Journal of General Virology (1997), **78**, 2067–2071. Printed in Great Britain SHORT COMMUNICATION

Complete genome structure and phylogenetic analysis of little cherry virus, a mealybug-transmissible closterovirus

Wilhelm Jelkmann,[1] Beate Fechtner[1] and Alexey A. Agranovsky[2]

[1] Federal Biological Research Centre for Agriculture and Forestry, Institute for Plant Protection in Fruit Crops, Schwabenheimer Str. 101, D-69221 Dossenheim, Federal Republic of Germany

[2] Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119899 Moscow, Russia

The 5′-terminal genomic region (8597 nt) of little cherry virus (LChV), a mealybug-borne closterovirus, was cloned from double-stranded RNA, and its sequence determined to complete the 16934 nt sequence of the monopartite LChV RNA genome. In the 5′ to 3′ direction, the sequence encompasses ORF 1a, encoding the conserved replicative domains of methyltransferase and helicase, and ORF 1b, encoding RNA polymerase. ORFs 1a and 1b partially overlap (in 0/ 1 configuration), and the LChV replicase is probably expressed by ribosomal frameshifting as a fusion product with a molecular mass of 318 kDa. The N-terminal part of the ORF 1a product contains a papain-like cysteine proteinase (PCP) domain with a predicted cleavage site between Gly-619 and Ser-620. The PCP and the upstream protein domains can be aligned with the equivalent parts of the leader proteins encoded by the whitefly-transmitted lettuce infectious yellows and sweet potato sunken vein closteroviruses. Phylogenetic reconstruction based on the aligned RNA polymerase sequences clearly suggests that the aphid-transmissible and whitefly-transmissible closteroviruses represent two distinct evolutionary lineages, with the mealybug-transmissible LChV being the most remote member of the 'whitefly' lineage.

Little cherry disease (LCD) is distributed worldwide in ornamental and sweet cherry and has a great impact on the fruit quality of infected trees. In infected phloem cells, characteristic vesicles and elongated virus-like particles were detected by electron microscopy, thus indicating the association of LCD with a clostero-like virus (Raine *et al.*, 1975, 1979; Ragetli *et al.*, 1982). This hypothesis was supported by the extraction of a high molecular mass double-stranded RNA (dsRNA) from LCD-affected sweet cherry (Hamilton *et al.*, 1980; Legrand & Verhoyen, 1986; Jelkmann, 1995; Eastwell & Bernardy, 1996). Recently, the sequence of the 3′-terminal 8337 nucleotides of the dsRNA replicative form associated with a German isolate of little cherry virus (LChV) was obtained and shown to encompass six open reading frames (ORFs) coding for proteins of 70 kDa (a homologue of HSP70 heat shock proteins), 61 kDa, 46 kDa (coat protein, CP), 76 kDa (diverged CP duplicate), 21 kDa and 27 kDa, which is typical for closteroviruses (Keim-Konrad & Jelkmann, 1996). Here we report the complete 16934 nucleotide sequence of LChV determined from cDNA clones obtained from the dsRNA template.

The field sample UW1, which was the source of the dsRNA used as a template for obtaining the 5′-terminal cDNA clones, had been described previously (Jelkmann, 1995; Keim-Konrad & Jelkmann, 1996). The 5′-terminal sequence was obtained from a total of 39 recombinant plasmids. To avoid errors due to the blunt-end cloning strategy used to obtain the cDNA from dsRNA, all sequences were covered by at least two independent cDNA clones. A few one-base ambiguities detected between individual clones were resolved by analysing at least one additional cDNA clone. As isolate UW1 was no longer available in 1996, the 5′-most sequence of the genome was determined for the two isolates 3/7 'Roth' and UW2. Both extended the sequence of the 5′-proximal cDNA clone (p109) for 17 nt. These were identical in the two isolates. As the anchor ligation method (Keim-Konrad & Jelkmann, 1996) failed to yield PCR products, tailing of the dsRNA with rATP followed by applying a 5′/3′ RACE protocol (Boehringer Mannheim) was used. The 5′-terminal residue determined from the minus strand of LChV dsRNA is a C. Genomic RNAs of other closteroviruses start with an A or G (Agranovsky *et al.*, 1994; Klaassen *et al.*, 1995; Karasev *et al.*, 1995), as is characteristic for the residues next to the 5′-m7G cap in eukaryotic messengers. However, Karasev *et al.* (1995)

Author for correspondence: Wilhelm Jelkmann.
Fax +49 6221 861222. e-mail Wilhelm.Jelkmann@urz.uni-heidelberg.de

The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ nucleotide sequence databases under the accession number Y10237.

0001-4688 © 1997 SGM

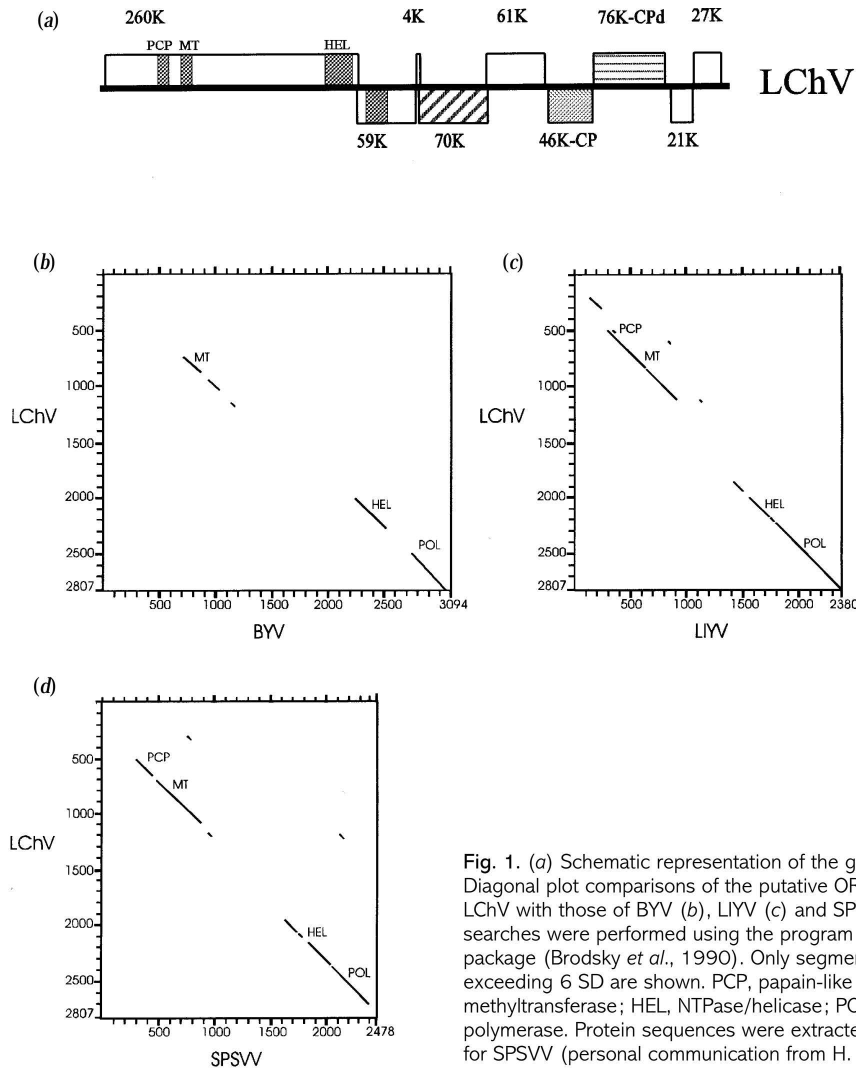

Fig. 1. (*a*) Schematic representation of the genome organization of LChV. Diagonal plot comparisons of the putative ORF 1ab fusion products of LChV with those of BYV (*b*), LIYV (*c*) and SPSVV (*d*). Local similarity searches were performed using the program ALIGN-2 from the GENEBEE package (Brodsky *et al*., 1990). Only segments with similarity levels exceeding 6 SD are shown. PCP, papain-like cysteine proteinase; MT, methyltransferase; HEL, NTPase/helicase; POL, RNA-dependent RNA polymerase. Protein sequences were extracted from the databases, except for SPSVV (personal communication from H. J. Vetten).

reported an extra unpaired G at the 3′ terminus of the citrus tristeza virus (CTV) genomic RNA minus strand. If the same was true for LChV, the 5′-terminal sequence would be GUUUU … instead of CGUUUU …, which is remarkably similar to the 5′ end of the beet yellows virus (BYV) genome.

The 5′-terminal 8597 nucleotides of the LChV genome were determined in this study. Together with the 3′-terminal 8337 nucleotides determined in previous work (Keim-Konrad & Jelkmann, 1996) this completes the sequence of the monopartite RNA genome of LChV, which consists of 16934 nt (EMBL accession no. Y10237). With respect to its monopartite genome size, LChV is intermediate between BYV (15480 nt; Agranovsky *et al*., 1994) and CTV (19296 nt; Karasev *et al*., 1995), and is the second largest plant virus sequenced to date. As the previous electrophoretic size estimations of the LCD-associated dsRNA of ca. 15 kbp (Keim-Konrad & Jelkmann, 1996) and 12·6 kbp (Eastwell & Bernardy, 1996) both differ from the complete LChV sequence determined in this study, we analysed the LChV dsRNA by agarose gel electrophoresis using BYV dsRNA as a size marker. The European and USA LChV isolates used in this and previous studies (Jelkmann, 1995; Keim-Konrad & Jelkmann, 1996) showed a prominent band of high molecular mass dsRNA with an electrophoretic mobility indistinguishable from that of BYV (data not shown), thus suggesting an anomalously fast migration of LChV dsRNA in the gels.

The first AUG in the LChV sequence (starting at position 77) is in an optimal context for translation, with G at the +4 and −3 positions (Kozak, 1986), and is thus likely to be the initiator codon for ORF 1a. This ORF (terminated by the UGA stop codon at position 6981) encodes a polypeptide of 2301 aa (M_r 259330). ORF 1b (nt 6951–8495) overlaps the last 10 triplets of ORF 1a and encodes a polypeptide of 515 aa (M_r 59118) (Fig. 1*a*). ORFs 1a and 1b encode the conserved domains of methyltransferase (MT), RNA helicase (HEL) and RNA-dependent RNA polymerase (POL) (Fig. 1*a*) identified by database screening with FASTA. The LChV ORFs 1a and 1b are found in 0/+1 configuration. A similar layout of the

```
LChV-L     204  GLVCEIYVSENLNKLLIKLVSSFDKNFPELLIAEVNIPKPNNSGKVHPTVNFVIAPSAGVDSKLPRSYYLYNPHGHFESLLSFLII
SPSVV-L    173  GYKLKYFENRDRIDIFFKFGSYVSDSYDFKIN--MQALYSQGRTIRDFRVTAHVCQSKKVTAPFCSSFNLNGRLNYRQFSTKVLIS
LIYV-L     132  GWSVTSYHNDSVSRLHVSMVNNIPNSNTKKLSWWVTRVRSKNPWKRPYKVIYQRDFRDDGIKPYVSEFILYAPKNRADLEMQSILL

cons 1          G&xxxx@xxxxxxxUx&x&xxx&xxxxxxxUxxxUxxxxxxxxxxxxxxVxxxxxxxxxxxxx&xxx@xLxxxxxxxBxxxxxUUx


LChV-L          HNTRPFEDIVDALNLF  79  SAIKFGDFEVFDLREPDNVIKSYAQALTSNLPVDDRILKLNIPVSSEVKTECPSTMNPSNPKSEI
SPSVV-L         SVIGLFGDYVYFTNLY      FGKKFRDFYPSAIKSVTENHKKYCDYVSRLSTRGAKIGTDSTKDCLDKNKTSYKDVVNSPDLTFG
LIYV-L          SSVECFQDIKYALNYW      TNLKFKEICNSALKRFTQTN----IIITPVKPRVVSV-----------NKA--REVSSSVE----

cons 1          xxxxxFxD&xxxxN&@      xxxKFxB&xxxxU+xxxBxxxxxxxxUsxxxxxxxxUxxxxxxxxxxxxxxxxxUxxSxxxxxx


LChV-L          IAKRAVELSSRATSDSSGEIYSKTFNQLLNIPTGGTIIRHFINGVCRSFTLNRSDGRVFNILNNRNAAKIMFNLSLKHKKFYVDGYC
SPSVV-L         GSLKGVKLTHNSTAVPSKPKSNIPLVTFVVDDVGHE-----------SFIIKEKNGFTKKITNDERALVTLFNATLTDKKYNIHPKC
LIYV-L          -----VKLQRY------EREGEV----FVVSKAGKE--------------------DTRVLNDDNAIRNLFNATLNGGKYKLHPEA

cons 1          xxxxxVxLxxxxxxxxxxxxxxxxxx&UxxxxGxxxxxxxxxxxxxxxxxxxxxxxxxUxNBxxAxxxUFNxsLxxxK@xUxxxx

                                 !                                                !
LChV-L          RITPSGKLFSEYDNGFCWLQAFAMFGKIIPTFVEFIPNLNVSCLLQAGLPKSCLKYMHKTGNALYHFDPNKINNIYHANLNFLVG/
SPSVV-L         K-VGKHGRLEDFPNDFCWIDAFACAGKKMPVDLKPYPQISVAYLFRCGLARVIQKHSCETGWRLFHFDRRQVVNLKRLPHGAVLG/
LIYV-L          K-SATGKRFKYYKDGFCWLDVFADANRRIPEWVKPHCLLTGSVLMSCGLWDFAKRKMVSVSHGLLHYDRKLERSSARAGVRDFVG/

cons 1          +xxxxxxx&xx@xBxFCWUBxFAxxx+xUPxxUxxxxxUxxxxL&xxGLxxxxx+xxxxxxxxL&H@Dxxxxxxxxxxxxxx&UG/
cons 2                       BGxCYUxHxxxxxRxxxxxUGx@Pxxxxxxx&xxx&GxxxL+xxURGxxxxFHCxxxxxxxxxS&xxxxxUG/

LChV-1a         SAMNEN
SPSVV-1a        VKGDMN
LIYV-1a         ASNEAV

cons 1          xxxBxx
cons 2          GxxBxx
```

Fig. 2. Alignment of the putative leader proteins of LChV, SPSVV and LIYV (significance score 14 SD, Doolittle distance 214). Non-conserved amino acids are indicated by numbers. Residues identical in two or more sequences are highlighted in bold. Exclamation marks (!) indicate the predicted catalytic Cys and His residues. Consensus 1 shows identical or similar positions in the proteins of LChV, SPSVV and LIYV; consensus 2 shows similar positions in the aligned C-terminal 85 amino acid portions in the proteins of BYV, CTV, LChV, SPSVV and LIYV. Within a consensus, similar amino acids are indicated as follows: U, a bulky aliphatic residue (I, L, V, M); @, an aromatic residue (F, Y, W); &, a bulky hydrophobic residue (aliphatic or aromatic); B, a negatively charged residue (D, N, E, Q); +, a positively charged residue (K, R); s, S or T; x, any residue. A slash (/) denotes a predicted scissible bond. According to the standard code (see Dougherty & Semler, 1993), the residues surrounding the PCP cleavage site are designated as P2, P1/P′1, P′2, P′3, P′4, P′5 and P′6 (VG/SAMNEN in the LChV sequence).

5′-proximal replication-associated genes is seen in the other closterovirus genomes so far sequenced, and it has been suggested that closterovirus replicases are expressed by +1 ribosomal frameshifting (reviewed in Agranovsky, 1996). If frameshift occurs at the ORF 1a stop codon, a 1ab fusion protein consisting of 2807 aa (M_r 317 548) would be produced. Pairwise comparisons of the putative 1ab fusion protein of LChV with those of the aphid-transmissible closteroviruses BYV (Fig. 1*b*) and CTV (not shown) revealed three regions of similarity corresponding to the MT, HEL and POL domains. The similarity regions were more extended and showed higher SD values when the LChV 1ab product was compared with those of the whitefly-transmissible lettuce infectious yellows virus (LIYV) and sweet potato sunken vein virus (SPSVV) (Fig. 1*c*, *d*).

Starting four nucleotides after the UAA stop codon of ORF 1b is ORF2 (nt 8503–8602), which codes for a small hydrophobic protein of 36 aa (4 kDa) containing the N-terminal stretch of non-polar amino acids (data not shown) predicted to form a transmembrane helix in the equivalent proteins of BYV, beet yellow stunt virus (BYSV), CTV and LIYV (Klaassen *et al.*, 1995). ORF 2 overlaps the first 11 nucleotides of ORF 3, which encodes a 70 kDa HSP70 homologue. ORF2 and the downstream ORFs 3, 4, 5, 6, 7 and 8 (nt 8603–10456, 10431–11981, 12064–13275, 13284–15275, 15426–15968 and 16008–16724, respectively, in the complete genome) represent a gene array typical of closteroviruses (Fig. 1*a*).

BYV ORF 1a encodes a papain-like cysteine proteinase (PCP) domain which has been proven to be active in an *in vitro* assay (Agranovsky *et al.*, 1994). Related PCP domains have been predicted in the ORF 1a products of CTV (Karasev *et al.*, 1995), LIYV (Klaassen *et al.*, 1995) and some other closteroviruses (reviewed in Agranovsky, 1996). Comparisons of the LChV ORF 1a product with those of BYV, LIYV and SPSVV allow prediction of a similar PCP domain with a cleavage site between Gly-619 and Ser-620 (Figs 1*b*–*d* and 2). Autoproteolysis at this site would yield a leader protein of 69 kDa, the largest among the cleaved leaders of closteroviruses (BYV 66 kDa, CTV 54 and 55 kDa, and LIYV 45 kDa). Previously, comparisons of the putative leader proteins encoded in

closterovirus genomes have revealed no statistically significant similarity among them, apart from the C-terminal PCP domain (Karasev *et al.*, 1995; Klaassen *et al.*, 1995). Our analysis also failed to identify any related sequences in the LChV and BYV proteins upstream of their PCP (Fig. 1*b*). In contrast, pairwise comparisons of the LChV and LIYV 1ab products revealed an additional segment of similarity located upstream of the PCP (positions 199–311 and 127–239 in the LChV and LIYV 1ab, respectively; Fig. 1*c*). Although this segment was not seen on the LChV/SPSVV plot (Fig. 1*d*), some lower SD segments could be found under less stringent comparison conditions allowing delineation of the borders of the conserved protein sequences. An alignment of the three leader proteins within these borders highlights the strong conservation of the PCP domains in the LChV, LIYV and SPSVV leader proteins, and the existence of related amino acid sequences located upstream of the PCP. These additional conserved domains are interrupted by deletions or insertions of variable length (Fig. 2). The putative cleavage sites of the closterovirus proteinases are found in non-conserved amino acid contexts, with the exception of positions P2, P1 and P′4 at which a bulky hydrophobic residue, a Gly residue and a negatively charged residue, respectively, are favoured (Fig. 2). Substantial variability is seen in the residues located rightward of the scissible bond (Gly in the BYV and CTV proteins, and Ser, Val and Ala in the proteins of LChV, SPSVV and LIYV, respectively; Fig. 2). Similar variability of the P′1 amino acids has been described for the leader PCP of arteriviruses and coronaviruses, which have Gly/Gly, Gly/Ala, Gly/Val or Gly/Tyr dipeptides as cleavage sites (Godeny *et al.*, 1993; Dong & Baker, 1994; den Boon *et al.*, 1995).

The recently published phylogenetic tree of the closteroviral CP and CPd suggested that the LChV sequences were most closely related to those of LIYV and SPSVV (Keim-Konrad & Jelkmann, 1996). Here, we studied the phylogenetic relationships among closteroviruses by comparing their putative RNA polymerases. The tentative closterovirus RNA polymerase tree (Fig. 3) showed branching of BYV, BYSV, CTV and LIYV compatible with that in the previous POL reconstructions (cf. Dolja *et al.*, 1994; Klaassen *et al.*, 1995). Further, our analysis suggested that the aphid-transmissible BYV, BYSV and CTV, on the one hand, and the whitefly-transmissible LIYV, SPSVV and cucumber chlorotic spot virus (CCSV) and the mealybug-transmissible LChV, on the other, represent two distinct evolutionary lineages (Fig. 3). Although this reconstruction clearly supports a rather recent divergence of the whitefly- and mealybug-transmitted closteroviruses, LChV appears to be the most remote member of this lineage (Fig. 3). Comparisons of the PCP domains of BYV and CTV versus LChV, LIYV and SPSVV show only limited amino acid sequence similarity (cf. consensus 1 and 2; Fig. 2). This agrees with the closterovirus grouping suggested by the RNA polymerase tree and implies evolution of the PCP and the replicative domains as an entity.

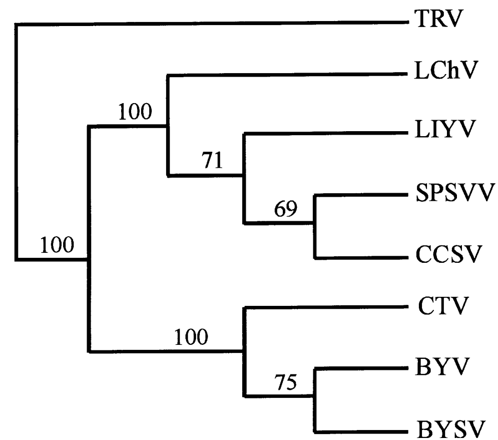

Fig. 3. Cluster dendrogram derived from aligned sequences of the closterovirus RNA polymerases (ORF 1b products). Multiple alignments were generated by the program OPTAL allowing calculation of the adjusted alignment scores as the number of standard deviations (SD) over the mean of 25 random simulations (Gorbalenya *et al.*, 1989).Trees were constructed from the alignments using the program PROTPARS in the PHYLIP package (Felsenstein, 1989), excluding the positions containing gaps. The number above each node shows the percentage of bootstrap replicates in which a given node was recovered. Branch lengths are arbitrary. The related sequence of tobacco rattle tobravirus (TRV) polymerase was used as the outgroup. Protein sequences were extracted from the databases, except for cucumber chlorotic spot virus (CCSV; personal communication from L. P. Woudt) and SPSVV (personal communication from H. J. Vetten).

This research was supported by a grant of the Deutsche Forschungsgemeinschaft (DFG), Bonn, FRG. We thank Dr H. J. Vetten for providing unpublished SPSVV sequence data.

References

Agranovsky, A. A. (1996). Principles of molecular organization, expression and evolution of closteroviruses: over the barriers. *Advances in Virus Research* **47**, 119–158.

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

Brodsky, L. I., Drachev, A. L., Tatuzov, R. L. & Chumakov, K. M. (1990). GENEBEE: a package of computer programs for biopolymer sequence analysis. *Biopolimeri i Kletka* **7**, 10–14.

den Boon, J. A., Faaberg, K. S., Meulenberg, J. J., Wassenaar, A. L., Plagemann, P. G., Gorbalenya, A. E. & Snijder, E. J. (1995). Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases. *Journal of Virology* **69**, 4500–4505.

Dolja, V. V., Karasev, A. V. & Koonin, E. V. (1994). Molecular biology and evolution of closteroviruses: sophisticated build-up of large RNA genomes. *Annual Review of Phytopathology* **32**, 261–285.

Dong, S. & Baker, S. C. (1994). Determinants of the p28 cleavage site recognized by the first papain-like cysteine proteinase of murine coronavirus. *Virology* **204**, 541–549.

Dougherty, W. G. & Semler, B. L. (1993). Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiological Reviews* **57**, 781–822.

Eastwell, K. C. & Bernardy, M. G. (1996). Association of high molecular weight double-stranded RNA with little cherry disease. *Canadian Journal of Plant Pathology* **18**, 203–208.

Felsenstein, J. (1989). *PHYLIP 3.2 Manual*. Herbarium, University of California at Berkeley.

Godeny, E. K., Chen, L., Kumar, S. N., Methven, S. L., Koonin, E. V. & Brinton, M. A. (1993). Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV). *Virology* **194**, 585–596.

Gorbalenya, A. E., Blinov, V. M., Donchenko, A. P. & Koonin, E. V. (1989). An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. *Journal of Molecular Evolution* **28**, 256–268.

Hamilton, R. I., Dodds, J. A. & Raine, J. (1980). Some properties of a nucleic acid associated with little cherry disease. *Acta Phytopathologica Academicae Scientiarum Hungaricae* **15**, 75–77.

Jelkmann, W. (1995). Cherry virus A: cDNA cloning of dsRNA, nucleotide sequence analysis and serology reveal a new plant capillovirus in sweet cherry. *Journal of General Virology* **76**, 2015–2024.

Karasev, A. V., Boyko, V. P., Gowda, S., Nikolaeva, O. V., Hilf, M. E., Koonin, E. V., Niblett, C. L., Cline, K., Gumpf, D. J., Lee, R. F., Garnsey, S. M., Lewandowski, D. J. & Dawson, W. O. (1995). Complete sequence of the citrus tristeza virus RNA genome. *Virology* **208**, 511–520.

Keim-Konrad, R. & Jelkmann, W. (1996). Genome analysis of the 3′ terminal part of the little cherry disease associated dsRNA reveals a monopartite clostero-like virus. *Archives of Virology* **141**, 1437–1451.

Klaassen, V. A., Boeshore, M. L., Koonin, E. V., Tian, T. & Falk, B. W. (1995). Genome structure and phylogenetic analysis of lettuce infectious yellows virus, a whitefly-transmitted, bipartite closterovirus. *Virology* **208**, 99–110.

Kozak, M. (1986). Bifunctional messenger RNAs in eukaryotes. *Cell* **47**, 481–483.

Legrand, G. & Verhoyen, M. (1986). Use of different methods to detect little cherry disease in ornamental cherry trees. *Acta Horticulturae* **193**, 283–289.

Ragetli, H. W. J., Elder, M. & Schroeder, B. K. (1982). Isolation and properties of filamentous viruslike particles associated with little cherry disease in *Prunus avium*. *Canadian Journal of Botany* **60**, 1235–1248.

Raine, J., Weintraub, M. & Schroeder, B. (1975). Flexuous rods and vesicles in leaf and petiole phloem of little-cherry diseased *Prunus* spp. *Phytopathology* **65**, 1181–1186.

Raine, J., Weintraub, M. & Schroeder, B. (1979). Hexagonal tubules in phloem cells of little cherry infected trees. *Journal of Ultrastructural Research* **67**, 109–116.

Received 29 January 1997; Accepted 11 April 1997