

Sequence of Subterranean Clover Stunt Virus DNA: Affinities with the Geminiviruses

P. BOEVINK, P. W. G. CHU, and P. KEESE*¹

C.S.I.R.O. Division of Plant Industry, G.P.O. Box 1600, Canberra, A.C.T. 2601; and *The Cooperative Research Centre for Plant Sciences, P.O. Box 475, Canberra, A.C.T. 2601, Australia

Received September 13, 1994; accepted December 16, 1994

The nucleotide sequences of seven circular, single-stranded DNA components of one isolate of subterranean clover stunt virus (SCSV) have been determined. Each component, of about 1 kb, appears to encode a single open reading frame in the same sense as the encapsidated DNA. Notably, the proteins encoded by two SCSV components are related. Each has a consensus nucleotide binding motif and shares about 40% amino acid identity with the other and with the putative replication proteins of banana bunchy top virus and coconut foliar decay virus. The noncoding regions of the five other SCSV components share a highly conserved noncoding sequence of about 160 nucleotides. All seven components were found to contain a sequence capable of forming a stable stem-loop structure in the noncoding region which contains a conserved 9-nucleotide sequence in the loop, very similar to that of the geminiviruses. In addition, the putative replication proteins of SCSV are similar to those of the geminiviruses. We suggest that the SCSV-like viruses and the geminiviruses share a common ancestor and that this ancestor was more like SCSV in particle structure and genome organisation. © 1995 Academic Press, Inc.

INTRODUCTION

Subterranean clover stunt is an obligately aphid-transmitted viral disease of *Trifolium subterraneum* L. (subterranean clover) and several other pasture and grain legumes in Australia (Grylls and Butler, 1956; Smith, 1966). It is transmitted in a persistent manner but does not appear to replicate in the aphid vector (Grylls and Butler, 1959). It is found throughout Australia and can cause yield losses of up to 65% (Grylls and Butler, 1956). The symptoms of infection include stunting, leaf distortion, yellowing of younger leaves, and reddening of older leaves (Grylls and Butler, 1959). Virus-like particles associated with the disease were found to be isometric, 17–19 nm in diameter and constructed from coat protein subunits of 19 K (Chu and Helms, 1988). Each particle appears to encapsidate one circular, single-stranded DNA molecule of about 1000 nucleotide residues and a short complementary primer sequence (Chu and Helms, 1988). The genome of the virus, subterranean clover stunt virus (SCSV), is composed of several distinct circular molecules (Chu and Helms, 1988).

SCSV was suggested to be a representative of a new group of plant DNA viruses (Chu and Helms, 1988) that may include banana bunchy top virus (BBTV; Harding *et al.*, 1991) and coconut foliar decay virus (CFDV; Rohde *et al.*, 1990). BBTV is transmitted by aphids and CFDV is transmitted by a leaf hopper. BBTV and CFDV have circular single-stranded DNA genome components and small isometric particles of 17–20 nm. Six DNAs have been

identified in BBTV (Burns *et al.*, 1993). The sequences of one DNA component each of BBTV and CFDV have been reported (Harding *et al.*, 1993; Rohde *et al.*, 1990). The only other group of plant viruses with circular single-stranded genomes are the geminiviruses (Francki *et al.*, 1991). However, the DNA components of geminiviruses are larger (about 2.7 kb) and they have a unique geminate particle structure (for reviews, see Davies and Stanley, 1989; Lazarowitz, 1992).

We report here the sequence of seven DNA components derived from one isolate of SCSV. The genome organisation and sequence, although distinctive, show distant similarity to geminiviruses, indicating common ancestry. SCSV may provide clues to the evolution of both viral groups.

MATERIALS AND METHODS

Cloning of SCSV DNA

The recombinant DNA techniques used were generally as described in Sambrook *et al.* (1989). The F isolate of SCSV (Chu *et al.*, 1993), previously isolated from a field-infected subterranean clover plant (Chu and Helms, 1988), was used for the sequencing. Full-length clones of the SCSV genome components were created from the replicative form (RF) DNA of the F isolate as described in Chu *et al.* (1993). Additional clones required to complete the sequencing were created using the polymerase chain reaction (PCR) with SCSV-specific primers complementary to known sequences flanking the gaps and RF DNA as the template. The PCR mixture contained 2 mM MgCl₂, 50 mM Tris-HCl, 50 mM KCl, 10 mM DTT, 0.5 mM dNTPs, and 0.2 units of Cetus Amplitaq. After addition of

The nucleotide sequences reported in this paper have been deposited with GenBank under Accession Nos. U16730 to U16736.

¹ To whom reprint requests should be addressed.

the DNA and primers the mixture was subjected to 30 cycles of denaturation for 45 sec at 95°, annealing for 45 sec at 50°, and extension for 2 min at 72°. The PCR products were either digested with restriction endonucleases (New England Biolabs) or end-filled with T4 DNA polymerase (New England Biolabs) and ligated into the Promega pGEM3Z and 5Z plasmids at appropriate sites. Recombinant plasmids were electroporated into *Escherichia coli* NM522 using the Bio-Rad Gene Pulser. Clones were analysed by restriction endonuclease digestion and agarose gel electrophoresis using *EcoRI*-restricted SPP-1 bacteriophage DNA (Bresatec) as size markers. The identity and orientation of the M13 clones were determined by hybridisation (Sambrook *et al.*, 1989) to probes derived from purified virion DNA. Probes for the virion strand were prepared by synthesising cDNA from the endogenous primer using the Klenow fragment of DNA polymerase I (New England Biolabs) and [α -³²P]dCTP. Probes for the complementary strand were prepared by end-labelling linear breakdown products in the preparations of viral DNA with T4 polynucleotide kinase (New England Biolabs) and [γ -³²P]dATP.

Clones representing component 7 were obtained by PCR using back-to-back primers. Adjacent primers matching SCSV DNAs give rise to 1-kb PCR products since the DNAs are circular. The primers were based on the sequence of one of a few random small clones from the original field isolate of SCSV (unpublished data), from which the F isolate was derived. The sequence of this clone did not correspond to any of the six components which had been found at that stage.

Sequencing of SCSV DNA

The dideoxynucleotide chain-termination method (Sanger *et al.*, 1977) was used to sequence either M13 single-stranded DNA templates (Sambrook *et al.*, 1989) or dsDNA templates prepared by the CTAB method (Del Sal *et al.*, 1989) from the clones described in the previous section. The U.S. Biochemicals Sequenase II DNA sequencing kit was used. Sequence analysis was carried out using the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package (Devereaux *et al.*, 1984).

Colony hybridisation

A SCSV library in *E. coli* NM522, generated by cloning SCSV RF DNA digested with *Eco47III* into pGEM5Z, was spotted onto a plate covered with a nitrocellulose filter and a replica plate. The plates were incubated overnight and the DNA was extracted from the colonies as in Sambrook *et al.* (1989). Probes were made using primer extension from viral DNA with random hexamers as primers with the Klenow fragment of DNA polymerase I (New England Biolabs) and [α -³²P]dNTPs, and by end-labelling component-specific primers with polynucleotide kinase (New England Biolabs) and [γ -³²P]dATP. Hybridisations

were carried out using Hybond N+ membranes and Hybaid tubes and ovens according to the methods in the Hybond manual.

RESULTS

Number of genome components

Seven components of the F isolate of SCSV have been identified. More than 20 restriction endonucleases were used to create the full-length clones of these components. Some partial-length clones were also created with *Sau3AI*, *HpaII*, and *HinPI* digests (Chu *et al.*, 1993). Given the small size of the DNAs, however, there may have been other viral components present which did not contain sites for any of those enzymes and would, therefore, have been missed. Three approaches were taken to determine whether more components were present in the F isolate.

The first approach, using back-to-back primers based on a random SCSV library obtained from the original field isolate and which led to the discovery of component 7, has been described. Second, a more comprehensive library was constructed from the RF DNA (Chu *et al.*, 1993), using the 4-base restriction enzyme *HinPI*, and sequenced. *HinPI* was chosen because its sites occurred at least once in all of the components. No novel SCSV-specific sequences were found. Finally, a unique *Eco47III* site is present in the highly conserved region shared between all of the components except those two which encode presumptive replication proteins. Other SCSV components might be expected to contain this common region. The RF DNA was, therefore, digested with this enzyme, end-labelled with [α -³²P]dCTP using the Klenow fragment of DNA polymerase I, and electrophoresed on a 4% polyacrylamide gel. The band corresponding to approximately 1 kb was extracted and cloned into the *EcoRV* site of pGEM5Z. The clones were probed with ³²P-labelled DNA from all known SCSV components. No new components were identified amongst the 134 clones screened.

Sequence of SCSV and identification of ORFs

Figure 1 contains the sequence for seven components of the F isolate of SCSV determined in both orientations with an average of sevenfold redundancy from clones described in Chu *et al.* (1993) and the other clones described in this paper.

A few nucleotide variations were found between different clones of the same DNA component in the F isolate. In component 4, position 132, in the noncoding region, there was a G rather than an A in one clone, and in position 588, in the coding region, there was a T instead of C in another clone, which changes the encoded amino acid from Pro to Arg. In component 7, position 775, in the coding region, there was an A rather than G in one clone, which changes the encoded amino acid from Gly

Component 1

TAGTATTACCCCGTGCCGGGATCAGAGACATTTGACCAATAGTTGACTAGTATAATAGCCCTTGGATTAAATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGCT 120
 GAACGAATCTGACGGAAGAGCGTTCACACTTAGATCTAGTTAGCGTACTTAGTACGCGTGTCTTGGGCTCTATAAATAGAGTGCCTTCTGAACAGATTGTTTCAAGATTTTCATAGCGAGATG 240
 D S G D G Y N T Y S Y E E G A G D A K K E V L Y K I G I I M L C I V G I V V L W
 GATTCTGGTGTGTTACAATACATACTCATATGAAGAAGGTGCTGGAGATGCGAAGAAGGAAGTTTATATAAAATAGGTATTATTATGTTATGTTAGGGATTGTAGTTTATGG 360
 V L I I L C C A V P R Y A K S T M D A W L S S S S I M K R K M A S R I T G T P F
 GTTTTAATATATTATGTTGTGCTGTTCTCGCTATGCTAAATCAACGATGGACGCTTGGTTATCTTCGTCTTCTATTATGAAGAGGAAGATGGCTTCAAGGATTACTGGTACTCCGTTT 480
 E E T G P H R E R R W A E R R T E A T N Q N N N D N V N R F S *
 GAAGAACTGGTCTCATCGTGAAGAAGATGGGCTGAAAGAAGAACTGAAGCGACGAACAGAAATAAATGATAATGTAATAGATTAGTTGATATGTTGTAATTTTATATGGATTA 600
 ATGAGAATTATTATTATCTGTTCTTCGCTGTGTTTTTAAGCTTTTCTGTGTTTTAATGGCGCTCGGAGAGAGAAAGGAATAATTGTAAGGTAGACGACGATGTAGTGGATTACAGT 720
 TGTCTTTACTTCGCTCGAAGAAGACACATTTCAAGTTGTGAGTGTATGCTTTTGAGGAAGCTTCTCGAAGCAGCGTATAACTTTAATTGAAATTTGGTTTGGCGCTTAGTGAA 840
 ATTGCGGCTGTAAACGTGTCAAGTTGTGAGTGGCTGAAATAAGATAATAGATATATTATTATTGTTTTAATTAAATTCGCGAAGCGATATGTTAAGTGATAAATGAAACGAAGCGTTTT 960
 GATGACGTCATATGCTCCGTGCCTACGTAGCACGGGGCT 1001
 ----->

Component 2 (replication protein)

TAGTATTACCCGACCTTGCACACCTCCTTGGAACTTTCTCTCTCTAGAAAGTGTGAGACTTTCTCTCTCTAAGCTTATGCTAGAGGTACTGTTTACATTAAATTACGCTACT 120
 E I E R E T F L S L F S Q D E L N Y F V V G D E T A T T G Q K H L Q G F V S F K
 GAGATAGAGAGAGAAACATTTCTCTCTCTCTCTCAAGACGAATTAACATTTTCTGTTGTCGGCGACGAACTGCAACTACTGGACAGAAACACCTCCAGGGATTGTTATCGTTCAAG 240
 N K I R L G G L K K K F G N R A H W E I A R G S D S Q N R D Y C C K E T L I S E
 AACAAATTCGCTCTGGTGGATTGAAGAAGAAATTTGGTAATCGAGCTCAGTGGGAAATTCGAGAGGCGAGCGATTCTCAGAAATCGGATTATTGCTGTAAAGAAACCCATTTCTGAA 360
 I G I P V M K G S N K R K T M E I Y E E D P E E M Q L K D P D T A L R C K A K K
 ATTGGGATTCGGTCAATGAAGGTTTCGAACAAGCGGAAGACGATGGAGATTATGAAGAGGATCCCGAAGAAATGCAATTGAAGGATCCAGATACTGCTCTTCGATGTAAGCGAAGAAA 480
 L K E E Y C S C Y D F Q K L R P W Q I E L H A A L M A E P D D R S I I W V Y G S
 TTGAAGAGGAATATTGTTCTGTTATGATTTTCAGAACTCCGTCATGGCAAATGAGCTTCACGAGGATTAAATGGCGGAACAGATGATCGGAGTATCATCTGGGTCTATGTTCA 600
 D G G E G K T S F A K E L I R Y G W F Y T A G G K T Q D V L Y M Y A Q D P E R N
 GACCGAGGAAGGAAGGAGCTTCGCGAAGGAATTAATCAGGTATGGATGGTTTTATACAGCGGAGGGAAGACCCAGGACGTATTATATATGATGCTCAAGACCCAGAGAGGAAT 720
 I A F D V P R C S S E M M N Y Q A M E M L K N R V F A S T K Y R P V D L C I R K
 ATTGCGTTGATGTTCCAGGTGTTCTTCGAGATGATGAATATCAGGCGATGGAGATGTTGAAGAACAGAGTTTTCGAAGTACAAATATAGGCTGTAGATCTTTGATTAGGAAG 840
 L V H L I V F A N V A P D P T R I S E D R L V I I N C *
 TTAGTTCATTTAATGTGTTTGGCAACGTGGCACCTGACCCACGCGCATAGTGGAGACAGACTTGAATTATCAATGTTGTAATAGGAATATATATTATTGTTTTAATTAAATTC 960
 GCGAAGCGGTAGCCGTCATAACACTGTTGCCCTTGGAACTATATATAGCAAGGTCGGCT 1022
 ----->

Component 3

TAGTATTACCCCGTGCCGGGATCAGAGACATTTGACCAATAGTTGACTAGTAAATAAGCCCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGC 120
 TGAACGAATCTGACGGAAGAGCGGACATACGCACATGATTATGGCCACATGTCTAAAGTGTATCTCTTTACAGCTATATGATGTGACGTAAGATGCTTTACTTCGCCCTCGAAGTAAA 240
 GTAGGAATTTGCTCGCTAAGTTATTCTTTCTGAAAGAAATTAATTTAATCTAAATTAATTAATAGAGTGGCTATATAATAGATGTTTCTGTTCTGTTTACAACGAAGCTTAGAA 360
 M A L R Y F S H L P E E L K E K I M N E H L K E I K K K E F L E N V I K A
 TCTTGTGTTAATGGCGTTAAGGTATTCTCTCATCTCTCTGAAGAACTGAAGGAGAAGATTATGAACGAGCACTTGAAGGAAATTAAGAAGAAGGAATTTCTAGAGAATGTAATTAAGC 480
 A C A V F E G L T K K E S V E E D D I L R F S G F L E G L S A Y Y A E A T K K K
 TGGCTGTCTGTGTTCAAGGTTTAAACAAGAAGGAGTCTGTTGAAGAAGACGACATACTACGCTTCTCTGGGTTTCTGGAAGGCTCTGCTGCATATTATGAGAGGCGACGAAGAGAA 600
 C L V R W K K S V A I N L K W R V M E E M H Y K L Y G F A D M E D L Y C S E L G
 GTGTTTAGTTAGATGAAGAAGAGCGTTGCAATAAATCTGAAATGGAGAGTTATGGAGGAGATGCATTACAGCTTTATGGATTTCAGACATGGAAGATTTATATTATTCAGAGTTAGG 720
 F P N Y G E D D V A Y H D G A I V N C K Q L E V V F D D L G I E F M S I V I D R
 GTTCTCTAATTACGGTGAAGACGATGTAGCTTATCACGATGGTCAATTGTAAATGTAAGCAATTAGAAGTTGTATTTGATGATTAGGTATTAGTTTATGTCTATTGTAATTGATAG 840
 G S I K I E L *
 AGGTTCTATTAGATAGAATTATGAGATGAATTGTGATTAAATGAATAAGAGTTGTTATTCTTTGAATTACTCCGCGAAGCGGTGTGTTATGTTTTTGTGGAGACATATGACGTC 960
 ATATGCTCGCCGACAGGCTGGCACGGGGCT 991
 ----->

FIG. 1. Sequence of SCSV. The sequence of each component is listed separately, with the deduced amino acid sequence of each ORF shown above. Known or possible functions encoded by a component are indicated in brackets. The stem-loop sequence is marked by the arrows, the potential TATA boxes and polyadenylation signals are underlined, and *denotes the translation stop codon. The amino acids comprising the nucleotide binding motif in the putative gene products of components 2 and 6 are shown in bold italics.

to Asp. In the variable positions the sequences presented in Fig. 1 show the nucleotide most commonly found. It is possible that some of these differences were artefacts of PCR.

Figure 2 contains a linear representation of the SCSV genome organisation showing the relative positions of the ORFs, gene regulation signals, and a highly con-

served common region. In addition to an AAUAAA-like polyadenylation signal, all components have a downstream UG-rich sequence: UG-rich sequences are found associated with many polyadenylation signals and may be required for efficient termination (Gil and Proudfoot, 1984; Conway and Wickens, 1985). The possible TATA boxes and polyadenylation signals are appropriately

TAGTATTACCCCGTGCCGGGATCAGAGACATTTCGACTAAATGTTGACTTGAATAATAGGCCCTTGATATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTAGAAGCGCTTAAGCT	120
GAACGAACTCTGACGGAAGAGCGGCACAAACGCACATGGGACTTAGGCCCACTGCTTTATTAAGAAGTGAATGACAGCTGTCTTTTGCTTCAAGACGAAGTAAAGAATAGTGGAAAACGCGTA	240
AAGAATAAGCGCTACTCAGTACGCTTCGTGGCTTTATAAATAGTGCTTCGCTCTATTCTTCGTTGTATCATCAACGAAGAAGTTAAGCTTTGTTCTGCGTTTAAATGGGGGACTGGTTTCA	360
S A L K T C T H V C D F S D I K A S S Q Q D F F C C D S M R G K L S E A P R K V L	
CAGTGGCGCTTAAGACATGTACTCATGTTCTGTGATTTTCAGATATTAAAGCGCTCTTCACAACAGGATTTCTCTGTTGTGATAGTATGCGAGGATAAATTATCTGAAACCTAGGAAGGTGTT	480
L V S C F V S F T G S F Y G S N R N V R G Q Q V Q L G M Q Q D D G V V R P I G Y I	
GTTAGTTAGTTGTTTGTAAAGTTTACTCGTAGTTTATTGGAAGTAATAGGAATGTTAGAGGTCAGATTTCAGTTGGGTATGCAGCAAGATGATGGCGTTGTTTCGTCCAATAGGATATAT	600
P C T T A G G G Y L Y H D D Y G Y Y Q G E K T F N L D I S E S D Y L K P D E D F W K R F T	
TCCTATGCGGGTATTGTATCATGATGATTATGGATATTATCAAGGAGAGAAGACGTTCAATCTGGACATCGAGTCAGATTATCTGAAGCCTGATGAAGATTTTGGGAAGAGATTAC	720
I N I V N D K G L D D R C D V K C Y V V H T T M R I K V *	
AATTAAATATTGTAATGATAAAGGATTAGATGATAGGTGTGATGTAATAATCTTATGTAGTTCATACGATGCGTATTAAAGGTGAATTGTTATTATCAATAAAGAATTTTATTGTTATT	840
GTGTTATTGTGTAATTTATGCTTATAAGTAATTTCTATGATTAAATGTGGAATTAATAAGACTAATGAGGATAATAATTAATGAAATTTGATTAAATTAACCTGCGGAAGCTATATGTCTTTCACG	960
TGAGAGTCACGTGATGTCTCCCGCAGCAGGCTGGCAGGGGGCT	1002

TACTGATACCCCGTGC CGCGGCTCAGAGACATTTGACTAAATATTGACTTGGAAATAATAGCCCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGCT GAACGAATCTGACGGAAAGCGTCTATGGTCCACATGTCTAAAGAATAATGCTTTACAGCTGTATTGATTGACTTTACGCGCTTTACTTTAATTGCTTTAAGTAAAGTAAGATGCTTTAC TTTGCTTCGCGCAGCAAGCAAAGTGATTTGTAGCTGCGAGAAATGTAGCTTTAATTACCGGGTAAACCGGTTTGATTGTGGGTATAAATATGTGTTCTGTTCGTTTCTTCTGTTGTCAATTTACA M V A V R W G R K G L R S Q R R K Y S R I A A Y K P P S S K V V S H V E S V L ACGAAGATGGTGTGCTGTTTCGATGGGGAAGAAAGGCTCTGAGGTCTCAAAGGAGAAAAATATTGCGGAATTGCTTACAAACCTCCTTCGTCTAAGGTTGTAAGTCATGTGGAGTCTGTCTCG AATAAGAGADAGTGTACTGGAGCGGAGGTTAAGCCATTGCTGATGTTCAAGGTATAGTATGAAGAAGGTAATGTTGATGCAACATTAACTATGGCTCCTGGAGAAATAGTTAATTAT L J V K S N S P I A N W S S F S N P S L M V K E S V Q D T V T I V G G G K L E CTTATTGTGAAGAGTAATTCGCCATTGCGGAATTGGAGTTGCTCTTTTCAGTAATCCTTCGTTGATGGTGAAGAGTCTGTTCAAGATACAGTTACGATTGTTGGAGGAGGAAAGCTTGAG TCTTCGTGACTGCTGGAAGDGTAACTAAGTCTTTTAGGAAGTTTGTTAAGCTCGGTTACGAGTATPAGTCAGACCAGCATTTGTATTAAATTATTATCCAGTGATGCGGATGAAG I T L E T R M Y I D V * ATCACACTGGACAGCAGAATGTATATGTATGATATAATTGTGATGATTAAATGAATAAGAGTGTGTTTTATTCTTTGAATTACTCGCGGAAGCGGTGTGTTATGTTTTGTTGGAGACATA TGACGTCATATGTCTCCGCGACAGGCTGGCACGGGGCT	120 240 360 480 600 720 840 960 998
--	--

CAGTATTACCGCACCTCGGCTTACCCCTCCCTCGCTTCCCTGGGCCCACTATGCCTACTAGACAAAGCACTAGTTGGGGTGTTCACACTTAACTTTGAGGGCGCAAAATCTCTATTTTGCCCTTTA
 120
 NESVQYQYACWQHHERVGVGHDLHQLGQGFIIQFKSRNTTLRQAKKYIFN
 ATGAAGNCSTTCAGTACGCTTGTGGCAGCAGTAGAGAGTGGGACACGATCATTACAGGGATTATACAACTTAAATCCGCAACTACATTCGGCTCAAGGTATATTTTAATG
 240
 GLNPHLEIARDVVEKAQQLYAMKEDSRVAVAGPWWEYGLFVKRGS
 GACTGAATCCTCATCTGGAAATTCGTAGGGATGTAGAGAAGCGCAATTTGACCGGATGAAGGAAGATAGTAGAGTAGCTGGTCCCTGGGAGTATGGGTGTGTTATTAAAGAGAGGATCGC
 360
 HKRKLMLMERFEEEDGEEEMKIIADPSSLYRRRLCTLSRKMMAEEQRCS
 ATAAGCGTAAGCTGGAGAGATTTGAAGAAGATGGAGAAGAGATGAAATTCGTGATCCCTCTCTATAGCGGTGTCTCTATCAAGGAAGATGGCTGAAGAACAACGCTGTCTCTG
 480
 EWNYYDLRFPWQEEVVMHLLLEESP DYRTIIWVYGFAGNFKST
 AGTGGAAATTATGACTTACGCCCTTGGCAAGAAGAGTGTGCAATTTGTAGAGGAAGAACAGATATAGAACGATAATCTGGGTGTATGGACCTGTCTGTAATGAAGGCAAAATCTACAT
 600
 FARHLLSLKDGWGYLPGGKKTQDDMMHLLVTAEBPKNNWVFPDIPR
 TTGCAAGACATCTGTGATTAAGAGATGGTTGGGGTTATCTGCCTGGAGGAAAGACACAAAGATATGATGATCTTGTGACTGCTGAGCCTAAGAATAATTTGGGTATTGACATACCCAGAG
 720
 VSSSEYVNYNGVIEQVKNRVMVTNTKYEPVCMRDDNHFPVHVIV
 TTAGTTTCAGAGTATGTGAATTTATGGTGAATAGAACAGGTTAAGAATAGGGTATGGTGAATACTAAGTATGAGCCATGTGTAATCGCGGATGATAATCATCCTGTTCTATGTAATTGTGT
 840
 FANVLPLPDLGKLSEDRIRKILRC*
 TTGCAAAATGTACTCCAGATTGGGAAAAATTAGTGAAGATAGAAATAAATAATTAATTCGTGTGTGAAAACCTCTGCGAAGGCAGAGTTATAAAAAAATGTGTTTGTGAGAGAAGTCCCA
 960
 TCGGTAGTTTCGCGAAACAGGTTGAGGGAAGCGAGCAATATAGGCGAGGTGCGTAT
 1017

TAGTATTACCCCGTGGCCGGATCAGAGACATTGACTAAATATTGACTTGGAAATAATAGCCCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGCT	120
GAACGAATCTCGACGGAAGAGCGGACATACGCACATGGATTATGGCCACATGTCTAAAGTGTATCTCTTTACAGCTATATTGATGTGACGTAAAGATGCTTTACTTCGCTTCGAAGTAAAG	240
TAGGAAATTGCTCGCTAAGTTATTCTTTTCTGAAAGAAATTAATTTAATCTAATTAATAATGAGTGGCTATAAAATAGTGTCGATGCTGCCTCACATCGTATTTCTTCTCGCATCGT	360
CTGTTCTGGTTTTTAAGCGATGGTCAGTTTAGTTTCTCGTAGATATACGATGTGAGCGACGATGTTCTTGTGAAGCGATAGCAGAAGAAGTGTAGCTGTTGAGTTCGAAGAGAAGGTTCAA	480
GTAINVKNVLRRLIEAIVEDRVGVKVMFRLCYRYRRELKILTLGGTGAATTAACGTCGAAGTACTGAGGTGATTGAAGCTGTGATGAAGATAGAGTTGGAGTGAAGGTTATGTTTCGTCGTGTACAGATACAGACGAGAGAATTAAGATTACGTTTGGGT	600
CKMELWTSLSLKS S GKY SVQSL LQRK L N G I C V S N Y C I G I D M F TGTAAAGTGGAGCTATGGAAGTCTCGTTGAAGTCTTCAGGCAAGTATTCAGTTCAATCTTTGTTGCAGAGGAAGCTTAATGGTATATGTGTTAGTAATTACTGTATAGGTATGTATATGTTT	720
GTAAAGTATGTTAAAGAGTTGATTAATAGATGATAAAGGATACATCTGTTCAAGGTTAACTCCATATGCTTGTATCATATGDCGAAGAGTAATAATAGTAAATTATGATTAAT	840
TATGAGATAAGAGTTGTATTAAATGCTTATGAGGATAAAGAATGATTAATAATTGTTAAATTTATTTCGCAAGCGGTGTGTTATGTTTTTGTGGAGACATCAGTGACTCTCACGTGA	960
TGTCCTCCGGACAGGCTGGCACGGGGCT	988

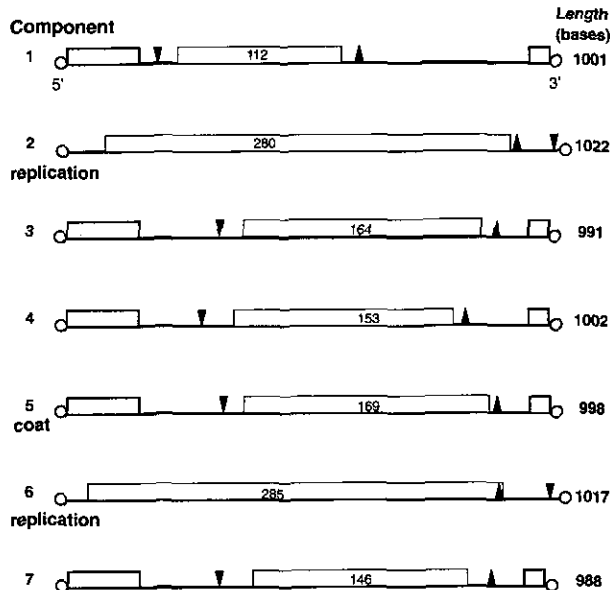


FIG. 2. Genome organisation of SCSV. The components are represented in linear form starting at the first nucleotide of the conserved 9-nucleotide sequence in the stem-loop. The numbers in the open boxes indicate the sizes of the putative protein products. Known or possible functions encoded by a component are indicated below that component number. ■, common region; □, open reading frame; O, conserved stem-loop sequence; ▼, TATA box; ▲, polyadenylation signal.

placed with respect to the largest ORF in each component. These ORFs are all read from the virion or + strand. Component 5 has been identified as encoding the coat protein by comparison with N-terminal amino acid sequence of protein purified from viral particles (Chu *et al.*, 1993). Components 2 and 6 both encode proteins containing consensus nucleotide binding motifs, GXGK(T/S) (Walker *et al.*, 1982), see Fig. 1. Additional ORFs (not shown) were detected using FRAMES; of these, the only ORF with the potential to encode a protein greater than 10K is on component 6, starting at position 937 in the same orientation as the major ORF. This ORF does not appear to have appropriate consensus TATA box or polyadenylation signal sequences.

Sequence relationships between the SCSV components

All components contain a sequence in the noncoding region which can form a hairpin structure with a GC-rich stem and an AT-rich loop. This sequence will be referred to as the stem-loop sequence, although the existence of the structure has not been proven. The sequence within the loop contains a conserved 9-nucleotide sequence, 5'-(T/C)AGTATTAC-3' (Fig. 3). Components 1, 3, 4, 5, and 7 also share sequence surrounding the stem-loop sequence, called the common region (see Fig. 2). This extended region of sequence matching varies in size between 153 and 260 nucleotides and is greater than 95% similar in pairwise comparisons. The common

region shown in Fig. 2 is the largest sequence which is shared among all five components. Components 2 and 6 may lack a similar region due to the large size of their coding regions.

Components 2 and 6 have similar base ratios, codon usage (data not shown), and ORF sizes, distinct from other components. The putative gene products of these two components were found to share 43% amino acid identity using GAP. The other components, despite some having similar ORF sizes and base ratios, were not found to share significant nucleotide or amino acid similarity using GAP.

Sequence relationships with other viral genomes

The SCSV conserved stem-loop sequence is very similar to that of geminiviruses; this structure is involved in viral replication (Lazarowitz *et al.*, 1992; Schneider *et al.*, 1992; Heyraud *et al.*, 1993). The loop sequence of the geminiviruses has eight of nine matches with the sequence from SCSV components 1–5 and 7, see Fig. 3. BBTv and CFdV have similar conserved loop sequences (Fig. 3; Harding *et al.*, 1993; Rohde *et al.*, 1990). Bipartite geminiviruses also have regions of about 200 bases in each component which are highly conserved between the DNA components of any one virus but not between viruses (Davies and Stanley, 1989). As in SCSV, this common region encompasses the conserved stem-loop sequence.

The protein sequences encoded by the major SCSV ORFs were compared with amino acid sequences in the SwissProt database using FASTA. No significant matches were found when any of the sequences were compared individually. The amino acid sequences of the putative products of components 2 and 6, the proposed replication proteins, were found to share about 40% homology with the equivalent proteins of BBTv (Harding *et al.*, 1993) and CFdV (Rohde *et al.*, 1990). A profile of these replication proteins was constructed with PROFILEMAKE and compared to the SwissProt database with PROFILESEARCH. The best matches occurred with the replication proteins of the bipartite geminiviruses. A subsequent alignment, created with PILEUP, of the SCSV-like viruses' (SCSV, BBTv, and CFdV) and representative geminivirus replication proteins revealed homology not only around the nucleotide binding motif, but also at the positions of three other motifs identified by Koonin and Ilyina (1992) in replication proteins involved in rolling-circle replication (Fig. 4).

DISCUSSION

The seven DNA components sequenced may represent the complete genome of the F isolate of SCSV. We were unable to detect the presence of additional components. However, the search was not exhaustive. Infectivity studies will be required to ascertain the number of essential genome components.

		5'			3'
SCSV components	1	AGCACGGGGCT	TAGTATTAC	CCCCGTGCC	
	3	GGCACGGGGCT	TAGTATTAC	CCCCGTGCC	
	4, 5, 7	GGCACGGGGCT	TAGTATTAC	CCCCGTGCC	
	2	GCAAGGTCGGCT	TAGTATTAC	CCGACCTTGC	
	6	GCGAGGTGCGTAT	CAGTATTAC	CGCACCTCGC	
CFDV		AGCCCGCGGGGC	TAGTATTAC	CCCCGCGGCT	
BBTV		AGCGCTGGGGCT	TATTATTAC	CCCCAGCGCT	
Geminiviruses	ACMV	GGGGCCAAACGTA	TAATATTAC	CGGTTGGCCCC	
	BCTV	GGGCCATCCGTTA	TAATATTAC	CGGATGGCCCC	
	MSV	GCAGGAAAAGAGCGCGCAC	TAATATTAC	CGCGCCTTCTTTCTCTGC	

FIG. 3. Comparison of the stem-loop sequences of SCSV, BBTV, CF DV, and three representative geminiviruses, African cassava mosaic virus (ACMV), beet curly top virus (BCTV), and maize streak virus (MSV). The conserved 9-nucleotide sequences are boxed. The complementary sequences which make up the stems are underlined.

Each DNA component of SCSV is predicted to have unidirectional transcription expressing a single ORF. The only ORF which has been firmly identified is that of component 5, the coat protein gene (Chu *et al.*, 1993). The

ORFs of components 2 and 6 both appear to encode replication proteins since they are homologous to the replication proteins of geminiviruses, as shown in Fig. 4. The relatedness of components 2 and 6 is no greater than either of them with the corresponding BBTV and CF DV genes, suggesting that they may have come from two different viruses in the SCSV-like group.

SCSV shares close similarity with BBTV and CF DV and more distant similarity with the geminiviruses. It does not appear to share sequence or noncoding region secondary structural features with chicken anemia virus (Todd *et al.*, 1990) or psittacine beak and feather disease virus (Ritchie *et al.*, 1989), which had been suggested to be related to the SCSV-like viruses.

BBTV and CF DV have several circular single-stranded DNAs and small isometric particles of 17–20 nm (Randles and Hanold, 1989; Harding *et al.*, 1991). BBTV and CF DV DNAs each contain a stem-loop structure with closely related loop sequences (Fig. 3). Their replication proteins (Rohde *et al.*, 1990; Harding *et al.*, 1993) share about 40% identity with SCSV. Virus transmission is limited to the insect vectors; SCSV and BBTV by aphids and CF DV by a plant hopper. The viral genes appear to be expressed unidirectionally.

These three viruses appear to form a distinct group which may include faba bean necrotic yellows virus (FBYV; Katul *et al.*, 1993), milk-vetch dwarf virus (MDV; Isogai *et al.*, 1992), and porcine circovirus (PCV; Tischer *et al.*, 1982). FBYV is an aphid-transmitted virus of legumes which appears to have 18-nm isometric particles and multiple 1-kb circular single-stranded DNA components (Katul *et al.*, 1993). MDV is aphid transmitted, causes a yellowing disease, and contains single-stranded DNA 1 kb in size (Isogai *et al.*, 1992). PCV has 17-nm icosahedral particles, made up of a 36K protein, which contain a 1.7-kb circular single-stranded DNA (Tischer *et al.*, 1982). It was reported by Rohde *et al.* (1990) that PCV ORF4 shares sequence homology with the replication protein-encoding ORF of CF DV.

The SCSV-like viruses also appear to be related to the geminiviruses. Both groups have circular single-stranded DNA genomes and a single stem-loop sequence in the noncoding region (Fig. 3). The loop contains a highly

ACMV	1	...MRTPRF	RVQAKNVPLT	YPNCSIPKEH	LLSFIQTLSL	PSNPKFIKIC	50
MSV		MASSSSNRQF	SHRANMTPLT	YPKCPENPEI	ACQMIWELVV	RWIFKYILCA	
SCSV2			..MARRYCPTL	NYATEIERET	FLSLFQDEL	NY...FVVG	
CFDV		MG	SSIRRMCTTL	NYETEEAAN	VVRRTESLNL	VY...AIVGD	
SCSV6		MPT	RQSTSWVFTL	NFEGEI...P	ILPFNESVQY	AC...WQHE.	
BBTV		MA	RYVVCWMTI	NNPT.....T	LPVMDIEIKY	MV...YQVER	
			*** **				
ACMV	51	RELHQNGEPH	LHALIQFEG.	KITITNNRLF	DCVHPSCTSN	FHPNIQGAKS	100
MSV		REAHKDGSLH	LHALLQTEK.	PVRISDSRFF	DI.....NG	FHPNIQSAKS	
SCSV2		ETA..TTGQKH	LQGFVSPFN.	KIRLGGLKKK	FGN	RAHWIARAG	
CFDV		EVAPSTGQRH	LQGFHILKT.	GRRLLQGLKT	LGN	DRHLEPTRG	
SCSV6		...RVGHDH	LQGFIFQKSR	NTTLRQAKYI	FNG	LNPHLEIARD	
BBTV		G...QEGTRH	VQGYVEMKRR	S.SLKQMRGF	FPG	..AHLEKRRG	

ACMV	101	S.SDVKSYLE	KDG...DTV.	EWGQF.....	...QIDGRSA	RG...GQOSAN	150
MSV		V.NRVRDYIL	KEP...LAVF	ERGTFIPR..	..KSPFLGKSD	SEVKEKPSK	
SCSV2		SDSQNRDYCC	KET...LIS	EIGLIPVMKGS	NKRRTMEIYE	EDPE.....	
CFDV		SDEQNRDYCS	KER...VLL	EHGVPTRPV	KRPRLAQRF	EEPD.....	
SCSV6		VE.KAQLYAM	KEDSRVAGPW	EYGLFTIKRGS	HKRKLMEPFE	EDGE.....	
BBTV		SQEARSYCM	KEDTRIEGPF	EFGSF...KLS	CNDNLFVDIQ	DMRETHRPL	
ACMV	151	DAYAKALNS.	..GSKSEALNV	IRELVKPDFV	LQPHNLNSNL	DRIFQEPFAP	200
MSV		DEIMRDLISH	ATSKEEYLSM	IQKELPFDM	TKLQVFEYS	NKLFPEIQEE	
SCSV2		EMQLKDPDTA	LRCRAK...KA			...KEE	
CFDV		ELRLDPGCGY	RRCVH...GA			...SVE	
SCSV6		EMKIADPSLY	RRCLSR...KM			...AEE	
BBTV		EYLYDCPNTF	DRSKDLYRV			...QAE	
ACMV	201	VVSFFPCSSF	DQVPDE.LEE	WVADNVDRSA	ARPWRPNSIV	IEG.DSRITGK	250
MSV		FTNPHFPSSP	DLTCSNENND	WLQPNIFQSS	DESRKQSLY	IVG.PTRITGK	
SCSV2		YC....SCY	DF...QLRP	WQIELHAALM	AEPDRSIIW	VYSGDGGEGK	
CFDV		WTRWAENPF	PF...P.YHN	WQIEVLGAIG	EPADRTILW	ICGRDGGEGK	
SCSV6		...QRCSSSEW	NY...D.LRP	WQEEVHMLE	EEPDRYRTI	WYGPAGNEGK	
BBTV		MNKTKAMNSR	R....TSFSA	WTSEVEN.IM	AQPCRRRIIW	VYGPNGGEGK	
ACMV	251	TIWARSLSGP	HNLYLQGLDL	SPKV...FNN	DAWYNIVDDV	DPHYLKHFK.	300
MSV		BTWARSLSGV	HNWQNNVDW	SS....YNE	DAIYNIVDDI	PFKPCPCWK.	
SCSV2		TSFAKEL.IR	YGFYFYTAGGK	TQDVLVYMAQ	DPERNIAFDV	PRCSSEPMNY	
CFDV		SVFAKYLGLK	PDWFFYTCGGT	RKDVLYQYIE	DPKRNLIILV	PRCNLEYLYN	
SCSV6		STFARHLSLK	DGWGYLPGGK	TQDMHMLVTA	EPKNNWVFDI	PRVSSEYVNY	
BBTV		TYAKHLMTK	RNIFYSPGGK	SLDTCRLVNY	E...DIVIFDI	PRCKEDYLYN	
ACMV		EFMGSDRDWQ	..SNTKYGK.P	VQI.KGGIPT	IFLCN..PGP	TSSYKEFLDE	
MSV		QLVCCQDFET	..VNPKYKG.K	KKVQKSKPT	IILAN..SD.	...EDWKE	
SCSV2		QAMEMLNKRV	PASTKYRPVD	LCIRKL.VHL	IVFANVADPD	TRISEDRLVI	
CFDV		ALLBCVKNRA	FSSDKYEPVS	YLGPDH.VHV	LVFANVLPDY	LKISRDRILK	
SCSV6		GVIEQVKNRV	MVNTKYEPVCV	MRDDNHVPV	IVFANVLPDL	IKISDRILK	
BBTV		GLLEEFKNKI	IQSGKYEPVL	KIV...EYVEV	IVMANFLPKE	GTFSEDRIKL	
ACMV	351	EKQEAALKAWA	LKNAIFI.TL	TEPLYSGSNQ	SQSQTIQEAS	HNA	393
MSV		MTFGQLEYFE	ANCIYIMSP	GEKWYSPPEL	PPTEAVHSR	S...	
SCSV2		INC.....				...	
CFDV		WNI.....				...	
SCSV6		IRC.....				...	
BBTV		VSC.....				...	

FIG. 4. Alignment, created using PILEUP, of the replication proteins of the SCSV-like viruses and two representative geminiviruses, African cassava mosaic virus (ACMV; bipartite) and maize streak virus (MSV; Nigerian isolate fusion protein; monopartite). The conserved motifs identified by Koonin and Ilyina (1992) are indicated by * and the nucleotide binding motifs are in bold.

conserved 9-nucleotide, nearly palindromic sequence, 5'-TA(G/A/T)TATTAC-3', which is involved in the initiation of replication (Lazarowitz *et al.*, 1992; Heyraud *et al.*, 1993). The presumptive replication proteins of SCSV components 2 and 6, BBTV, and CFDV show closest similarity with those of geminiviruses. The conservation of motifs identified by Koonin and Ilyina (1992) and the overall similarity to the geminivirus replication proteins, see Fig. 4, suggest that SCSV-like viruses may also replicate using a similar rolling-circle mechanism.

There are a number of other similarities between the SCSV-like viruses and geminiviruses. Both monopartite geminiviruses and SCSV have primer sequences associated with the encapsidated DNA (Donson *et al.*, 1984; Chu and Helms, 1988). The highly conserved noncoding common region of the SCSV components 1, 3, 4, 5, and 7 is similar to that of bipartite geminiviruses. Small spherical particles have been found in preparations of African cassava mosaic virus and chloris striate mosaic virus (see Stanley, 1985), in addition to the normal geminate, or "twinned," particles. These appear to be associated with defective minicircles and thus mimic SCSV-like viruses in genome size and particle structure. Finally, CFDV and monocot infecting geminivirus are transmitted by related vectors, that is, leaf hopper species (Bock, 1974; Francki and Hatta, 1980; Julia, 1982).

The SCSV-like viruses and the geminiviruses appear to share a common heritage. A single recombination event between an SCSV-like coat protein-encoding DNA component and the complementary strand of a replication protein-encoding component could give rise to a virus with many similarities to a monopartite geminivirus. The unique geminate particle structure of geminiviruses might be derived from a basic spherical unit in response to encapsidation of a larger DNA component. We suggest that the common ancestor of the geminivirus group and the SCSV-related viruses may have been more like the latter in genome organisation and particle structure.

ACKNOWLEDGMENT

P. Boevink was supported by an Australian Wool Research and Promotion Organisation Postgraduate Fellowship.

REFERENCES

- Bock, K. R. (1974). Maize streak virus. CMI/AAB "Descriptions of Plant Viruses," no. 133.
- Burns, T. M., Harding, R. M., Hafner, G., Beetham, P., and Dale, J. L. (1993). Single-stranded DNA genome organisation of banana bunchy top virus. IXth Int. Congr. Virology, Abstract W62-8.
- Chu, P. W. G., and Helms, K. (1988). Novel virus-like particles containing circular single-stranded DNAs associated with subterranean clover stunt disease. *Virology* **167**, 38-49.
- Chu, P. W. G., Keese, P., Qiu, B., Waterhouse, P. M., and Gerlach, W. L. (1993). Putative full-length clones of subterranean clover stunt virus and identification of the component coding for the viral coat protein. *Virus Res.* **27**, 161-171.
- Conway, L., and Wickens, M. (1985). A sequence downstream of A-U-A-A-A is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes. *Proc. Natl. Acad. Sci. USA* **82**, 3949-3953.
- Davies, J. W., and Stanley, J. (1989). Geminivirus genes and vectors. *Trends Genet.* **5**, 77-81.
- Del Sal, G., Manfioletti, G., and Schneider, C. (1989). The CTAB-DNA precipitation method: A common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. *BioTechniques* **7**, 514-158.
- Devereaux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.
- Donson, J., Morris-Krsinich, B. A. M., Mullineaux, P. M., Boulton, M. J., and Davies, J. W. (1984). A putative primer for second-strand synthesis of maize streak virus is virion-associated. *EMBO J.* **3**, 3069-3073.
- Francki, R. I. B., and Hatta, T. (1980). Chloris striate mosaic virus. CMI/AAB "Descriptions of Plant Viruses," no. 221.
- Francki, R. I. B., Fauquet, C. M., Knudson, D. L., and Brown, F. (1991). "Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses." Springer-Verlag, Wein.
- Gil, A., and Proudfoot, N. J. (1984). A sequence downstream of AAUAAA is required for rabbit β -globin mRNA 3'-end formation. *Nature (London)* **312**, 473-474.
- Grylls, N. E., and Butler, F. C. (1956). An aphid transmitted virus affecting subterranean clover. *J. Aust. Inst. Agric. Sci.* **22**, 73-74.
- Grylls, N. E., and Butler, F. C. (1959). Subterranean clover stunt, a virus disease of pasture legumes. *Aust. J. Agric. Res.* **10**, 145-159.
- Harding, R. M., Burns, T. M., and Dale, J. L. (1991). Virus-like particles associated with banana bunchy top disease contain small single-stranded DNA. *J. Gen. Virol.* **72**, 225-230.
- Harding, R. M., Burns, T. M., Hafner, G., Dietzgen, R. G., and Dale, J. L. (1993). Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. *J. Gen. Virol.* **74**, 323-328.
- Heyraud, F., Matzeit, V., Schaefer, S., Schell, J., and Gronenborn (1993). The conserved nonanucleotide motif of the geminivirus stem-loop sequence promotes replicational release of virus molecules from redundant copies. *Biochimie* **75**, 605-615.
- Isogai, M., Sano, Y., and Kojima, M. (1992). Identification of the unique DNA species in the milk-vetch dwarf virus-infected leaves. *Ann. Phytopathol. Soc. Jpn* **58**, 631-632.
- Julia, J. F. (1982). *Myndus taffini* (Homoptera Cixiidae), vector of foliar decay of coconuts in Vanuatu. *Oléagineux* **37**, 409-414.
- Katul, L., Vetten, H. J., Maiss, E., Makkouk, K. M., Lesemann, D.-E., and Casper, R. (1993). Characterisation and serology of virus-like particles associated with faba bean necrotic yellows. *Ann. Appl. Biol.* **123**, 629-647.
- Koonin, E. V., and Ilyina, T. V. (1992). Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *J. Gen. Virol.* **73**, 2763-2766.
- Lazarowitz, S. G. (1992). Geminiviruses: Genome structure and gene function. *Crit. Rev. Plant Sci.* **11**, 327-349.
- Lazarowitz, S. G., Wu, L. C., Rogers, S. G., and Elmer, J. S. (1992). Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. *Plant Cell* **4**, 799-809.
- Randles, J. W., and Hanold, D. (1989). Coconut foliar decay virus particles are 20 nm icosahedra. *Intervirology* **30**, 177-180.
- Ritchie, B. W., Niagro, F. D., Lukert, P. D., Steffens, W. L., and Latimer, K. S. (1989). Characterisation of a new virus from cockatoos with psittacine beak and feather disease. *Virology* **171**, 83-88.
- Rohde, W., Randles, J. W., Langridge, P., and Hanold, D. (1990). Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. *Virology* **176**, 648-651.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual." Second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

- Schneider, M., Jarchow, E., and Hohn, B. (1992). Mutational analysis of the conserved region of maize streak virus suggests its involvement in replication. *Plant Mol. Biol.* **19**, 601–610.
- Smith, P. R. (1966). A disease of French beans (*Phaseolus vulgaris* L.) caused by subterranean clover stunt virus. *Aust. J. Agric. Res.* **17**, 875–883.
- Stanley, J. (1985). The molecular biology of geminiviruses. *Adv. Virus Res.* **30**, 139–177.
- Tischer, I., Gelderblom, H., Vettermann, W., and Koch, M. A. (1982). A very small porcine virus with circular single-stranded DNA. *Nature* **295**, 64–66.
- Todd, D., Creelan, J. L., Mackie, D. P., Rixon, F., and McNulty, M. S. (1990). Purification and biochemical characterisation of chicken anaemia agent. *J. Gen. Virol.* **71**, 819–823.
- Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the α and β subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951.