses, gene-end sequences comprise an AU-rich region and a cytidylate and a poly(U) tract (Conzelmann, 1998). The poly(U) tract is thought to be a template for the mRNA poly(A) tail by reiterative

transcription or slippage (Barr *et al.*, 2002). The gene-end sequences of LBVV were either 3'-UAUNCAUUUUUUU-5' (type A) or 3'-AUAACUUUUUUU-5' (type B), reminiscent of those observed in rhabdoviruses (Fig. 4). Interestingly, when the transcription termination/polyadenylation signal of LBVV was 3'-UAUNCAUUUUUUU-5', the transcription of the next gene initiated immediately after the mononucleotide G (IGR), while for the type B sequence, the two genes (genes 5 and L) were located at the 5'-terminal ends of each RNA (viral RNA) or the transcription was re-initiated after a long 42 nt IGR (gene 4). These gene-end sequences of LBVV might play a role not only in the termination of transcription and polyadenylation but also in re-initiating transcription of the downstream genes. The role of gene-end sequences in signalling initiation of the downstream mRNA synthesis was recently reported for *Vesicular stomatitis Indiana virus* (VSIV) (Hinzman *et al.*, 2002).

The IGRs of LBVV, except for the gene junction between genes 4 and 5, were highly conserved, consisting of the mononucleotide G (Fig. 4). The IGR of a fish rhabdovirus in the genus *Novirhabdovirus* also consists of a single nucleotide, G or A (Walker *et al.*, 2000). The IGRs of rhabdoviruses either are highly conserved in sequence and length (genera *Vesiculovirus* and *Nucleorhabdovirus*) or are not conserved in either sequence or length (genera *Lyssavirus*, *Ephemerovirus* and *Cytorhabdovirus*), whereas the first residue of the IGRs in the most rhabdoviruses is G (Conzelmann, 1998; Walker *et al.*, 2000). The G residue also followed immediately downstream of the last genes, gene 5 on RNA2 and the L polymerase gene on RNA1. The G residue might play an important role in the termination of LBVV mRNA transcripts, as reported for VSIV in which the first G residue in the IGR is required for efficient termination of the upstream transcript (Barr *et al.*, 1997; Stillman & Whitt, 1997, 1998). The IGR between genes 4 and 5 differed from the other LBVV IGRs and consisted of 42 nt. Such a long IGR is also found in *Rabies virus* (RABV) (Tordo *et al.*, 1986), *Bovine ephemeral fever virus* (Walker *et al.*, 2000) and LNYV (Wetzel *et al.*, 1994). The long IGR of RABV was reported to correlate with transcriptional attenuation (Finke *et al.*, 2000). However, because of the varying specific activities of our probes, Northern blot analysis failed to show a different degree of transcription attenuation between genes 4 and 5 (Fig. 3).

## Conclusion

This study completes the sequence analysis of the entire LBVV genome, which is the first report of the entire sequence of a member of the genus *Varicosavirus*. The LBVV genome is composed of 12 878 nt, divided into two segments. LBVV RNA1 consists of 6797 nt and contains a small gene 6 and L polymerase gene (Sasaya *et al.*, 2002). LBVV RNA2 consists of 6081 nt and contains five genes. The genomic organization of LBVV and the amino acid sequences of LBVV CP and L polymerase were similar to those in rhabdovirus genomes. At the LBVV gene junctions,

there were transcription termination/polyadenylation and initiation signals comparable with those of rhabdoviruses, and LBVV transcribed capped and polyadenylated monocistronic RNAs. LBVV and rhabdoviruses may utilize a similar mechanism to express individual genes differentially from a contiguous viral genome. These observations provide further evidence that LBVV is most closely related to viruses in the family *Rhabdoviridae* of the order *Mononegavirales*, even though LBVV has a genome that is divided into two segments and does not have enveloped virions. Furthermore, compared with viruses in the family *Rhabdoviridae*, purified LBVV contains a relatively large amount of positive-sense RNA (Sasaya *et al.*, 2001), which is also a unique property of LBVV.

## ACKNOWLEDGEMENTS

We are very grateful to Dr Mike A. Mayo for a critical reading of the manuscript and helpful advice, and to Mrs Yoshiko Aono for technical assistance. This research was funded by a grant from the Ministry of Agricultural, Forestry and Fisheries.

## REFERENCES

- Banerjee, A. K., Barik, S. & De, B. P. (1991). Gene expression of non-segmented negative strand RNA viruses. *Pharmacol Ther* **51**, 47–70.
- Barr, J. N., Whelan, S. P. J. & Wertz, G. W. (1997). Role of the intergenic dinucleotide in vesicular stomatitis virus RNA transcription. *J Virol* **71**, 1794–1801.
- Barr, J. N., Whelan, S. P. J. & Wertz, G. W. (2002). Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Biochim Biophys Acta* **1577**, 337–353.
- Chen, X. Y., Luo, Z. L. & Fang, R. X. (1998). Sequence analysis of the RYSV gene 3. *Chin Sci Bull* **43**, 745–748 (in Chinese).
- Conzelmann, K.-K. (1998). Nonsegmented negative-strand RNA viruses: genetics and manipulation of viral genomes. *Annu Rev Genet* **32**, 123–162.
- Duggal, R., Lahser, F. C. & Hall, T. C. (1994). *cis*-Acting sequences in the replication of plant viruses with plus-sense RNA genomes. *Annu Rev Phytopathol* **32**, 287–309.
- Fazeli, C. F. & Rezaian, M. A. (2000). Nucleotide sequence and organization of ten open reading frames in the genome of Grapevine leafroll-associated virus 1 and identification of three subgenomic RNAs. *J Gen Virol* **81**, 605–615.
- Finke, S., Cox, J. H. & Conzelmann, K.-K. (2000). Differential transcription attenuation of rabies virus genes by intergenic regions: generation of recombinant viruses overexpressing the polymerase gene. *J Virol* **74**, 7261–7269.
- Gupta, K. C. & Kingsbury, D. W. (1984). Complete sequences of the intergenic and mRNA start signals in the Sendai virus genome: homologues with the genome of vesicular stomatitis virus. *Nucleic Acids Res* **12**, 3829–3841.
- Hamilton, W. D. O., Boccara, M., Robinson, D. J. & Baulcombe, D. C. (1987). The complete nucleotide sequence of tobacco rattle virus RNA-1. *J Gen Virol* **68**, 2563–2575.
- Heaton, L. A., Zuidema, D. & Jackson, A. O. (1987). Structure of the M2 protein gene of sonchus yellow net virus. *Virology* **161**, 234–241.
- Heaton, L. A., Hillman, B. I., Hunter, B. G., Zuidema, D. & Jackson, A. O. (1989). Physical map of the genome of sonchus yellow net

- virus, a plant rhabdovirus with six genes and conserved gene junction sequences. *Proc Natl Acad Sci U S A* **86**, 8665–8668.
- Hillman, B. I., Heaton, L. A., Hunter, B. G., Modrell, B. & Jackson, A. O. (1990). Structure of the gene encoding the M1 protein of sonchus yellow net virus. *Virology* **179**, 201–207.
- Hinzman, E. E., Barr, J. N. & Wertz, G. W. (2002). Identification of an upstream sequence element required for vesicular stomatitis virus mRNA transcription. *J Virol* **76**, 7632–7641.
- Huijberts, N., Blystad, D.-R. & Bos, L. (1990). Lettuce big-vein virus: mechanical transmission and relationships to tobacco stunt virus. *Ann Appl Biol* **116**, 463–475.
- Kusaba, S., Fukumoto, M., Honda, C., Yamaguchi, I., Sakamoto, T. & Kano-Murakami, Y. (1998). Decreased GA<sub>1</sub> content caused by the overexpression of *OSH1* is accompanied by suppression of GA 20-oxidase gene expression. *Plant Physiol* **117**, 1179–1184.
- Kuwata, S., Kubo, S., Yamashita, S. & Doi, Y. (1983). Rod-shaped particles, a probable entity of lettuce big vein virus. *Ann Phytopathol Soc Jpn* **49**, 246–251.
- Liu, X. & Gorovsky, M. A. (1993). Mapping the 5' and 3' ends of *Tetrahymena thermophila* mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). *Nucleic Acids Res* **21**, 4954–4960.
- Lot, H., Campbell, R. N., Souche, S., Milne, R. G. & Roggero, P. (2002). Transmission by *Olpidium brassicae* of *Mirafiori lettuce virus* and *Lettuce big-vein virus*, and their roles in lettuce big-vein etiology. *Phytopathology* **92**, 288–293.
- Luo, Z. L. & Fang, R. X. (1998). Structure analysis of the rice yellow stunt rhabdovirus glycoprotein gene and its mRNA. *Arch Virol* **143**, 2453–2459.
- Luo, Z. L., Chen, X. Y., Gao, D. M. & Fang, R. X. (1998). The gene 4 of rice yellow stunt rhabdovirus encodes the matrix protein. *Virus Genes* **16**, 277–280.
- Masters, P. S. & Samuel, C. E. (1984). Detection of *in vivo* synthesis of polycistronic mRNAs of vesicular stomatitis virus. *Virology* **134**, 277–286.
- Mayo, M. A. (2000). Genus *Varicosavirus*. In *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 521–523. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.
- Natsuaki, K. T., Morikawa, T., Natsuaki, T. & Okuda, S. (2002). Mirafiori lettuce virus detected from lettuce with big vein symptoms in Japan. *Jpn J Phytopathol* **68**, 309–312 (in Japanese).
- Roggero, P., Ciuffo, M., Vaira, A. M., Accotto, G. P., Masenga, V. & Milne, R. G. (2000). An *Ophiovirus* isolated from lettuce with big-vein symptoms. *Arch Virol* **145**, 2629–2642.
- Rose, J. K. & Whitt, M. A. (2001). *Rhabdoviridae*: the viruses and their replication. In *Fields Virology*, 4th edn, pp.1221–1244. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins.
- Sambrook, J. & Russell, D. W. (2000). Northern hybridization. In *Molecular Cloning: a Laboratory Manual*, 3rd edn, pp. 7.42–7.50. Edited by J. Sambrook & D. W. Russell. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sasaya, T., Ishikawa, K. & Koganezawa, H. (2001). Nucleotide sequence of the coat protein gene of *Lettuce big-vein virus*. *J Gen Virol* **82**, 1509–1515.
- Sasaya, T., Ishikawa, K. & Koganezawa, H. (2002). The nucleotide sequence of RNA1 of *Lettuce big-vein virus*, Genus *Varicosavirus*, reveals its relation to nonsegmented negative-strand RNA viruses. *Virology* **297**, 289–297.
- Scholthof, K.-B. G., Hillman, B. I., Modrell, B., Heaton, L. A. & Jackson, A. O. (1994). Characterization and detection of sc4: a sixth gene encoded by sonchus yellow net virus. *Virology* **204**, 279–288.
- Stillman, E. A. & Whitt, M. A. (1997). Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts. *J Virol* **71**, 2127–2137.
- Stillman, E. A. & Whitt, M. A. (1998). The length and sequence composition of vesicular stomatitis virus intergenic regions affect mRNA levels and the site of transcript initiation. *J Virol* **72**, 5565–5572.
- Tanno, F., Nakatsu, A., Toriyama, S. & Kojima, M. (2000). Complete nucleotide sequence of Northern cereal mosaic virus and its genome organization. *Arch Virol* **145**, 1373–1384.
- Tordo, N., Poch, O., Ermine, A., Keith, G. & Rougeon, F. (1986). Walking along the rabies genome: is the large G–L intergenic region a remnant gene? *Proc Natl Acad Sci U S A* **83**, 3914–3918.
- van Rossum, C. M. A., Brederode, F. T., Neeleman, L. & Bol, J. F. (1997). Functional equivalence of common and unique sequences in the 3' untranslated regions of alfalfa mosaic virus RNAs 1, 2, and 3. *J Virol* **71**, 3811–3816.
- Vetten, H. J., Lesemann, D.-E. & Dalchow, J. (1987). Electron microscopical and serological detection of virus-like particles associated with lettuce big vein disease. *J Phytopathol (Berl)* **120**, 53–59.
- Walker, P. J., Benmansour, A., Dietzgen, R. & 7 other authors (2000). Family *Rhabdoviridae*. In *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 563–583. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.
- Wetzel, T., Dietzgen, R. G. & Dale, J. L. (1994). Genomic organization of lettuce necrotic yellows rhabdovirus. *Virology* **200**, 401–412.
- Zuidema, D., Heaton, L. A. & Jackson, A. O. (1987). Structure of the nucleocapsid protein gene of sonchus yellow net virus. *Virology* **159**, 373–380.