View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

VIROLOGY 208, 99–110 (1995)

Genome Structure and Phylogenetic Analysis of Lettuce Infectious Yellows Virus, a Whitefly-Transmitted, Bipartite Closterovirus

VICKI A. KLAASSEN,*[,1] MAURY L. BOESHORE,† EUGENE V. KOONIN,‡ TONGYAN TIAN,* and BRYCE W. FALK*

*Department of Plant Pathology, University of California, Davis, California 95616; †Experimental Plant Genetics, Asgrow Seed Company, Kalamazoo, Michigan 49001; and ‡National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

Received October 31, 1994; accepted December 29, 1994

We report the complete nucleotide sequences of lettuce infectious yellows virus (LIYV) RNAs 1 and 2. LIYV RNA 1 is 8118 nucleotides and includes three open reading frames (ORFs). Computer-assisted analysis of LIYV RNA 1 ORFs identified domains for a papain-like protease, methyltransferase (MTR), RNA helicase (HEL), and RNA-dependent RNA polymerase (RdRp). We suggest that the RdRp domain is expressed independently of the other replication-associated domains via a +1 ribosomal frameshift. Amino acid sequences of the MTR, HEL, and RdRp show highly significant similarity to the homologous sequences from other closteroviruses and lower similarity to the respective proteins of tobamoviruses, tobraviruses, hordeiviruses, bromoviruses, and furoviruses. LIYV RNA 2 is 7193 nucleotides and includes six ORFs. These ORFs include a gene array that is characteristic of the closteroviruses: ORFs encoding a small membrane protein, a homologue of the HSP70 family of chaperone proteins, a protein whose function is unknown, the coat protein, and a diverged duplicate of the coat protein. LIYV is distinguished from the monopartite closteroviruses in the following ways: its genome consists of two RNAs, the positions of the coat protein gene and its diverged duplicate are reversed, and LIYV includes ORFs that are unrelated to ORFs found in other closteroviruses. © 1995 Academic Press, Inc.

INTRODUCTION

Virions of lettuce infectious yellows virus (LIYV) are flexuous, filamentous particles measuring 1800 to 2000 nm in length (Duffus *et al.*, 1986). Francki *et al.* (1991) tentatively included LIYV in the closterovirus group based on particle morphology and the observation that LIYV induces intracellular vesicles similar to those induced by beet yellows closterovirus (BYV; Duffus *et al.*, 1986). Classification of LIYV as a closterovirus is supported by comparison of the LIYV coat protein amino acid sequence with those of other filamentous plant viruses. This comparison suggests that the LIYV coat protein sequence is most similar to the coat protein sequences of BYV and citrus tristeza closterovirus (CTV; Klaassen *et al.*, 1994). However, LIYV differs in several respects from other viruses within the closterovirus group. Aphid transmission is typical of the closteroviruses; LIYV is transmitted semipersistently by the sweet potato whitefly, *Bemisia tabaci* (Gennadius). The closteroviruses characterized to date contain one ssRNA. We have evidence for two high-molecular-weight RNAs (LIYV RNAs 1 and 2) in LIYV-infected plants and purified virions (Klaassen *et al.*, 1994).

Agranovsky *et al.* (1991a,b, 1994) and Boyko *et al.* (1992) determined the genome sequence and organization of BYV. They identified a unique gene array; it includes open reading frames (ORFs) for a small membrane protein, a homologue of the HSP70 family of chaperone proteins, a protein (p64) whose function is unknown, a diverged duplicate of the coat protein, and the coat protein. Sekiya *et al.* (1991), Boyko *et al.* (1992), and Pappu *et al.* (1994) determined the sequence of part of the CTV genome and identified ORFs for the HSP70 homologue, a protein (p61) with sequence similarity to the p64 protein of BYV, a diverged copy of the coat protein, and the coat protein.

We initiated further studies of LIYV RNAs 1 and 2 with two objectives. First, we wanted to determine if LIYV RNAs 1 and 2 include the gene modules characteristic of BYV and CTV. Identification of these modules in LIYV would provide a basis for classification beyond that of particle morphology, disease symptoms, vector, and host range. Second, we wanted to determine how the LIYV genome would be organized between the two LIYV RNAs. In this paper, we report the complete nucleotide sequence and genome structure of LIYV RNAs 1 and 2. In addition, we show preliminary evidence that the expression of the LIYV genome occurs via mechanisms similar to those of BYV.

MATERIALS AND METHODS

cDNA cloning and nucleotide sequence analysis

LIYV was propagated, virions were purified, and LIYV virion RNAs, total ssRNAs, and dsRNAs were extracted and analyzed as previously described (Klaassen *et al.*,

[1] To whom reprint requests should be addressed.

0042-6822/95 $6.00
Copyright © 1995 by Academic Press, Inc.
All rights of reproduction in any form reserved.

1994). Overlapping cDNA clones representing LIYV RNAs 1 and 2 (Klaassen *et al.*, 1994) were identified using the polymerase chain reaction (PCR; Saiki *et al.*, 1988) and synthetic oligonucleotides based on the 5'- and 3'-terminal nucleotide sequences of individual cDNA clones. Additional cDNA clones representing LIYV RNA 1 were obtained using oligonucleotide primer RMM402 (5'-ATAAGAATGCGGCCGCCGACCTTTATTCTATATATA-3', complementary to the 3'-terminal 20 nucleotides of LIYV RNA 1 and including a *NotI* site) to prime first-strand cDNA synthesis. cDNA was cloned as previously described (Klaassen *et al.*, 1994).

DNA sequencing reactions were performed on both strands of selected cDNA clones by the dideoxynucleotide chain-termination method of Sanger *et al.* (1977) using the Sequenase kit (U.S. Biochemicals). The sequences of internal regions of the cDNA inserts were obtained from deletion clones generated by exonuclease III and nuclease S1 digestion (Henikoff, 1984) or by using synthetic oligonucleotides complementary to internal regions of the cDNA to prime sequencing reactions.

The nucleotide sequences of the 5'-terminal regions of LIYV RNAs 1 and 2 were determined in two ways. First, primers RMM420 (5'-CAACTTCGAGTGCCATCGAA-3', complementary to nucleotides 94 to 113 of LIYV RNA 1) and RMM1 (5'-ACGTCACTCAAGTGCCATCG-3', complementary to nucleotides 134 to 153 of LIYV RNA 2) were hybridized to LIYV RNAs 1 and 2 and extended using AMV reverse transcriptase (Promega) in the presence of dideoxynucleotides (Huiet *et al.*, 1991). Second, reverse transcriptase extension products were synthesized using RMM420 and RMM1 in the absence of dideoxynucleotides. The largest extension product for each RNA was gel-isolated, G-tailed, and amplified by PCR using the original cDNA primer and primer BG07 (5'-GAGCTCGGATC$_{(10)}$-3'; Huiet *et al.*, 1993). The PCR products were cloned into the *Sna*BI site of pBluescript (SK$^+$) (Stratagene). Eight recombinant colonies were selected for each LIYV RNA and the nucleotide sequences of the plasmid inserts were determined.

The 3'-terminal sequences of LIYV RNAs 1 and 2 were determined from cDNA synthesized from polyadenylated LIYV virion RNAs and amplified by PCR (Huiet *et al.*, 1992). BG06 (5'-CTCGAGAAGCT$_{(15)}$-3') was used as a primer for cDNA synthesis; RMM227 (5'-CGAATAGGGTGTTTTGATAG-3', identical to nucleotides 7948 to 7967 of LIYV RNA 1) or RMM18 (5'-CACAAGCGGGTAAGTTGGAT-3', identical to nucleotides 7005 to 7024 of LIYV RNA 2) plus BG06 were used in PCR. PCR products were cloned into the *Sna*BI site of pBluescript (SK$^+$). Eight recombinant colonies were selected for each LIYV RNA and the nucleotide sequences of the plasmid inserts were determined.

RNA hybridization for the analysis of subgenomic RNAs

LIYV virion RNAs, total ssRNAs, and dsRNAs from healthy and LIYV-infected plants were denatured with glyoxal, separated by agarose gel electrophoresis, and transferred to Hybond N+ (Amersham) as previously described (Klaassen *et al.*, 1994). The filters were rinsed in 2× SSC (0.3 *M* NaCl, 30 m*M* trisodium citrate) and hybridized according to the procedure of Amasino (1986) to ^{32}P-labeled single-stranded RNA probes. Negative-sense transcript probes were generated *in vitro* with SP6 RNA polymerase (Stratagene) and linearized LIYV cDNA clones. After washing, filters were treated with RNase A to reduce nonspecific binding of the RNA probe.

Computer analysis of nucleotide and amino acid sequences

The nucleotide sequences of LIYV RNAs 1 and 2 were assembled and analyzed using the University of Wisconsin Genetics Computer Group (GCG) package (Devereux *et al.*, 1984). Amino acid and nucleotide versions of the nonredundant sequence database (NRDB) at the National Center for Biotechnology Information (NIH) were searched for similarity to LIYV proteins using the programs BLASTP and TBLASTN based on the BLAST algorithm (Altschul *et al.*, 1990, 1994). Alignments of amino acid and nucleotide sequences were constructed using the program MACAW (Schuler *et al.*, 1991). Transmembrane segments in proteins were predicted using the algorithms of Eisenberg *et al.* (1984) and Rao and Argos (1986) as implemented in the programs HELIXMEM and RAOARGOS in the PCGENE package (Moore *et al.*, 1988). RNA secondary structure was predicted using the RNAFOLD program (Zuker, 1989) in the PCGENE package. Phylogenetic trees were constructed using the protein parsimony and neighbor joining algorithms as implemented in the PHYLIP package (Felsenstein, 1989).

RESULTS AND DISCUSSION

cDNA cloning and nucleotide sequence analysis

By using PCR to screen a cDNA library, we identified overlapping LIYV cDNA clones for about one-half of LIYV RNA 1 (pSP10 and pSP182, Fig. 1A) and nearly all of LIYV RNA 2 (pSP6, pSP222, and pSP105, Fig. 1B). Additional cDNA clones of the 5'-terminal half of LIYV RNA 1 (e.g., pSP55, Fig. 1A) were obtained using primer RMM402, complementary to the 3'-end of LIYV RNA 1.

We used the cDNA clones shown in Fig. 1 to determine the nucleotide sequences of LIYV RNAs 1 and 2. Four sequence heterogeneities in coding regions were observed between these clones. Three of the heterogeneities were silent. One sequence heterogeneity, position 5140 of LIYV RNA 2, was a 5-bp insertion, resulting in changes in the downstream ORFs. We determined the sequence of this region from four independent clones and directly from LIYV virion RNA. Three of the four clones did not contain the insertion; in addition we could not detect the insertion by directly sequencing LIYV RNA 2.

A comparison of pSP55 and pSP144 nucleotide se-

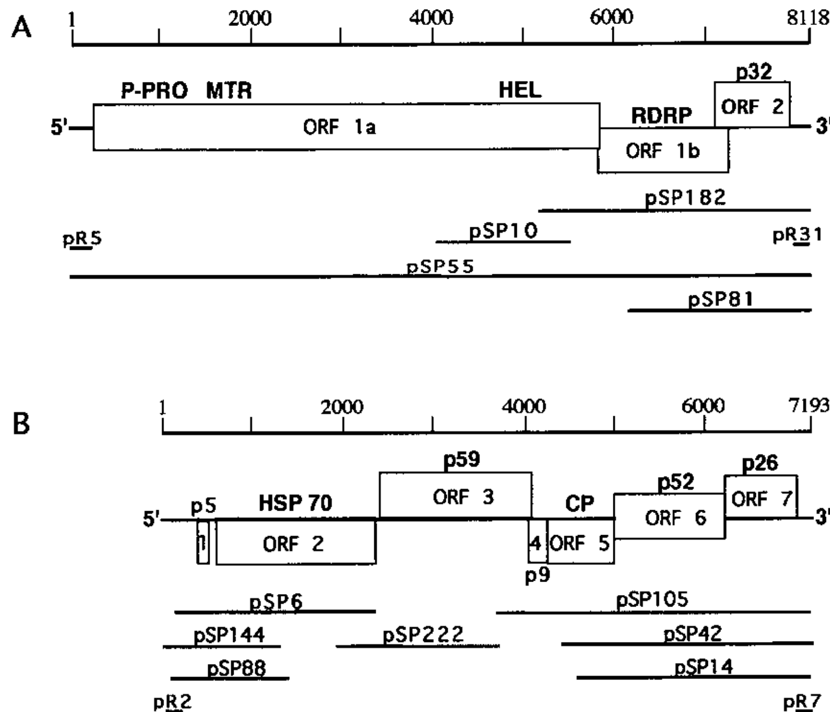

FIG. 1. Genome organization of lettuce infectious yellows closterovirus (LIYV) RNAs (A) 1 and (B) 2. Numbered lines represent nucleotide coordinates, beginning from the 5′ terminus for each RNA. The amino acid sequences of the predicted ORFs in LIYV RNAs 1 and 2 were determined by computer sequence analysis using GCG. P-PRO, papain-like protease; MTR, methyltransferase; HEL, RNA helicase; RDRP, RNA-dependent RNA polymerase; HSP70, HSP70-related homologue; and CP, LIYV coat protein. The start of ORF 1b (LIYV RNA 1) is drawn assuming the suggested +1 frameshift. ORF 2 of LIYV RNA 1, and ORFs 1, 3, 4, 6, and 7 of LIYV RNA 2 are identified by the approximate molecular weights of their putative products. The lines below each LIYV RNA represent the cDNA clones used to determine the nucleotide sequences of LIYV RNAs 1 and 2. Lightly shaded regions represent the negative-sense transcripts used in Northern blot hybridization analysis of LIYV virion RNAs and dsRNAs and total ssRNAs from LIYV-infected and healthy plants (Fig. 8).

quences and primer extension sequencing, using LIYV virion RNAs and primers RMM420 and RMM1, indicated that these clones were 11 and 13 nucleotides short of the 5′-termini of their respective RNAs. However, we still could not identify the 5 5′-terminal nucleotides of LIYV RNA 1 and the 3 5′-terminal nucleotides of LIYV RNA 2 using primer extension sequencing. Therefore, primer extension products were cloned and the sequences of eight independent clones (e.g., pR5 and pR2, Fig. 1A and B) for each LIYV RNA were determined. Sequence data from these clones indicated that the 5 5′-terminal nucleotides of both LIYV RNAs 1 and 2 are 5′-GGUAA-3′.

To identify the probable 3′-ends of LIYV RNAs 1 and 2, we determined the 3′-terminal sequence of a number of clones from the cDNA library and PCR-amplified cDNA. Since the virion RNAs which were used as templates for cDNA synthesis had been polyadenylated *in vitro,* we anticipated that many of the clones would begin at the 3′-ends of LIYV RNA 1 or 2. We identified 4 LIYV RNA 1 cDNA clones (e.g., pSP182 and pSP81, Fig. 1A) which had identical 3′-termini. In addition, the 3′-termini of 6 of 8 PCR clones (e.g., pR31, Fig. 1A) were identical to pSP182 and pSP81. No clones were identified which extended downstream of these clones. These results suggest that the 10 cDNA and PCR clones represent the 3′-end of LIYV RNA 1.

For LIYV RNA 2, we identified 5 cDNA clones (e.g., pSP105, pSP42, and pSP14, Fig. 1B) which had identical 3′-termini. In addition, the 3′-termini of 5 of 8 PCR clones (e.g., pR7, Fig. 1B) were identical to pSP105, pSP42, and pSP14. These results suggest that the 10 cDNA and PCR clones represent the 3′-end of LIYV RNA 2. We could not detect any homology between the 3′-terminal regions of LIYV RNAs 1 and 2.

Together, these nucleotide sequence data indicate that LIYV RNAs 1 and 2 are 8118 and 7193 nucleotides, respectively. The complete nucleotide sequences of both RNAs have been deposited with the Genbank Database under Accession No. U15440 for LIYV RNA 1 and U15441 for LIYV RNA 2 and are not duplicated here.

LIYV RNA 1 — the principal replicative domains, leader protease, and putative ribosomal frameshift

LIYV RNA 1 includes a proposed 97-nucleotide 5′ untranslated region followed by two ORFs, 1a and 1b (Fig. 1A). The proposed methionine start codon for ORF 1a, starting at position 98, is the 5′-most AUG and is in the most favorable context for translational initiation (Kozak, 1986). ORF 1a encodes a 1873-amino-acid product with a predicted M_r of 217,254. ORF 1b, starting at position 5808, encodes a 515-amino-acid product with a predicted

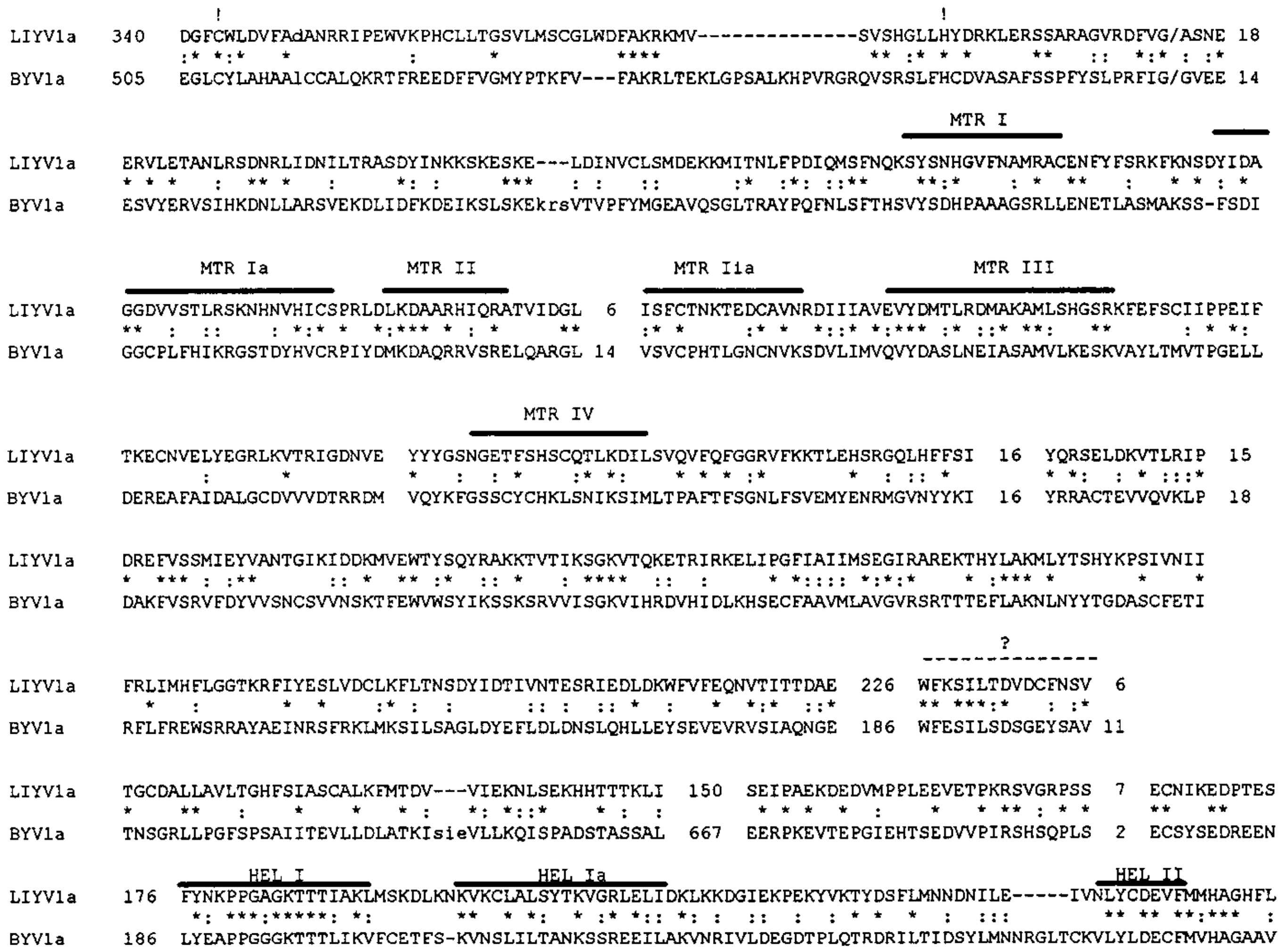

FIG. 2. Alignment of the putative 1a/b proteins of lettuce infectious yellows closterovirus (LIYV) and beet yellows closterovirus (BYV). The alignment was constructed using the MACAW program. The intervals where the alignment was uncertain were replaced by the number of amino acid residues. Asterisks show identical residues and colons show similar residues. Exclamation marks indicate the predicted catalytic residues of the leader papain-like protease; slashes indicate the predicted cleavage sites. The conserved motifs of the MTR domain (Rozanov *et al.*, 1992), the HEL domain (Gorbalenya *et al.*, 1988; Hodgman, 1988), and the RdRp domain (Koonin, 1991a) are overlined and designated as in the original references. Question marks indicate the putative catalytic residue of an aspartic protease. The frameshift regions are indicated.

M_r of 55,486. A database search using the deduced amino acid sequences of ORFs 1a and b revealed a highly significant similarity to the nonstructural proteins of other closteroviruses (BLAST scores over 100 for ORF 1a and over 200 for ORF 1b, with the probability of occurring by chance below 10^{-20} in each case). Lower, but also statistically significant, similarity was detected with the nonstructural proteins of tobamoviruses, tobraviruses, hordeiviruses, furoviruses, and tricornaviruses.

Alignment of the amino acid sequence of LIYV ORF 1a with the homologous sequence of BYV indicated that the highest similarity occurred in the methyltransferase (MTR) and RNA helicase (HEL) motifs (Fig. 2). Twenty-four percent identity was found in a 267-amino-acid residue alignment of the MTR domain; a 299-residue alignment of the HEL domain showed 30% identity. Alignment of LIYV ORF 1b with the homologous sequence of BYV showed 36% identity in 324 residues. LIYV ORF 1b contains the eight RNA-dependent RNA polymerase (RdRp) motifs (Koonin, 1991a).

Comparison of the ORF 1a region of LIYV and BYV upstream of the MTR domain did not reveal statistically significant similarity. Nevertheless, both contained the motif GxC[WY]U (Fig. 2; x — any amino acid residue; U — a hydrophobic amino acid residue; the residues shown in brackets are alternates), which is a characteristic signature of papain-like proteases and the helper-component proteases of potyviruses (Gorbalenya *et al.*, 1991; Koonin and Dolja, 1993; Agranovsky *et al.*, 1994). Site-directed mutagenesis of BYV indicated that the catalytic cysteine residue of the leader protease is contained in this motif (Agranovsky *et al.*, 1994). We suggest the same role for the homologous cysteine residue of LIYV. Downstream of the putative catalytic cysteine residue, a counterpart to the predicted catalytic histidine residue of the BYV leader protease could be identified in the LIYV se-

```
                       HEL III                                        HEL IV
LIYV1a     TLLTKIAYQNGYCYGDVNQIPFINRDPYTPAYLSREFFRKQDLN--YDTYTYRCPLDTCYLLSNLKDEMGNIIYAGGVKNVNEVYPTIRSLNLFGINVVG
             :      :     :**  ** :           * **        :*    *   :**** * *  **               *   *   * *      *:::  *: *
BYV1a      ACIEFTKCDSAILFGDSRQIRYGRCSELDTAVLSDLNRFVDDESRVYGEVSYRCPWDVCAWLSTF--------YPKTVATTNLVSAGQSSMQVREIESVD

                                                  HEL V                              HEL VI
LIYV1a     EVPVEYNAKYLTFTQDEKLNLQRHIDSQGGC---RNAVSTVNEAQGCTFSEVNLVRLVQFDNPVMSDINQFVVAISRHTTTFKYFTPHSRLNDRVSNAIS
           :*    :  **   * ** :* :              :  * ******* *:  *****     ::        *   **:***      *     *: :* :::**
BYV1a      DVEYSSEFVYLTMLQSEKKDLLKSFGKRSRSSVEKPTVLTVHEAQGETYRKVNLVRTKFQEDDPFRSENHITVALSRHVESLTYSVLSSKRDDAIAQAIV

                 * frameshift
LIYV1a/b   SLQSVSDFVLKTIILDSAFDVYTLNLT    8  CRPPAARYESIIDLMDIISPGVAFYNY  36  VSKILGKGERNRPNTWKQVIQSLSHRNFN  43
              : *     :  ::*    **:::        *:  :* ** *    :: : **     ::       ** * :       *   :   :  : *  **:*
BYV1a      KAKQLVDAYR---VYPTSFGGSTLDVS   10  CKASSAPYEVINSFLESVVPGTTSVDF  37  VSSIRSQAIPKRKPSLQENLYSYESRNYN  45
                                                                            frameshift

                                          RdRp I                       RdRp II                     RdRp III
LIYV1b     IGKWLDDRDGSKYRMLNRRLDFA-SLADKFKTLNLMVKGETKPKMDLSTYDSYNAPANIVYYQQIVNLYFSPIFLECFARLTYCLSDKIVLYSGMNTDVL
           :  *:  *: *  *  *    *:     *      *  ****   * *:* *             **::    **  **  * *   *:  *     ** :: *    *
BYV1b      VRTWMSKREPSQLRALSSDLQKPLNLEEEITTFKLMVKRDAKVKLDSSCLVKHPPAQNIMFHRKAVNAIFSPCFDEFKNRVITCTNSNIVFFTEMTNSTL

                               RdRp IV                                                         RdRp V
LIYV1b     AELIESKLPLGLNAYHTLEIDFSKFDKSQGTCFKLYEEMMYKMFGFSPELYDRDFKYTEYFCRAKA-TCGVDLELGTQRRTGSPNTWLSNTLVTLGMMLS
           * :      *      *   ***********     * :*   *  ***   :* *        **    *        *     :  **::*  ***::*:: ***::
BYV1b      ASIAKEMLGSE-HVYNVGEIDFSKFDKSQDAFIKSFERTLYSAFGFDEDLLDVWMQ-GEYTSNATTLDGQLSFSVDNQRKSGASNTWIGNSIETLGILSM

                RdRp VI                                      RdRp VII       RdRp VIII
LIYV1b     SYDIDDIDLLLVSGDDSLIFSRKHLPNKTQEINKNFGMEAKYIEKSSPYFCSKFIVELNGKLKVIPDPIRFFEKLSIPIRQEDFVNGSVVKERFISFKDL
            *  :     * **********     : *   :  :   : * * *::  * ******* :        :  :*** :    **:           * *:    : * * **:**
BYV1b      FYYTNRFKALFVSGDDSLIFSESPIRNSADAMCTELGFETKFLTPSVPYFCSKFFVMTGHDVFFVPDPYKLLVKLGA---SKDEVDDEFLFEVFTSFRDL

LIYV1b     MKEYDNDVAVIRIDEAVCYRYSIPVGCSYAALCYIHCCMSNFVSFRRIY  8
            *   ::  :  :    *   :*:    * :***** ***   *** **:::*
BYV1b      TKDLVDERVIELLTHLVHSKYGYESGDTYAALCAIHCIRSNFSSFKKLY 47
```

FIG. 2–*Continued*

quence (Fig. 2). The BYV papain-like protease has been shown to cleave the BYV ORF 1a polyprotein between residues Gly_{588} and Gly_{589} to generate a 66,000 M_r N-terminal peptide (Agranovsky *et al.*, 1994). This cleavage site aligns with a glycine–alanine dipeptide in LIYV (Fig. 2). Cleavage at this site would result in a leader protein of 412 amino acid residues (45,320 M_r). Despite the relatively low amino acid sequence conservation in the protease domains, the distance between the catalytic histidine and the cleavage site was identical in BYV and LIYV (Fig. 2); this suggests mechanistic similarities in the autocleavage reactions. Upstream of the catalytic cysteine, the leader proteins of LIYV and BYV showed very limited sequence similarity.

```
(a)

1871    S  D  F  V  L  K  D  Y  H  F  R  Q  C  L  *
5664   AGUGAUUUUGUUUUGAAAGACUAUCAUUUUAGACAGUGCCUUUGA
                      *  K  U  I  I  L  D  S  A  F  D

(b)
                K(1a)                          K(1b)
LIYV RNA   5'..AAAG..3'   slippage        5'..AAAG..3'
                 |||      ---------->          ||•
tRNALys    3'..UUU...5'                   3'...UUU..5'
```

FIG. 3. The putative +1 frameshift region in LIYV RNA 1. (a) Overlap between ORF 1a and ORF 1b. Termination codons are indicated by asterisks and a tandem direct repeat is indicated by arrows. (b) Putative tRNA slippage on two overlapping lysine codons.

The region between the MTR and the HEL domains of ORF 1a showed very limited sequence similarity between BYV and LIYV (Fig. 2). However, LIYV ORF 1a includes a possible counterpart to the predicted retroviral aspartic protease found in BYV ORF 1a (Agranovsky *et al.*, 1994). The respective portions were located approximately the same distance downstream of the MTR domain, and the predicted catalytic aspartic acid was conserved (Fig. 2). However, the significance of this conservation remains uncertain as the similarity between the LIYV and BYV sequences in this region was low. Moreover, there were deviations from the aspartic protease consensus in the LIYV sequence.

LIYV ORF 1b encodes a putative RdRp (Fig. 1A). Agranovsky *et al.* (1994) and Dolja *et al.* (1994) suggest that the RdRp domains of BYV and CTV are expressed directly from the genomic RNA via +1 ribosomal frameshifting. To determine if LIYV RNA 1 might contain a frameshift region analogous to the one in BYV, we compared the aligned amino acid sequences of LIYV and BYV around

```
                 transmembrane
BYV  6K  15  LICLFLFCLVVFIWFVYKQILFRTTAQSNEARHNHS  3
BYSV 6K  15  VICFFIAVAAYFFAFFVKNT--HSQDTDVDIRQEDL  2
CTV  6K  16  VFCAFA-GLIIIVITIYRCT--IKPVRSASPYGTHA  2
LIYV 5K   3  LLFFLMSILVWFIFTILKLLFVNTDSEVNIPNKSRF  0
consensus    U&hh&h..hhh&&h.&&J..................
```

FIG. 4. Alignment of the small hydrophobic proteins of closteroviruses. Within the consensus, U indicates a bulky aliphatic residue (I, L, V, or M); & indicates any bulky hydrophobic residue (I, L, V, M, F, Y, or W); h indicates any hydrophobic residue (I, L, V, M, F, Y, W, C, or A); J indicates a positively charged residue (K or R); and dots indicate any residue. The predicted, conserved transmembrane helix is shown. The distances from the protein termini are indicated by numbers.

the overlap of ORFs 1a and 1b (Fig. 2). This showed that the C-terminal portion of BYV ORF 1a aligned with the N-terminal portion of LIYV ORF 1b; the frameshift sites in these two viruses are not homologous. Analysis of the nucleotide sequence in the overlapping region of LIYV ORFs 1a and 1b did not reveal any "slippery" heptanucleotide or stable secondary structure which might be indicative of frameshifting (for a review, see Atkins *et al.*, 1990). However, mechanisms have been described for +1 frameshifting in retrotransposons which appear to be much simpler and need not be associated with distinct structural features (Farabaugh *et al.*, 1993). In addition, potential secondary structure similar to that in −1 frameshifting is not conserved in CTV (Dolja *et al.*, 1994; A. V. Karasev, personal communication). A mechanism for the expression of LIYV ORF 1b might be slippage by $tRNA_{Lys}$ on the sequence AAAG, resulting in a +1 frameshift (Fig. 3).

The putative product of the 3′-terminal ORF of LIYV RNA 1 (ORF 2, Fig. 1A), starting at position 7075, shows no similarity to any proteins in the current databases. This ORF is analogous, with respect to size and location, to ORF 2 of CTV and beet yellow stunt closterovirus (BYSV) (Karasev *et al.*, 1994; Dolja *et al.*, 1994). However, we were unable to detect any conservation between the amino acid sequence of the LIYV product and those of CTV and BYSV. LIYV RNA 1 includes a putative 3′ untranslated region of 219 nucleotides. Our analysis indicates that it does not possess extensive secondary structure.

LIYV RNA 2 — a gene array unique for closteroviruses

LIYV RNA 2 includes a proposed 5′ untranslated region of 326 nucleotides which does not show significant similarity to LIYV RNA 1 except for two regions: the first 5 nucleotides (5′-GGUAA-3′) of LIYV RNAs 1 and 2 are identical, and a stretch of 23 nucleotides (5′-UCUUGGAGAAUUUCGAUGGCACU-3′) immediately surrounding the first AUG of RNA 1 is identical to a 23-nucleotide stretch surrounding the AUG starting at position 136 of LIYV RNA 2 (results not shown). The AUG starting at position 136 is in a favorable context for translation but is followed by a stop codon four codons downstream. The occurrence, by chance, of a 23-nucleotide identity in the 5′ untranslated regions of LIYV RNAs 1 and 2 is unlikely (the probability of such an event is about 10^{-5} as computed using the MACAW program). There are several possible explanations for the presence of this identity. It could be important in regulating the translational efficiency of an ORF starting at position 136 and expressed via ribosomal frameshifting. A +1 frameshift immediately downstream of this AUG in LIYV RNA 2 would result in an ORF consisting of 45 codons which could encode a leader peptide with a M_r of 4950. Alternatively, the 23-nucleotide identity could be an evolutionary vestige, once important in regulating the translational efficiency of an ORF that is no longer expressed due to the introduction, by mutation, of a stop codon at position 148. Finally, the existence of the 23-nucleotide identity could be due to recombination between LIYV RNAs 1 and 2.

ORF 1, starting with the AUG at position 327 of LIYV RNA 2, encodes a small protein consisting of 39 amino acids (4650 M_r, Fig. 1B). The deduced amino acid sequence includes a stretch of 17 nonpolar amino acids terminated by a positively charged residue (Fig. 4), which is predicted to form a transmembrane helix. A similar small, hydrophobic protein is encoded in BYV, CTV, and BYSV (Agranovsky *et al.*, 1991a; Karasev *et al.*, 1994).

ORF 2, starting at position 693, encodes a 554-amino-acid protein (62,300 M_r; Fig. 1B) homologous to the HSP70 chaperone proteins. The HSP70 homologue is highly conserved among closteroviruses, with the level of conservation comparable to that found in the RdRp domains of all characterized closteroviruses (Dolja *et al.*, 1994). Inspection of the alignment of four closterovirus HSP70-related proteins (Fig. 5) shows the presence of the conserved motifs that characterize the HSP70 ATPase domain (Bork *et al.*, 1992). This indicates that the closterovirus HSP70 chaperone-like proteins may possess ATPase activity. Additional conservation between the HSP70-related proteins of closteroviruses was observed in the C-terminal regions (Fig. 5). These domains may be involved in protein–protein interactions important for chaperone activity.

ORF 3 starts at position 2449 and encodes a 514-amino-acid protein with a M_r of 59,265 (Fig. 1B). The deduced amino acid sequence does not show any signif-

FIG. 5. Alignment of the HSP70-related proteins of closteroviruses. The complete sequences are shown for lettuce infectious yellows closterovirus (LIYV), beet yellows closterovirus (BYV), and citrus tristeza closterovirus (CTV). For beet yellow stunt closterovirus (BYSV), only the N-terminal sequence is available. The conserved motifs comprising defined parts of the ATP-binding domain in cellular HSP70 proteins are overlined and designated after Bork *et al.* (1992). In the consensus, @ designates an aromatic residue (F, Y, or W), and B designates a negatively charged residue (D or E); the other designations are as in Fig. 4.

```
                   phosphate 1

LIYV  p62   MRDCKVGLDFGTTFSTVSTLVNNSMYVLRLGDSAYIPTCIAITPGGEAIIGGA-AEVLSG
BYV   p65   M--VVFGLDFGTTFSSVCAYVGEELYLFKQRDSAYIPTYVFLHSDTQEVAFGYDAEVLSN
CTV   p65   M--VLLGLDFGTTFSTVAMATPSELVILKQFNSSYIPTCLLLHAEPNSVSYGYDAEYLAA
BYSV  p?    M--VVFGVDFGTTFSSVCVFNSGRLHVFKQQNSAYIPTCLFLYSDTMAMSFGYDAETASL
consensus   M....&GUDFGTTFS$V.......U.U&J..BS.YIPT.U.U......U..G..AE....

LIYV  p62   D-DTPHCFFYDLKRWVGVDDNTFKFAMNKIRPKYVAELVEGEV------YLTGINKGFSI
BYV   p65   DLSVRGGFYRDLKRWIGCDEENYRDYLEKLKPHYKTELLKVAQSSKSTVKLDCYSGTVPQ
CTV   p65   S-GESGSFYKDLKRWVGCTAKNYQTYLHKLSPSYKVIVKEFGTKSVPVPYLSPLNNDLGL
BYSV  p?    DPNVKGGFFRDLKRWVGCDETNIEEYKSKLKPHYSVTLSNFGKGSRKIPTLGSYSGSVQM
consensus   .......F@.DLKRWUG...........KL.P.Y...U............L..&...&....

LIYV  p62   KLSVKQLIKAYIETIVRLLASSYSLRVIDLNQSVPADYKNAQRLAARSVLKALSFPCRRI
BYV   p65   NATLPGLIATFVKALISTASEAFKCQCTGVICSVPANYNCLQRSFTESCVNLSGYPCVYM
CTV   p65   SVALPSLIASYAKSILSDAERVFNVSCTGVICSVPAGYNTLQRAFTQQSISMSGYSCVYI
BYSV  p?    SGSLSGLIALFIQALVKSAAIEFKCECTELIVSVPANCDCMQRLFTENCVNLSGFTCVHM
consensus   ...U..LI..@...UU......@......U..SVPA.....QR......U....@.C..U

             connect 1             phosphate 2

LIYV  p62   INEPSAAAVYCVSRY-PNYNYFLVYDFGGGTFDVSLIGKYKSYVTVIDTEGDSFLGGRDI
BYV   p65   VNEPSAAALSACSRIKGATSPVLVYDFGGGTFDVSVISALNNTFVVRASGGDMNLGGRDI
CTV   p65   INEPSAAAYSTLPKLNSADKYLAVYDFGGGTFDVSIVSVRLPTFAVRSSSGDMNLGGRDI
BYSV  p?    MNEPSAAALSTCGRTDMSARNLLVYDFGGGTFDVSVLSSLNQTFTVRASGGDMNLGGRDV
consensus   UNEPSAAA&....J.......&.VYDFGGGTFDVSUU......&.V..$.GD..LGGRDU

            adenosine

LIYV  p62   DKSIEDYLVGKYNIKKVIPATYLALIKEECNNTNKSI-FTILFDDGSVQVVEFSKSELEK
BYV   p65   DKAFVEHLYNKAQLPVNYKIDISFL-KESLSKKVSFLNFPVVSEQGVRVDVLVNVSELAE
CTV   p65   DKKLSDKIYEMADFVPQKELNVSSL-KEALSLQTDPVKY-TVNHYGMSETVSIDQTVLRE
BYSV  p?    DRAFKAKIYQMANLPFDEEADISSL-KESLSKIDYPITYTVKTKDGESKTVVVSRGLLAE
consensus   DJ.&...U&....&.......&..L.KE........U.&......G....V.&....L..

                                                  connect 2

LIYV  p62   CVRPFVERSIKLINDVVVRNKLTSGV--------IYMVGGSSLLQPVQDMVRSYASTKGL
BYV   p65   VAAPFVERTIKIVKEVYEKYCSSMRLEPNV-KAKLLMVGGSSYLPGLLSRLSSIPFVDE-
CTV   p65   IASVFINRTIDILTQVKVKSSMPESQSL-----KLVVVGGSSYLPGLLDALATVPFVSG-
BYSV  p?    VIVPFVDRTIKVMKRVFELYVKNMNLKAQDAKAALVLVGGSSYLPGLK
consensus   ....FU.R$I.UU..V..................U&UUGGSS&L..U......&......

LIYV  p62   TLVADQDMRSAVSYGCSVLHK-LEVNKEIVYIDCNSHPLSDISFNCDPEPIIRKPMSIPY
BYV   p65   -CLVLPDARAAVAGGCALYSACLRNDSPMLLVDCAAHNLSISSKYCESIVCVPAGSPIPF
CTV   p65   -IVPVEDARTAVATGCALYSECLDGRSKALLIDCITHHLSVTTFSADSVVVAAAGSPIPF
consensus   ..U...D.R.AV..GC.U&...L......U&UDC..H.LS..$..............IP@

LIYV  p62   THTVKMRHDRPLKTI---VNIYEGSNLFMPENDWLISSNINTTDFAKVGEEYSKVYEYDI
BYV   p65   TGVRTVNMTGSNASAVYSAALFEGDFVKCRLNKRIFFGDVVLGNVGVTGSATR-------
CTV   p65   EGEQKLTLRKCVSTSNYQARMFEGDYEKVFRNERIYAASISLFTLGVNWSVPN-------
consensus   .....U..............U@EG.......N..I&...U....&...............

LIYV  p62   DGIITLKIRNEVTGKM-FTLPNSFTKSDNIKPITF   0 KLTQLSNTDDLATLTSLLGYH
BYV   p65   TVPLTLEINVSSVGTISFSLVGPTGVKKLIGGNAA  38 KLTDGDKALFLKRLTADYRRE
CTV   p65   DVEMTLVTKVDSMGKVEFYLKGPSGELVNVQGTSH  41 LRTLFDTLLADLRKTASLGEY
consensus   ...UTL.......G.U.F.L.........U.....     ..T..........T..&...

LIYV  p62   DKNFERFY----GLFNVPTILIKEI---DKLGGFKTLYRRLKSMNANF
BYV   p65   ARKFSSYDD---AVLNSSELLLGRIIPKILRGSRVEKLDV
CTV   p65   SKKYPITRNDIDVVSSRMGIVVSKVL----RGSDLERIPL
consensus   ...@.........U.....UUU..U......G.....&..
```

```
LIYV   p59   56 DYYISGGQIVVSPKDSNAYVKLLIVYLKYCYINYSAKTKYPPQSLLAVLD
BYV    p64   43 DFRFSDGVILSRKTFGESTGESFVREFSL-LLTFPKTYEVCKLCGVAMEL
CTV    p61   38 DHTLSSGVVVRRQSLLNAPQGTFENELAL-LYNSVVINDFVELTGMPLKS
consensus       D..&S.G.UU............&...&.&.&&.......&.....U.U..

LIYV   p59      YDSFKAKWVKYLDKSLTDYLDDNKTEGCSFTEQQVVEKYPQV-DSLVAKI
BYV    p64      ALNGMNRLSDYNVSEFN--IVDVKTVGCKFNIQSVTEFVKKINGNVAEPS
CTV    p61      LMTGIEDRK--VPDELI--SVDPHEVGCRFTLNDVESYLMSRGEDFADLA
consensus       ...............&.....D....GC.F....V...&......&....

LIYV   p59      LYRVCNSL----GKLLDLKDFENKNISGF-EINTAQDSPTVADDNESNDF
BYV    p64      LVEHCWSLSNSCGELINPKDTKRFVSLIFKGKDLAESTDEAIVSSSYLDY
CTV    p61      AVEHSWCLSNSCGRLLSSTEIDAYKTLVF-TKNFDSNVSG--VTTKLETY
consensus       .&.....L....G.LU...B........F....................@

LIYV   p59      FRECVNDQRYYSSLSGSKLGKAKLEANAYIFKILLKS-ASGEFDIDRLSR
BYV    p64      LSHCLN---LYETCNLSS-NSGKKSLYDEFLKHVIDYLENSDLEYRSPSD
CTV    p61      LSYCIS---LYKKHCM-K-DDDYFNLILPMFNCLMKVLASLGLFYEKHAD
consensus       &..CU....&Y...................&&..UU.......&.&.....

LIYV   p59      NPLAISKFMNLYTNHVTDSETFKSKFEALKSIKTPFASFIKKAFGIRLNF
BYV    p64      NPLVAGILYDMCFEYNTLKSTYLKNIESFDCFLSLYLPLLSEVFSMNWER
CTV    p61      NPLLTGMLIEFCLENKVYYSTFKVNLDNVRLFKSKVLPVVLTVWDISEPD
consensus       NPL....&&.&.........T@...&..&..&...&..&U...@.U....

LIYV   p59      E--DSKIFYALPKERQSDVLSDDMMVESIVRDAASFTVVSDNNYLPERVD
BYV    p64      PAPDVRLLFELDAAELLLKVPTINMHDSTFLYKNKLRYL--ESYFEDDSN
CTV    p61      DPMDERVLIPFDPTDFVLDLPKLNIHDTMVVVGNQIRQL--EYVVESDAL
consensus       ...D..U&&.&........U....U.B$.&.....&..U....&&.....

LIYV   p59      RFVTQLLLELFPKTKASFPNKIMF-GFLHYFALSTTNSKRFNDTQESTIE
BYV    p64      ELIKVKVDSLLTRDNPELKLAQRWVGFHCYYGVFRTAQTR-KVKRDAEYK
CTV    p61      DDLSQHVDLRLAADNPDLRVGLRWAGMFVYYGVYRCVVDR-AVERPTLFR
consensus       ..U...U...&......&.....@.G&..Y@.U......R........&.

LIYV   p59      IEGETL----KISLKFITSYLRNAIQSQHPDYADSNIVRLWCNKRSNLAL
BYV    p64      LPPAL-----GEF-VINMSGVEEFFEELQKKMPSISVRRRFCGSLSHEAF
CTV    p61      LPQKLLSQDDGESCSLHMGSVEALFNLVQKVNKDINVRRQFMGRHSEVAL
consensus       U..............&....U...&...........U.R.@....S..A&

LIYV   p59      GYFKSRNIQLYLYS--KYPRLLNYMRFDYFKGLDMGKLTDEERLSIQTLR   24
BYV    p64      SVFKRFGVGFPPITRLNVPVKYSYLNVDYYRHVKRVGLTQDELTILSNIE   76
CTV    p61      RLYRNLGLRFPPISSVRLPAHHGYLYVDFYKRVPDGAVTADELESLRQLR   63
consensus       .&@J...U.&..&....&P....YU.&D@@J.U....UT.BE...U..U.
```

FIG. 6. Alignment of the ORF 3 product of lettuce infectious yellows closterovirus (LIYV) RNA 2 with the related proteins of beet yellows closterovirus (BYV) and citrus tristeza closterovirus (CTV). The designations are as in Fig. 4.

icant similarities to sequences in the available database. However, direct comparison with the deduced amino acid sequences of CTV p61 (Pappu *et al.*, 1994) and BYV p64 (Agranovsky *et al.*, 1991a) revealed several distinct conserved motifs (Fig. 6). A three-way alignment was statistically significant (probability of matching by chance below 10^{-4}) for some of these motifs. In previous analyses, sequence similarity was observed between regions of the p64 and p61 proteins and another family of chaperone proteins, HSP90 (Koonin *et al.*, 1991b; Karasev *et al.*, 1994). However, these regions were not highly conserved in the LIYV protein.

ORF 4 overlaps ORF 3 by 19 nucleotides and encodes a putative protein of 80 amino acid residues (9000 M_r). The deduced amino acid sequence did not show any significant similarity to other sequences in the current database. Detailed comparisons with other closteroviruses also did not reveal any significant similarities. Unusual segments, such as long runs of nonpolar residues that would be indicative of membrane localization, were not detected.

ORF 5 (249 amino acid residues; 27,800 M_r) starts 4 nucleotides downstream of ORF 4 (Fig. 1B) at position 4221 and encodes the LIYV coat protein (Klaassen *et al.*, 1994). As previously shown (Klaassen *et al.*, 1994), comparison of the deduced amino acid sequence of the LIYV coat protein to other filamentous plant virus coat protein sequences suggests that the LIYV coat protein sequence is most similar to the coat protein sequences of the two closteroviruses, BYV and CTV.

ORF 6 overlaps ORF 5 by 7 nucleotides, starting at position 4964, and encodes a protein of 454 amino acid

```
                                                                !
LIYV CP    69  SAND-VQSFREAMINFMRDK-DPNRNQPSDKLIIAMEVGVYQMVINLGTS-AKL-G-NANNL
BYV  CP    24  HGED-CDKLRKNFEECLKLK-----GVPEDNLGIALGLCLYSCAT-IGTS-NKVNV-QPTS-
CTV  CP    48  STQQ-NAALNRDLFLTLKGK-HP--NLPDKDKDFRIAMMLYRLAV-KSSS-LQSDD-DATGI
LIYV p52c 229  SSDPLTKKQCDQLMLSLIKWFEKFGITKDNARLLIF-----QFGISFSTS--KENLNNITNN
BYV  p24   36  NPNKLNRKETDELLGVIRERFKSELVITDEDFVKHLAFALIRAAN-ITTS-VKVNY--VGAY
CTV  p26   58  STEKFTGEHLKYVMVTMDTFLLENYKTKTEDLLVHLTMIQKRLYT-ISTS-TKTKFRDKGCI
consensus                  &   U                  &            $S
LIYV p52n  21  FKTRITDNFTGDLTLNINTSNLIKFKTCSFFICYGDDKDRYELGW-TSTSTSRSIFQHYKDG

                                                !
LIYV CP    EFTIA-YDQETRTY-KVADFVNYM-QSR-MRNSPNVVRQYARAMEKTINNIRSAGI-INSNGVLAA
BYV  CP    TFIKASFGGGKELYLTHGELNSFLGSQKLLEGKPNKLRCFCRTFQKDYISLRKEYRGKLPPIARAN
CTV  CP    TYTR----EGVEVDLSDKLWTDVVFNSKGIGNRTNALRVWGRTNDALYLAFCRQNR-NLSYGGRPL
LIYV p52c  IVVENDK-GGFV-KILKIDYLNKLYGSIPESHTHNLERVLLRHYAQEILILLRSKV-LEWPRKLAR
BYV  p24   EYTI----GGKKFLVKDAWVFPLIKECMKKFNKPNPVRTFCATFEDAYIVIARSLP-KLFLNRTIG
CTV  p26   SYVQ----GGLRYKLLDKVVFPFIISKFTDRETPNALRKFACTFEELHLCMARLRP-DLYENKRTT
consensus   &                 &   U          N  R &          &
LIYV p52n  KYIRDFRIQDPFPILSGSTFPVVISKIIANRVAFRMSRRLNNVIVDKLKNNIIEFL-FVVYLDVDT

                        !
LIYV CP    DAGIPAGYHYLCADFL--TGAGLTDAQLTSLMLARKQALCK--GEGGSVEHYN  11
BYV  CP    KHGVLASYRNSYSDFAVGFGNDTTDLQQSRLLLARENATHT--EFSSESPVTS  12
CTV  CP    RHGLPAEDHYLAADFI-STSTELTDLECAVYIQAKEQLLKK--RGADDVVVTN  10
LIYV p52c  RAGTPHLKGYLSADFLSGSLPGYTEAETTALTTVKSWTMNH---YKKKRQIVN   5
BYV  p24   NKGIFEQYAYMACDFFDTAELELNDHEKAIVLQASRAAIDRAVSSSVDGKIVS   8
CTV  p26   KRGIPSGYEFLGADFLTATSVCLSEHERGIILRASESMLARRQGYEEATELLN   8
             G          DF         B     &
LIYV p52n  GKIKPNTI-LKNLDLSSLFIVFSNNGNNKINLPYEIELQTK-DRGIVYTKMGN 254
```

FIG. 7. Alignment of the coat proteins of closteroviruses and their diverged copies. CP designates coat protein; p52c indicates the C-terminal portion of the lettuce infectious yellow closterovirus (LIYV) ORF 6 product and p52n indicates the N-terminal portion of the same protein. The consensus shows the amino acid residues conserved in the six aligned sequences. Exclamation marks indicate the amino acid residues universally conserved in the coat proteins of filamentous plant viruses (Dolja et al., 1991).

residues (52,300 M_r; Fig. 1B). Alignment of the amino acid sequence of LIYV ORF 6 with the LIYV, BYV, and CTV coat protein sequences and the diverged copies of the BYV and CTV coat proteins (Boyko *et al.*, 1992; Pappu *et al.*, 1994) indicated that the C-terminal one-half of the LIYV ORF 6 contains the 3 amino acid residues (S, R, and D) invariant in all filamentous virus coat proteins (Fig. 7; Dolja *et al.*, 1991). This suggests that the C-terminal one-half of LIYV ORF 6 contains a diverged copy of the coat protein, analogous to the genomes of BYV and CTV. A more distant similarity to the LIYV coat protein was also observed in the N-terminal part of the LIYV ORF 6 amino acid sequence (Fig. 7). This observation could be explained by two successive duplications of the LIYV coat protein. The order of the genes for the coat protein and its proposed diverged copy are reversed in LIYV when compared to BYV and CTV.

The 3′-terminal ORF (ORF 7, starting at position 6319) in LIYV RNA 2 could encode a protein of 227 amino acid residues (26,000 M_r; Fig. 1B), but it has no obvious counterpart in other closterovirus genomes. It shows no similarity to any sequences in the current databases. LIYV RNA 2 includes 187 nucleotides of 3′ untranslated sequence. Our analysis indicates that this region does not have extensive secondary structure and does not show any homology with the 3′-terminus of LIYV RNA 1. The absence of a poly(A) tract at the 3′-termini of LIYV RNAs 1 and 2 (Klaassen *et al.*, 1994) is consistent with the 3′-terminal structures of BYV (Karasev *et al.*, 1989) and CTV (German *et al.*, 1990).

Expression of internal ORFs via subgenomic RNAs

We have previously detected, by Northern blot hybridization analysis, multiple RNAs in dsRNAs and total ssRNAs extracted from LIYV-infected plants (Klaassen *et al.*, 1994). These data, together with the organization of the LIYV genome, suggest that the LIYV internal ORFs are expressed via subgenomic messenger RNAs (sgRNAs). To identify any putative LIYV sgRNAs, we hybridized denatured LIYV virion RNAs, dsRNAs, and total ssRNAs to negative-sense, ^{32}P-labeled SP6 transcripts generated from cDNA clones representing different regions of LIYV RNAs 1 and 2 (Figs. 1 and 8).

Transcripts of pSP10, complementary to a 1500-nt region upstream from ORF 2 of LIYV RNA 1 (Fig. 1A), hybridized to LIYV RNA 1 but not to any smaller subgenomic-like RNAs or to any RNAs in healthy samples (Fig. 8a, lanes 1–5). Transcripts of pSP182, complementary to the 3′-terminal region of LIYV RNA 1 (Fig. 1A; Fig. 8a, lanes 6–10), also hybridized to LIYV RNA 1 (Fig. 8a, lanes 6, 8, and 10). In addition, a single smaller RNA of approxi-

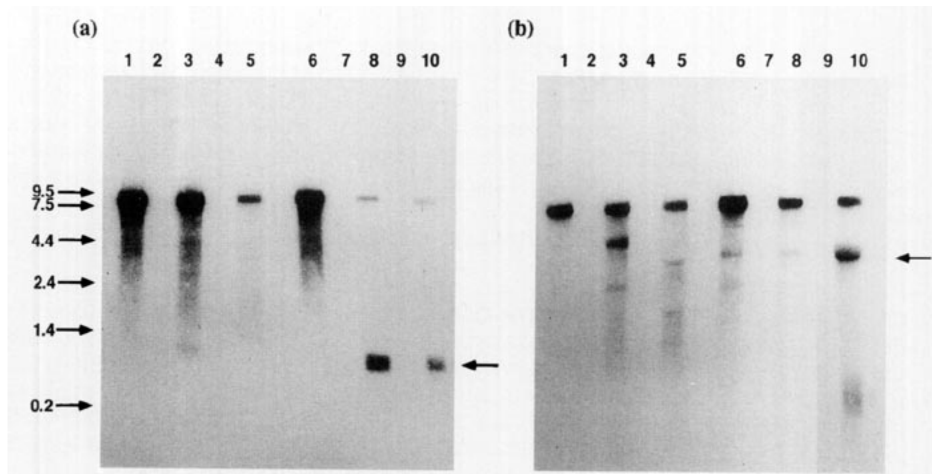

FIG. 8. Northern blot hybridization analysis of denatured lettuce infectious yellows closterovirus (LIYV) virion RNAs and dsRNAs and total ssRNAs from LIYV-infected and healthy plants. Negative-sense probes were generated by *in vitro* transcription using SP6 RNA polymerase and linearized cDNA clones representing (a) LIYV RNA 1 (pSP10, lanes 1–5; pSP182; lanes 6–10) and (b) LIYV RNA 2 (pSP144; lanes 1–5; pSP105, lanes 6–10). Lanes 1 and 6, LIYV virion RNAs; lanes 2 and 7, dsRNAs from healthy tissue; lanes 3 and 8, dsRNAs from LIYV-infected tissue; lanes 4 and 9, total ssRNAs from healthy tissue; and lanes 5 and 10, total ssRNAs from LIYV-infected tissue. The arrows to the far left indicate positions and sizes of marker RNAs (BRL) in Kb. The arrow to the right of (a) points to the putative subgenomic RNA for LIYV RNA 1. The arrow to the right of (b) points to the putative subgenomic RNA for LIYV RNA 2.

mately 1000 nt was detected in dsRNAs and total ssRNAs from LIYV-infected plants (Fig. 8a, lanes 8 and 10).

Transcripts of pSP144, complementary to the 5′-terminal 1400 nt of LIYV RNA 2 (Fig. 1B; Fig. 8b, lanes 1–5), hybridized to LIYV RNA 2 (Fig. 8b, lanes 1, 3, and 5) and two smaller RNAs of approximately 4200 and 2400 nucleotides in LIYV dsRNAs (Fig. 8b, lane 3). These RNAs were not present in healthy dsRNA preparations (Fig. 8b, lane 2) or total ssRNAs from healthy or LIYV-infected plants (Fig. 8b, lanes 4 and 5), and therefore appeared to be LIYV-specific dsRNAs. To further characterize these two RNAs, we used transcripts of pSP222 complementary to the 5′-terminal region of ORF 3 (Fig. 1B). Only a 4200-nucleotide RNA was detected (results not shown), indicating that the 2400-nucleotide RNA terminates upstream of ORF 3. Transcripts of pSP105, complementary to the 3′-terminal region of LIYV RNA 2 (Fig. 1B; Fig. 8b, lanes 6–10), did not hybridize to a 4200- or a 2400-nucleotide RNA in LIYV dsRNA (Fig. 8b, lane 8). These hybridization results indicate that the 4200- and 2400-nucleotide RNAs represent the 5′-terminal region of LIYV RNA 2 and are specific to LIYV dsRNAs.

When transcripts of pSP144 were hybridized to LIYV total ssRNAs, two smaller RNAs were detected (Fig. 8b, lane 5). However, these RNAs did not have the same mobility as the two RNAs specific to LIYV dsRNAs (Fig. 8b, lane 3). To determine the origin of these RNAs, we added LIYV virion RNAs to total ssRNAs from healthy plants. Two RNAs with mobility identical to that in LIYV total ssRNAs were detected by transcripts of pSP144 (results not shown). This suggests that these signals are artifacts due to trapping by the large amount of ribosomal RNA present in these extracts (Dougherty, 1983).

A subgenomic-like RNA of approximately 3000 nucleotides was detected by transcripts of pSP105 (Fig. 8b, lanes 6, 8, and 10). Unlike the other smaller RNAs, the 3000-nucleotide RNA was detected in LIYV virion RNA, in addition to LIYV dsRNAs and total ssRNAs.

The above hybridization results suggest that LIYV utilizes sgRNAs for expressing internal ORFs, an expression strategy similar to that of BYV (Dolja *et al.*, 1990) and CTV (Hilf *et al.*, 1994). The 1000-nucleotide RNA detected for LIYV RNA 1 (Fig. 8a, lanes 8 and 10) maps to the 3′-terminal region and is the correct size to serve as a sgRNA for ORF 2.

In contrast to the results for LIYV RNA 1, the hybridization patterns for LIYV RNA 2 suggested the presence of multiple smaller RNAs in LIYV dsRNAs and total ssRNAs. However, the two smaller RNAs present in LIYV dsRNA preparations mapped to the 5′-terminal region. Double-stranded RNAs representing the 5′-terminal region of apple chlorotic leafspot carlavirus genomic RNA have been observed (German *et al.*, 1992). German *et al.* (1992) suggest that these dsRNAs could arise by premature termination during replication due to the synthesis of sgRNAs via internal initiation. This model could explain the presence of the 4200- and 2400-nucleotide RNAs in LIYV dsRNAs (Fig. 8b, lane 3) if synthesis of sgRNAs for ORF 3 and ORF 4 or 5 caused premature termination of LIYV RNA 2 during replication.

One smaller RNA was detected for LIYV RNA 2 which mapped to the 3′-terminal region. This 3000-nucleotide RNA (Fig. 8b, lanes 6, 8, and 10) is the correct size to serve as a sgRNA for the LIYV coat protein. The 3000-nucleotide RNA was not always detected in purified LIYV virion RNAs (results not shown). However, we cannot exclude the possibility that this RNA may be encapsidated.

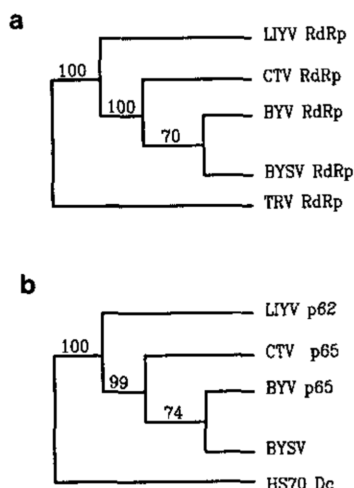

FIG. 9. Tentative phylogenetic trees for the RdRp- and the HSP70-related proteins of closteroviruses. The RdRp tree (a) was constructed from an alignment of 324 amino acid residues (with positions containing gaps excluded) using the protein parsimony algorithm. For each internal node, the number of bootstrap runs in which the given node appeared is indicated. The homologous sequence of tobacco rattle virus (TRV) was used as the outgroup. The N-terminal portion of the beet yellow stunt closterovirus (BYSV) sequence was provided by personal communication from Dr. A. V. Karasev. The HSP70 tree (b) was generated from an alignment of 236 amino acid residues in the N-terminal part of the protein (the C-terminal sequence is unavailable for BYSV). The HS70 sequence from carrot (*Daucus carota,* Dc) was used as the outgroup. The branch lengths are arbitrary. The use of the neighbor joining algorithm resulted in the same topology for both trees.

If it is assumed that the internal ORFs of LIYV RNA 2 are expressed via sgRNAs, one would expect to find a set of nested 3′-coterminal RNAs in total ssRNA and dsRNA extracts. However, we did not detect any other RNAs which mapped to the 3′-terminal region of RNA 2. One explanation is that these RNAs may be present in levels too low to detect using our methods. In addition, sgRNAs may be obscured by the ribosomal RNA present in these extracts.

Evolution of the bipartite closterovirus genome

The LIYV genome organization described above has a number of similarities to the monopartite closteroviruses, including the proposed use of a +1 frameshift to express the RdRp, the presence of a leader papain-like protease, and the distinctive gene array in RNA 2. On the other hand, the differences are also notable: the division of the genome into two RNA molecules, the reversal of the order of the genes for the coat protein and its putative diverged copy, the different size of the putative diverged copy (possibly due to an additional duplication), and two unique ORFs in RNA 2. What is the relationship between these features of the genome organization and the phylogeny of closteroviruses as derived from comparative analysis of conserved domains? To address this question, we constructed tentative phylogenetic trees for the RdRp- and the HSP70-related proteins using the appropriate outgroups and the bootstrap test to assess the statistical significance of the tree topology. Both trees placed the four closteroviruses in the same topology (Fig. 9); this supports the close evolutionary relationship between LIYV, BYV, and CTV that is indicated by the similarities in their genome organizations. The phylogenetic analysis also suggests that the split of the ancestral monopartite genome (Koonin and Dolja, 1993; Dolja *et al.,* 1994), leading to the LIYV bipartite genome, occurred early in the evolution of the closteroviruses, i.e., before the radiation of different monopartite closteroviruses.

ACKNOWLEDGMENTS

The authors thank Valerian Dolja, Elizabeth Estabrook, and Boni Passmore for their support and helpful discussions. Thanks are also due to Jim Duffus and Rebecca Creamer for providing some of the LIYV isolates. This work was supported in part by the California Iceberg Lettuce Research Program and The Asgrow Seed Company.

REFERENCES

Agranovsky, A. A., Boyko, V. P., Karasev, A. V., Lunina, N. A., Koonin, E. V., and Dolja, V. V. (1991a). Nucleotide sequence of the 3′-terminal half of beet yellows closterovirus RNA genome: Unique arrangement of eight virus genes. *J. Gen. Virol.* **2,** 15–23.

Agranovsky, A. A., Boyko, V. P., Karasev, A. V., Koonin, E. V., and Dolja, V. V. (1991b). Putative 65 kDa protein of beet yellows closterovirus is a homologue of HSP70 heat shock proteins. *J. Mol. Biol.* **217,** 603–610.

Agranovsky, A. A., Koonin, E. V., Boyko, V. P., Maiss, E., Frotschl, R., Lunini, N. A., and Atabekov, J. G. (1994). Beet yellows closterovirus: Complete genome structure and identification of a leader papain-like thiol protease. *Virology* **198,** 311–324.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

Altschul, S. F., Boguski, M. S., Gish, W., and Wootton, J. C. (1994). Issues in searching molecular sequence databases. *Nat. Genet.* **6,** 119–129.

Amasino, R. M. (1986). Acceleration of nucleic acid hybridization rate by polyethylene glycol. *Anal. Biochem.* **152,** 304–307.

Atkins, J. F., Weiss, R. B., and Gesteland, R. F. (1990). Ribosome gymnastics — Degree of difficulty 9.5, style 10.0. *Cell* **62,** 413–423.

Bork, P., Sander, C., and Valencia, A. (1992). An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin and HSP70 heat shock proteins. *Proc. Natl. Acad. Sci. USA* **89,** 7290–7294.

Boyko, V. P., Karasev, A. V., Agranovsky, A. A., Koonin, E. V., and Dolja, V. V. (1992). Coat protein gene duplication in a filamentous RNA virus of plants. *Proc. Natl. Acad. Sci. USA* **89,** 9156–9160.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12,** 387–395.

Dolja, V. V., Karasev, A. V., and Agranovsky, A. A. (1990). Organization of the beet yellows closterovirus genome. *In* "New Aspects of Positive Strand RNA Viruses" (R. Rueckert and M. Brington, Eds.), pp. 31–35. Am. Soc. Microbiol., Washington, DC.

Dolja, V. V., Boyko, V. P., Agranovsky, A. A., and Koonin, E. V. (1991). Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: Two families with distinct patterns of sequence and probably structure conservation. *Virology* **184,** 79–86.

Dolja, V. V., Karasev, A. V., and Koonin, E. V. (1994). Molecular biology and evolution of closteroviruses: Sophisticated build-up of large RNA genomes. *Annu. Rev. Phytopathol.* **32,** 261–285.

Dougherty, W. G. (1983). Analysis of viral RNA isolated from tobacco leaf tissue infected with tobacco etch virus. *Virology* **131,** 473–481.

Duffus, J. E., Larsen, R. C., and Liu, H. Y. (1986). Lettuce infectious yellows virus — A new type of whitefly-transmitted virus. *Phytopathology* **76**, 97–100.

Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125–142.

Farabaugh, P. J., Zhao, H., and Viamaladithan, A. (1993). A novel programmed frameshift expresses the POL3 gene of retrotransposon Ty3 of yeast: Frameshifting without tRNA slippage. *Cell* **74**, 93–103.

Felsenstein, J. (1989). PHYLIP — Phylogeny inference package (Version 3.2). *Cladistics* **5**, 164–166.

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. *Arch. Virol.* (Suppl. 2).

German, S., Candresse, T., Lanneau, M., Huet, J. C., Pernollet, J. C., and Dunez, J. (1990). Nucleotide sequence and genomic organization of apple chlorotic leaf spot closterovirus. *Virology* **179**, 104–112.

German, S., Candresse, T., Le Gall, O., Lanneau, M., and Dunez, J. (1992). Analysis of the dsRNAs of apple chlorotic leafspot virus. *J. Gen. Virol.* **73**, 767–773.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M. (1988). A novel superfamily of nucleoside triphosphate-binding motif-containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett.* **239**, 16–24.

Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. C. (1991). Putative papain-related proteases of positive-strand viruses. *FEBS Lett.* **288**, 201–205.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.

Hilf, M. E., Karasev, A. V., Pappu, H., Gumpf, D., Niblett, C., and Garnsey, S. M. (1994). Northern blot analysis of citrus tristeza virus subgenomic RNAs. (Abstr.) Am. Soc. Virol. 13th Annual Meeting, P1-17.

Hodgman, T. C. (1988). A new superfamily of replicative proteins. *Nature* **333**, 22–23.

Huiet, L., Klaassen, V., Tsai, J. H., and Falk, B. W. (1991). Nucleotide sequence and RNA hybridization analyses reveal an ambisense coding strategy for maize stripe virus RNA 2. *Virology* **182**, 47–53.

Huiet, L., Tsai, J. H., and Falk, B. W. (1992). Complete sequence of maize stripe virus RNA 4 and mapping of its subgenomic RNAs. *J. Gen. Virol.* **73**, 1603–1607.

Huiet, L., Feldstein, P. A., Tsai, J. H., and Falk, B. W. (1993). The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. *Virology* **197**, 808–812.

Karasev, A. V., Agranovsky, A. A., Rogov, V. V., Miroshnichenko, N. A., Dolja, V. V., and Atabekov, J. G. (1989). Virion RNA of beet yellows closterovirus: Cell-free translation and some properties. *J. Gen. Virol.* **70**, 241–245.

Karasev, A. V., Nikolaeva, O. V., Koonin, E. V., Gumpf, D. J., and Garnsey, S. M. (1994). Screening of the closterovirus genome by degenerate primer-mediated polymerase chain reaction. *J. Gen. Virol.* **75**, 1415–1422.

Klaassen, V. A., Boeshore, M. L., Dolja, V. V., and Falk, B. W. (1994). Partial characterization of the lettuce infectious yellows virus genomic RNAs, identification of the coat protein gene, and comparison of its amino acid sequence with those of other filamentous RNA plant viruses. *J. Gen. Virol.* **75**, 1525–1533.

Koonin, E. V. (1991a). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* **72**, 2197–2207.

Koonin, E. V., Mushegian, A. R., Tyabov, E. V., and Dolja, V. V. (1991b). Diverse groups of plant RNA and DNA viruses share related movement proteins that may possess chaperone-like activity. *J. Gen. Virol.* **72**, 2895–2903.

Koonin, E. V., and Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* **28**, 375–430.

Kozak, M. (1986). Regulation of protein synthesis in virus-infected animal cells. *Adv. Virus Res.* **31**, 229–292.

Moore, J., Engelberg, A., and Bairoch, A. (1988). Using PC/GENE for protein and nucleic acid sequence analysis. *BioTechniques* **6**, 566–572.

Pappu, H. R., Karasev, A. V., Anderson, E. J., Pappu, S. S., Hilf, M. E., Febres, V. J., Eckloff, R. M. G., McCaffery, M., Boyko, V., Gowda, S., Dolja, V. V., Koonin, E. V., Gumpf, D. J., Cline, K. C., Garnsey, S. M., Dawson, W. O., Lee, R. F., and Niblett, C. L. (1994). Nucleotide sequence and organization of eight 3′-open reading frames of the citrus tristeza closterovirus genome. *Virology* **199**, 35–46.

Rao, M. J. K., and Argos, P. (1986). A conformation preference parameter to predict helices in integral membrane proteins. *Biochim. Biophys. Acta* **869**, 197–214.

Rozanov, M. N., Koonin, E. V., and Gorbalenya, A. E. (1992). Conservation of the putative methyltransferase domain: A hallmark of the "Sindbis-like" supergroup of positive-strand RNA viruses. *J. Gen. Virol.* **73**, 2129–2134.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erich, H. A. (1988). Primer-directed enzyme amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

Schuler, G. D., Altschul, S. F., and Lipman, D. J. (1991). A workbench for multiple alignment construction and analysis. *Proteins: Struct. Funct. Genet.* **9**, 180–190.

Sekiya, M. E., Lawrence, S. D., McCaffery, M., and Cline, K. (1991). Molecular cloning and nucleotide sequencing of the coat protein gene of citrus tristeza virus. *J. Gen. Virol.* **72**, 1013–1020.

Zuker, M. (1989). Computer prediction of RNA structure. *Methods Enzymol.* **180**, 262–288.