Journal of General Virology (2000), **81**, 1865–1869. Printed in Great Britain

Sugarcane yellow leaf virus: a novel member of the *Luteoviridae* that probably arose by inter-species recombination

Grant R. Smith,[1,2] Zara Borg,[1,2] Ben E. L. Lockhart,[3] Kathryn S. Braithwaite[1,2] and Mark J. Gibbs[4]

[1] David North Plant Research Centre, Bureau of Sugar Experiment Stations, PO Box 86, Q 4068 Indooroopilly, Australia
[2] Cooperative Research Centre for Tropical Plant Pathology, The University of Queensland, Q 4070 St Lucia, Australia
[3] Department of Plant Pathology, University of Minnesota, MN 55108 St Paul, USA
[4] Research School of Biological Sciences, Australian National University, GPO Box 475, ACT 2601 Canberra, Australia

The 5895 nucleotide long single-stranded RNA genome of Sugarcane yellow leaf virus Florida isolate (SCYLV-F) includes six major ORFs. All but the first of these are homologous to genes of known function encoded by viruses of the three newly defined genera in the *Luteoviridae* ('luteovirids'), i.e. poleroviruses, luteoviruses and the enamoviruses. SCYLV-F ORFs 1 and 2 are most closely related to their polerovirus counterparts, whereas SCYLV-F ORFs 3 and 4 are most closely related to counterparts in luteovirus genomes, and SCYLV-F ORF5 is most closely related to the read-through protein gene of the only known enamovirus. These differences in affinity result from inter-species recombination. Two recombination sites in the genome of SCYLV-F map to the same genomic locations as previously described recombinations involving other luteovirids. A fourth type of luteovirid, *Soybean dwarf virus*, has already been described. Our analyses indicate that SCYLV-F represents a distinct fifth type.

Introduction

Yellow leaf syndrome of sugarcane (*Saccharum* L. interspecific hybrids) has been reported from Hawaii (Schenck, 1990) Brazil, continental USA (Comstock *et al.*, 1994), Australia (Smith *et al.*, 1995), Mauritius (Anon., 1995) and South Africa (Bailey *et al.*, 1996). Small icosahedral particles were found in symptomatic plants (Anon., 1995), and the aphid *Melanaphis sacchari* was shown to transmit a virus, now commonly called Sugarcane yellow leaf virus (SCYLV), from infected plants (Scagliusi & Lockhart, 1997, 2000). A weak serological reaction, and the physico-chemical properties of the particles, suggested that SCYLV was a member of the *Luteoviridae* ('luteovirid') (Scagliusi & Lockhart, 1997; Vega *et al.*, 1997), and this was confirmed by Irey *et al.* (1997) when a short stretch of sequence from the SCYLV genome was found to be most similar to sequence from *Barley yellow dwarf virus*-PAV (BYDV-PAV; *Luteovirus*). Here, we present an analysis of the genome of the Florida isolate of Sugarcane yellow leaf virus (SCYLV-F) and show that only part of its genome is related to that of BYDV-PAV. Other parts of the genome have closer affinities to regions in the genomes of other luteovirids. The results of our analyses suggest that the unusual affinities of SCYLV-F result from previously unrecognized inter-species recombinational events.

Author for correspondence: Grant Smith (at David North Plant Research Centre). Fax +61 7 3871 0383.
e-mail GSmith@bses.org.au

The nucleotide sequence of the SCYLV-F genome has been deposited in EMBL, accession no. AJ249447.

An isolate of SCYLV was obtained from sugarcane cultivar CP65-357 from Florida, which had been maintained in glasshouses on the St Paul Campus, University of Minnesota. Viral particles were purified as described by Scagliusi & Lockhart (2000) and RNA was extracted by treatment with Proteinase K and SDS. First strand cDNA was synthesized using MMLV-Superscript II (Gibco-BRL), second strand with RNaseH, *E. coli* DNA polymerase and *E. coli* DNA ligase. The products were fractionated on a CL-4B spun column (Pharmacia), blunt-end ligated into pBluescript SK(−) (Stratagene) and sequenced. Rapid amplification of cDNA ends (RACE) was performed using a 5′/3′ RACE kit (Boehringer) and the DNA fragments were cloned into plasmid pCR2.1 (Invitrogen), and sequenced. The entire 5895 nucleotide sequence of the SCYLV-F genome was determined in both orientations from two or more independently generated clones (EMBL accession no. AJ249447).

There were no detectable similarities between the 5′ and 3′ untranslated regions (UTRs) and any sequences in the databases. The 5′ UTR starts with the sequence ACAAAA, which is consistent with the 5′-terminal motif of many

0001-6913 © 2000 SGM

Table 1. Major ORFs encoded on the positive strand of the SCYLV genome and their homologues

Note that Expect values (E values) of less than 1×10^{-2} usually indicate homology and those less than 1×10^{-6} almost always indicate homology (Altschul *et al.*, 1997; Bork & Gibson, 1996).

ORF	Position	Predicted mass (kDa)	Closest homologue	E value
0	67–837	30·2	None found	
1	242–2194	72·5	ORF1 PLRV	1×10^{-70}
2*	1735–3453	64·4	ORF2 PLRV	0
1/2†	242–3453	120·6	Not tested	
3	3648–4238	21·8	ORF3 BYDV-PAV	2×10^{-34}
4	3679–4131	16·6	ORF4 BYDV-PAV	1×10^{-8}
5	4239–5678	52·1	ORF5 PEMV-1	3×10^{-41}
3/5‡	3648–5678	73·4	Not tested	

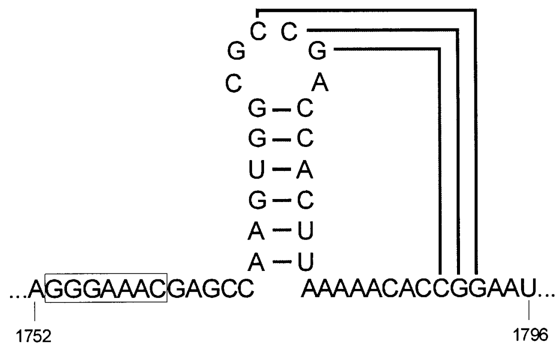
* Stop codon to stop codon.
† Frame-shift fusion protein.
‡ Read-through fusion protein.

poleroviruses. The 3′ UTR is 217 nucleotides long, which is slightly longer than most polerovirus 3′ UTRs but considerably shorter than the 600+ nucleotides of the 3′ UTRs of luteoviruses.

The SCYLV-F genome is arranged like those of the poleroviruses. SCYLV-F ORF0 begins at the first AUG codon in the sequence and encodes a 30·2 kDa protein (Table 1). No significant similarities were detected when the databases were searched with the SCYLV-F ORF0 amino acid sequence using the programs BLASTX and BLASTP (Altschul *et al.*, 1997). It has been noted that the equivalent genes in polerovirus genomes are similarly poorly conserved (Miller *et al.*, 1995). SCYLV-F ORF1 encodes a 72·5 kDa protein which was found to be most similar to the equivalent genes in the genomes of the poleroviruses (Table 1) and *Pea enation mosaic virus*-1 (PEMV-1; *Enamovirus*). ORF1 of *Potato leaf roll virus* (PLRV; *Polerovirus*) has been shown to be a serine protease (Hulanicka *et al.*, 1999) and a protease-motif, **H**X$_{29-34}$[**D**/**E**]X$_{62-63}$**T**XK-**G**Y**SG** (Gorbalenya *et al.*, 1989), is conserved in all polerovirus ORF1 amino acid sequences and the equivalent sequence carried by PEMV-1. SCYLV-F also encodes the protease-motif and hence it is likely that SCYLV-F ORF1 encodes a protease. SCYLV-F ORF2 was found to be most similar to the RNA-dependent RNA polymerase (RdRp) genes of the poleroviruses (Table 1), sobemoviruses, barnaviruses and PEMV-1. The 5′ terminus of the SCYLV-F RdRp overlaps ORF1 in the −1 reading frame by 460 nucleotides. There is a likely frame-shifting slippery heptamer, 5′ GGGAAAC 3′, in this overlap at position 1753, and sequence on the 3′ side of the heptamer motif could form a pseudoknot (Fig. 1). This arrangement with a slippery heptamer and a pseudoknot in the overlap of ORFs 1 and 2 is conserved in the genomes of the poleroviruses and PEMV-1, and permits −1 ribosomal frameshifting from ORF1 to the RdRp gene (Prüfer *et al.*, 1992). The heptamer and the pseudoknot probably have the same role in the SCYLV-F genome and thus ORFs 1 and 2 of SCYLV-F may be translated together by frameshifting to produce a single 120·6 kDa fusion protein (Table 1). The pseudoknot predicted in the SCYLV-F sequence has a similar structure to those of the polerovirus and PEMV-1 genomes. However, the sequence that forms the pseudoknot in the SCYLV-F genome includes only 12 of the 19 conserved nucleotides identified by Miller *et al.* (1995) in polerovirus and PEMV-1 sequences (Fig. 1). Furthermore, the likely 3′ stem in the SCYLV-F pseudoknot is the shortest so far predicted for a luteovirid, consisting of only 3 base pairs. Together, these differences suggest that the selective forces acting on the SCYLV-F frameshifting site are probably somewhat different from those acting on the equivalent sites in polerovirus genomes.

SCYLV-F ORF3 was found to be most similar to the coat protein (CP; 22 kDa) gene of BYDV-PAV (Table 1). Likewise, SCYLV-F ORF4 was most similar to the movement protein (MP) gene (Miller & Rasochová, 1997) of BYDV-PAV. As found in the genomes of other luteovirids, the likely MP gene (ORF4) of SCYLV-F is completely contained within the sequence encoding ORF3, in the +1 reading frame. The N terminus of the likely CP (ORF3) of SCYLV-F is arginine rich and the SCYLV-F CP gene is terminated with an amber stop codon, like all other luteovirid CP genes, and this codon is followed by the first codon of the read-through (RT) protein gene (ORF5). The sequence surrounding the amber stop codon is highly conserved (Miller *et al.*, 1995). The SCYLV-F sequence differs from the consensus sequence at the +4 and +5 base positions, and this changes the conserved aspartic acid residue in the RT protein to either a serine or a glycine residue. The proline-rich domain at the N terminus of the RT protein is encoded by a repeated nucleotide pattern of the form [CCNNNN]$_{7-16}$ in the genomes of all previously described luteovirids (Miller *et al.*, 1995). However, the equivalent SCYLV-F sequence follows a different pattern described by the motif CCC[CCNNNN]$_5$GCU[CCNNNN]$_3$. The SCYLV-F RT protein gene (ORF5) encodes a 52 kDa protein which is most similar, over the first 270 amino acid residues, to the RT protein sequence of PEMV-1 (Table 1). No significant similarities were found between the 3′ region of this ORF and any protein in the databases.

To test the significance of the contrasting database search results (Table 1), multiple alignments of the RdRp, CP and RT protein gene sequences were constructed and analysed (Fig. 2). The program SiScan (M. J. Gibbs, J. S. Armstrong & A. J. Gibbs, Australian National University; http://life.anu.edu.au/molecular/software/siscan/) was used to calculate total pair-wise identity within a window that was passed over a nucleotide alignment of SCYLV-F, BYDV-PAV, PEMV-1 and PLRV sequences. SiScan was also used to calculate Z scores from the identity scores using 100 equivalent sets of




```
              *          *         *        *   *           **
SCYLV      ...ACCAGGGAAACGAGCCAAG-UGGCGCCGACCACUUAAAAACACCGGAAU...
BYDV-RPV   ...AGUCGGGAAACGGGAAGG--CGGCGGCGUCCGCCGU--AACAAACGCAG...
BWYV       ...UGUCGGGAAACGGAGUGCG-CGGCACCGUCCGCGG---AACAAACGGAG...
PLRV-A, C  ...AGCCUUUAAAUGGGCAAG--CGGCACCGUCCGCCAA--AACAAACGGCA...
PEMV-1     ...GCUCGGGAAACGGAUUAUUCCGGUCGACUCCGGAGA--AACAAAGUCAA...
BYDV-PAV   ...CUGUGGGUUUUUAGAGGGGCUCUGUACCGCCUCUGGUUUUGAGAGCCCA...
```

Fig. 1. Potential pseudoknot structure in the SCYLV genome at the beginning of the region where ORFs 1 and 2 overlap and a comparison with polerovirus and PEMV-1 sequences. The likely frameshifting heptamer is boxed in the predicted structure. In the alignment, the shifty heptamer sequence is in bold type and bases identified by Miller *et al.* (1995) as conserved between all luteovirids with sobemo-like RdRps are underlined. Conserved bases in the SCYLV sequence are also underlined and differences are highlighted with an asterisk above the base. The heptamer sequence of BYDV-PAV is also aligned for comparison.

randomized sequences. The raw scores (Fig. 2A) and Z scores (Fig. 2B) confirmed that recombination had taken place and suggested that at least two inter-species recombinations contributed to the evolution of SCYLV-F. Direct inspection of the alignments showed that SCYLV-F ORF2 is most similar to the RdRp gene of PLRV over its entire length and database searches using the program BLASTP (Altschul *et al.*, 1997) showed that the arginine-rich N-terminal domain of the SCYLV-F CP is also most similar to its counterpart in the PLRV CP. The affinities of the SCYLV-F CP change at about amino acid residue 40. Most of the remainder of the CP sequence is closest to that of BYDV-PAV. A closer affinity to the PEMV-1 sequence becomes apparent in the last 15 amino acid residues of the CP or in the short proline-rich stretch at the N terminus of the RT protein. Alignments also suggested that the internal UTR sequence of SCYLV-F is more closely related to that of PLRV than to that of BYDV-PAV and we identified two short stretches of identity in the alignment of the SCYLV-F and PLRV sequences (5′ UAGCGGG; 5′ CGCAAUCCC).

Some previously identified recombinations in luteovirid genomes may have involved strand-switching at the subgenomic promoter that lies within the internal UTR (Miller *et al.*, 1995). A recombinational event close to this location in an ancestor of SCYLV-F might also have involved this mechanism. However, there is an alternative explanation. Base-pairing between positive- and negative-strand RNAs from different luteovirid species within conserved repetitive sequences may have induced recombination (Gibbs & Cooper, 1995). There are two such repetitive sequences close to the likely recombination sites in the SCYLV-F genome: the sequence that encodes the arginine-rich N terminus of the CP includes a pattern of repeated cytosine and guanine residues, and there is a repetitive cytosine-rich sequence that encodes the proline-rich N-terminal end of the RT protein. The proximity of these sequences to the likely recombination sites supports this alternative explanation.

Minimum evolution and maximum parsimony trees were found for a subset of luteovirus sequences (Fig. 2D–F) by heuristic searching with the program PAUP version 4d64 (kindly supplied by David L. Swofford). Minimum evolution trees were calculated using the Jukes–Cantor correction. Bootstrap values were calculated from neighbour-joining trees that were found using 1000 bootstrapped samples of the data. As the trees illustrate (Fig. 2D, E), no other luteovirid so far characterized has a polerovirus-like RdRp gene and a CP gene most closely related to those of viruses of the newly defined *Luteovirus* genus. Thus, it is clear that our new evidence of recombination at a site between the RdRp and CP genes cannot be explained by any of the previously identified recombinations that map to this region in luteovirid genomes (Veidt *et al.*, 1988; Rathjen *et al.*, 1994; Guilley *et al.*, 1995). Similarly, the evidence of recombination at a site close to the boundary between the CP and RT protein genes cannot be explained by the recombinational event previously identified close to this site in the evolution of *Cucurbit aphid-borne yellows virus* (CABYV; Gibbs & Cooper, 1995). Phylogenetic analyses show that the RdRp and CP genes of SCYLV-F and CABYV have different affinities (Fig. 2D, E), and hence the ancestors of these viruses must have diverged before the respective recombinational events and those events were independent.

The fact that the inter-species recombinations evident within the SCYLV-F sequence map to the same genomic locations as other previously described recombinations suggests that selection against recombination occurs at other

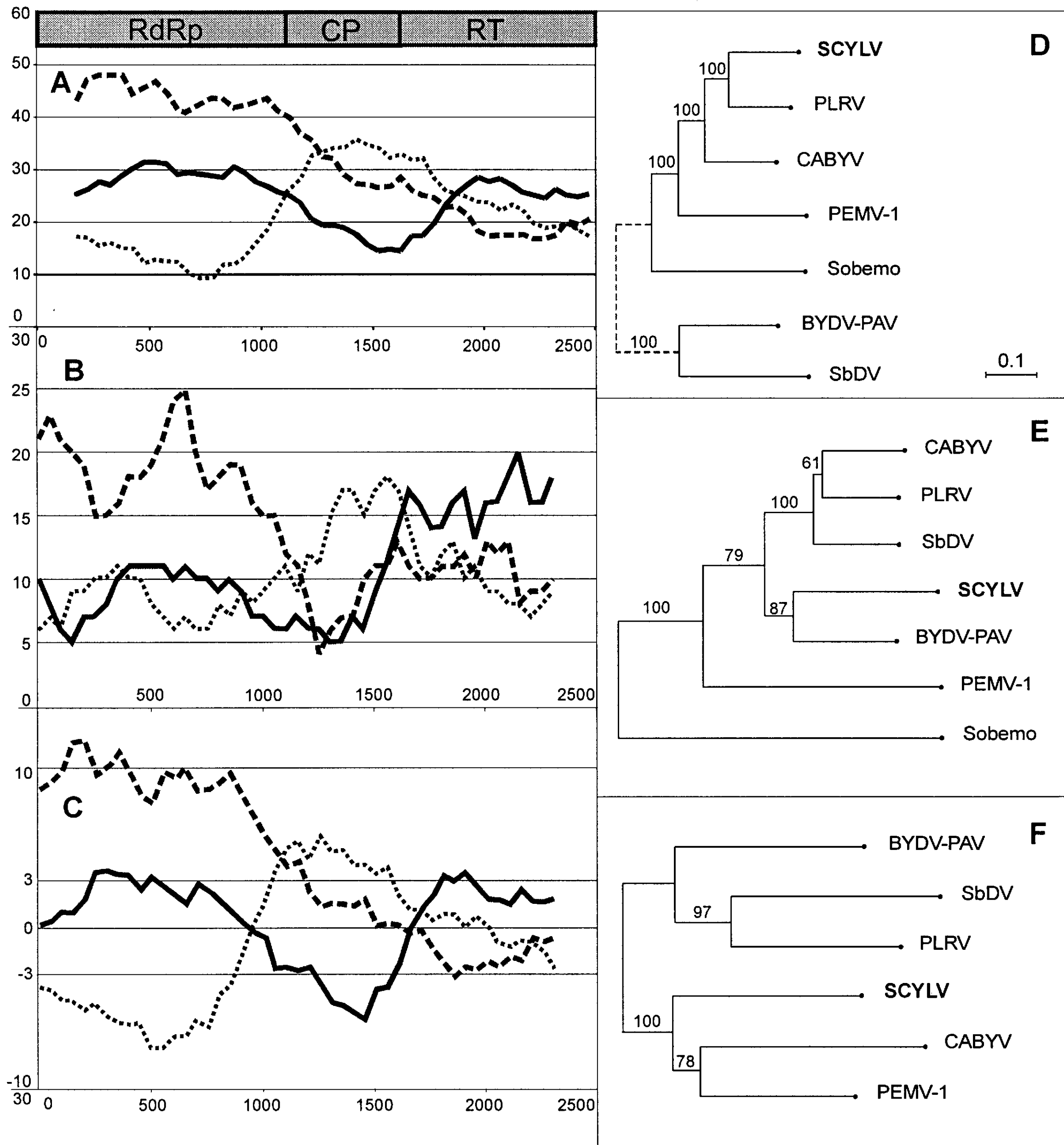


Fig. 2. Phylogenetic analyses of aligned nucleotide sequences of the RdRp, CP and RT genes. Percentage nucleotide identity was calculated using SiScan for pairs of sequences within a window covering 350 nucleotide positions (A). The window was moved through the alignment with a step length of 50 nucleotides. Values are shown for three sets of comparisons: SCYLV and PLRV (dashed lines), SCYLV and BYDV-PAV (dotted lines), and SCYLV and PEMV-1 (unbroken lines). Positions (columns) including gaps were deleted from the alignment before the analysis, and hence nucleotide positions shown on the *x*-axis refer to the condensed alignment. The number of informative sites counted within the window for the same three pairs of relationships was also calculated (B) as were Z scores calculated for total nucleotide identity scores (C). The internal UTR was not included in the alignment analysed with SiScan because it is poorly conserved. Panels (D)–(F) show minimum evolution trees found for the RdRp, CP and RT protein genes respectively. Maximum parsimony trees found using the same datasets were identical to the trees shown. Sequences of two sobemoviruses, *Southern bean mosaic virus* cowpea strain and *Rice yellow mottle virus*, were used as outliers to locate the roots of the RdRp and CP trees. The sobemovirus sequences are represented by the terminal node labelled 'Sobemo'. The scale bar relates the branch lengths to the estimated number of substitutions per site. The branch linking the RdRp sequence lineages of BYDV-PAV and SbDV to the rest of the tree is represented by a dashed line in (D) because the RdRp sequences of these viruses may not be homologous to the RdRp sequences of the other viruses.

sites, or operates in favour of certain new combinations of genes. One possibility is that the recombination events have been linked to changes in the combination of hosts and vectors used by the luteovirids (Gibbs, 1994) as determined by new combinations of RdRp, MP and RT protein genes. Alternatively, there may be strong selection against recombination at other sites.

Clearly inter-species recombination has occurred very frequently in the evolution of luteovirids. For this reason, it is difficult to distinguish the recombinant and parental lineages. We cannot assume that SCYLV-F is a recombinant simply because its sequence provides new evidence of recombination. The evidence depends on the availability of sequences from several species and the order in which the sequences were obtained could influence its interpretation. However, given that we have evidence of not one, but two, recombinational events that appear to have connected sequences from a polerovirus, a luteovirus and an enamovirus, which are otherwise largely distinct lineages, we think it is likely that SCYLV-F is a recombinant.

We thank Dr Neil Olszewski, Department of Plant Biology, University of Minnesota and his group, and the staff and students of the Department of Plant Pathology, University of Minnesota for assistance during the cloning of the viral genome. We also thank the Cooperative Research Centre for Tropical Plant Pathology, Sugar Research and Development Corporation and the Bureau of Sugar Experiment Stations for financial and logistical assistance during this study.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Anon. (1995). Yellow leaf syndrome. *Mauritius Sugar Industry Research Institute Annual Report* 1995, p. 51.

Bailey, R. A., Bechet, G. R. & Cronjé, C. P. R. (1996). Notes on the occurrence of yellow leaf syndrome of sugarcane in southern Africa. *Proceedings of the South African Sugar Technologists Association* **70**, 3–6.

Bork, P. & Gibson, T. J. (1996). Applying motif and profile searches. *Methods in Enzymology* **266**, 162–183.

Comstock, J. C., Irvine, J. E. & Miller, J. D. (1994). Yellow leaf syndrome appears on the United States mainland. *Sugar Journal*, March 1994, 33–35.

Gibbs, M. J. (1994). Risks in using transgenic plants. *Science* **264**, 1650–1651.

Gibbs, M. J. & Cooper, J. I. (1995). A recombinational event in the history of luteoviruses probably induced by base-pairing between the genomes of two distinct viruses. *Virology* **206**, 1129–1132.

Gorbalenya, A. E., Koonin, E. V., Blinov, V. M. & Donchenko, A. P. (1989). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Letters* **236**, 287–290.

Guilley, H., Richards, K. E. & Jonard, G. (1995). Nucleotide sequence of beet mild yellowing virus RNA. *Archives of Virology* **140**, 1109–1118.

Hulanicka, M. D., Sadowy, E., Juszczuk, M. & Gronenborn, B. (1999). ORF1 of potato leafroll virus encodes a serine proteinase indispensable for viral replication. Abstract VW15.04, *XIth International Congress of Virology*, p. 48.

Irey, M. S., Baucum, L. E., Derrick, K. S., Manjunath, K. L. & Lockhart, B. E. (1997). Detection of the luteovirus associated with yellow leaf syndrome of sugarcane (YLS) by a reverse transcriptase polymerase chain reaction and incidence of YLS in commercial varieties in Florida. *Proceedings of the ISSCT 5th Pathology and 2nd Molecular Biology Workshop*.

Miller, W. A. & Rasochová, L. (1997). Barley yellow dwarf viruses. *Annual Review of Phytopathology* **35**, 167–190.

Miller, W. A., Dinesh-Kumar, S. P. & Paul, C. P. (1995). Luteovirus gene expression. *Critical Reviews in Plant Science* **14**, 179–211.

Prüfer, D., Tacke, E., Schmitz, J., Kull, B., Kaufmann, A. & Rohde, W. (1992). Ribosomal frameshifting in plants: a novel signal directs the −1 frameshift in the synthesis of the putative viral replicase of potato leafroll luteovirus. *EMBO Journal* **11**, 1111–1117.

Rathjen, J. P., Karageorgos, L. E., Habili, N., Waterhouse, P. M. & Symons, R. H. (1994). Soybean dwarf luteovirus contains the third variant genome type in the luteovirus group. *Virology* **198**, 671–679.

Scagliusi, S. M. & Lockhart, B. E. (1997). Transmission, characterization and serology of sugarcane yellow leaf luteovirus. *Proceedings of the ISSCT 5th Pathology and 2nd Molecular Biology Workshop*.

Scagliusi, S. M. & Lockhart, B. E. (2000). Transmission, characterization and serology of a luteovirus associated with yellow leaf syndrome of sugarcane. *Phytopathology* **90**, 120–124.

Schenck, S. (1990). Yellow leaf syndrome – a new sugarcane disease. *Experiment Station Hawaiian Sugar Planters' Association Annual Report* 1990, p. 38.

Smith, G. R., Fraser, T. A., Braithwaite, K. S. & Harding, R. M. (1995). RT–PCR amplification of RNA from sugarcane with yellow leaf syndrome using luteovirus group-specific primers. *Proceedings of the 10th Australasian Plant Pathology Society Conference*, p. 84.

Vega, J., Scagliusi, S. M. M. & Ulian, E. C. (1997). Sugarcane yellow leaf disease in Brazil: evidence of association with a luteovirus. *Plant Disease* **81**, 21–26.

Veidt, I., Lot, H., Leiser, D., Scheidecker, D., Guilley, H., Richards, K. & Jonard, G. (1988). Nucleotide sequence of beet western yellows virus RNA. *Nucleic Acids Research* **16**, 9917–9932.

Received 10 January 2000; Accepted 3 March 2000