Analysis of the RNA of Potato yellow vein virus: evidence for a tripartite genome and conserved T-terminal structures among members of the genus Crinivirus

Livieratos, I.C.; Eliasco, E.; Muller, G.; Olsthoorn, R.R.C.L.; Salazar, L.F.; Pleij, C.W.A.; Coutts, R.H.A.

Analysis of the RNA of *Potato yellow vein virus*: evidence for a tripartite genome and conserved 3′-terminal structures among members of the genus *Crinivirus*

I. C. Livieratos,[1]† E. Eliasco,[1]† G. Müller,[2] R. C. L. Olsthoorn,[3] L. F. Salazar,[2] C. W. A. Pleij[3] and R. H. A. Coutts[1]

Correspondence
R. H. A. Coutts
r.coutts@imperial.ac.uk

[1]Department of Biological Sciences, Imperial College London, Sir Alexander Fleming Building, Imperial College Road, London SW7 2AZ, UK

[2]The International Potato Center, Apartado 1558, Lima, Peru

[3]Leiden Institute of Chemistry, Gorlaeus Laboratories, Einsteinweg 55, 2300 RA Leiden, The Netherlands

Received 16 December 2003
Accepted 17 February 2004

Double-stranded RNA preparations produced from potato plants graft-inoculated with a Peruvian isolate of *Potato yellow vein virus* (PYVV; genus *Crinivirus*, family *Closteroviridae*) contain five RNA species denoted RNA 1, RNA 2, RNA 3, x and y of approximately 8, 5·3, 3·8, 2·0 and 1·8 kbp, respectively. The complete nucleotide sequences of PYVV RNAs 1, 2 and 3 and Northern hybridization analysis showed that PYVV RNA 1 contained the replication module and an additional open reading frame (p7), while two distinct species, RNAs 2 and 3, contain the *Closteroviridae* hallmark gene array. Pairwise comparisons and phylogeny of genome-encoded proteins showed that PYVV shares significant homology with other criniviruses but is most closely related to the *Trialeurodes vaporariorum*-vectored *Cucumber yellows virus*. Secondary structure prediction of the 3′-untranslated regions of all three PYVV RNAs revealed four conserved stem–loop structures and a 3′-terminal pseudoknot structure, also predicted for all fully characterized members of the genus *Crinivirus* and some members of the genera *Closterovirus* and *Ampelovirus*.

INTRODUCTION

The virus family *Closteroviridae* contains positive-stranded, RNA plant viruses with long, flexuous virions and elongated genomes of up to 20 kb, which are transmitted by specific Homopteran vectors in a semi-persistent manner (Karasev, 2000; Martelli *et al.*, 2002). Based on the number of genomic RNAs and mode of transmission, the closteroviruses are currently divided into three genera (Karasev, 2000). The genus *Closterovirus*, with *Beet yellows virus* (BYV) as the prototype, includes aphid-transmitted closteroviruses with unipartite genomes. The genus *Ampelovirus*, with *Grapevine leafroll-associated virus 3* (GLRaV-3) as the prototype, includes mealybug-transmitted closteroviruses with unipartite genomes. The genus *Crinivirus*, with *Lettuce infectious yellows virus* (LIYV) as the prototype, includes whitefly-transmitted closteroviruses (WTCs) with bipartite genomes.

The genomic organization of all members of the family *Closteroviridae* shares two distinct features. Firstly, all closteroviruses possess genomes with a replication module that contains domains for a papain-like leader proteinase (L-Pro), a methyltransferase (MTR), a helicase (HEL) [open reading frame 1a (ORF1a)] and an RNA-dependent RNA polymerase [RdRp (ORF1b)] (Karasev, 2000). Secondly, downstream of the replication module, closteroviruses contain a hallmark gene array encoding a putative small hydrophobic protein, a heat-shock protein homologue (Hsp70h), a protein of 50–60 kDa, depending on the virus, the coat protein (CP) and the minor CP (CPm), arranged in a polar configuration (Karasev, 2000). For BYV (Alzhanova *et al.*, 2000; Napuli *et al.*, 2000, 2003) and *Citrus tristeza virus* (CTV) (Satyanarayana *et al.*, 2000), these genes encode proteins responsible for virus assembly and movement.

Because of the increasing number of WTCs and their significance as potential pandemic pathogens of agriculturally important crops (Wisler *et al.*, 1998), the molecular characterization of members of the genus is proceeding rapidly. The first WTC genome to be sequenced and characterized in its entirety was LIYV (Klaassen *et al.*, 1995).

The GenBank accession numbers for the nucleotide sequences determined in this work are AJ557128 (RNA 1), AJ557129 (RNA 2) and AJ508757 (RNA 3).

†Both authors contributed equally to this work.

LIYV RNA 1 (8·1 kb) encodes the replication module (Klaassen *et al.*, 1996) and an additional p32 ORF, the product of which enhances replication of LIYV RNA 2 (Yeh *et al.*, 2000). LIYV RNA 2 (7·2 kb) contains the *Closteroviridae* hallmark gene array, an additional ORF (p9) immediately upstream of the CP ORF and a putative 3′-terminal protein (p26) of unknown function (Klaassen *et al.*, 1995).

Diversity in the genus *Crinivirus* was unknown until the recent publication of the complete genomic sequences of three other criniviruses, *Sweet potato chlorotic stunt virus* (SPCSV; Kreuze *et al.*, 2002), *Cucurbit yellow stunting disorder virus* (CYSDV; Aguilar *et al.*, 2003; Coutts & Livieratos, 2003a; Livieratos & Coutts, 2002) and *Cucumber yellows virus* (CuYV; Hartono *et al.*, 2003). While LIYV, SPCSV, CYSDV RNA 2 and CuYV RNA 1 contain similarly sized and positioned ORFs, significant divergence has been reported regarding ORFs downstream of the replication module. For instance, in SPCSV RNA 1, ORFs that putatively encode an RNase III-like protein, a small hydrophobic protein (p7) and a 3′-terminal protein (p22) have been described downstream of the RdRp ORF (Kreuze *et al.*, 2002). In CYSDV RNA 1, ORFs that putatively encode three proteins (p5, p25 and p22) with no apparent amino acid similarities to equivalent putative proteins of SPCSV or LIYV have been described (Aguilar *et al.*, 2003; Coutts & Livieratos, 2003a). In CuYV, the replication module has been found on the smaller of the genomic RNAs, CuYV RNA 2, with no further ORFs downstream of the RdRp ORF (Hartono *et al.*, 2003).

Potato yellow vein virus (PYVV) is a WTC vectored by *Trialeurodes vaporariorum* (Westwood) known to cause a yellowing disease in potato crops in South America for over 50 years (Salazar *et al.*, 2000). The genomic sequence of PYVV is unknown except for a partial sequence of the *Hsp70h* gene (Salazar *et al.*, 2000) and the complete sequence of the *CPm* gene (Livieratos *et al.*, 2002). Over a period of 3 years, we have examined numerous dsRNA extracts from field isolates of PYVV from both Peru and Colombia. Single-strand conformation polymorphism analysis has strongly suggested low genetic variation within and between Peruvian and Colombian PYVV isolates (Offei *et al.*, 2004). In this study, we have cloned the three largest PYVV RNAs present in extracts of potato plants (17·3 kbp in total) infected with a graft-maintained Peruvian PYVV isolate. Our data suggest that PYVV possesses a tripartite genome and reveals conserved secondary structures in the 3′-untranslated regions (UTRs) for all crinivirus genomes.

METHODS

Plant material and virus isolate. The PYVV isolate used in this investigation was obtained from diseased potatoes (i.e. *Solanum tuberosum* ssp. *tuberosum* × *Solanum tuberosum.* ssp. *andigena*, cv. Yungay) originally growing in Chota, Cajamarca, Peru. The isolate was maintained in an insect-proof greenhouse in La Molina by graft inoculation of tissue culture plantlets cvs Atzimba × R-128.6 or DTO-33 (Salazar *et al.*, 2000). To confirm that the graft maintained PYVV isolate was biologically active, Koch's postulates for a virus were performed, which included back transmission of the isolate to potato plants using the whitefly vector.

Isolation and analysis of PYVV dsRNAs. PYVV dsRNA was isolated from infected potato leaves as described by Valverde *et al.* (1990). PYVV dsRNAs were treated with DNase I, separated by agarose gel electrophoresis and individual species purified using the RNAid kit (BIO 101) as described previously (Livieratos *et al.*, 1998). PYVV dsRNA preparations were analysed by Northern hybridization using ^{32}P-labelled riboprobes (Bringloe *et al.*, 1998) generated with a Riboprobe kit (Promega) from cloned RT-PCR amplification products.

Cloning and sequencing of PYVV-specific reverse transcription products and PCR amplicons. Specific oligonucleotides designed from the sequences of ORF 1a (L-Pro motif), the Hsp70h ORF (Salazar *et al.*, 2000) and the CPm ORF (Livieratos *et al.*, 2002) were used to produce clones in a genome-walking procedure as described previously (Livieratos *et al.*, 1999) using PYVV dsRNA as template. To clone the 5′ and 3′ termini of the individual RNAs, a modified RNA ligase-mediated rapid amplification of cDNA ends procedure was utilized based on that of Liu & Gorovsky (1993) using methylmercuric hydroxide-denatured dsRNA as substrate (Coutts & Livieratos, 2003b). All amplicons were cloned into the pGEM-T Easy system II vector (Promega) and transformed into *Escherichia coli* JM109 cells and sequenced in both directions. Sequence data were assembled and compared with databases using the BLAST server on the NCBI Web library (http://www.ncbi.nlm.nih.gov/blast) (Altschul *et al.*, 1997). Pairwise comparisons between closterovirus amino acid sequences were performed using the GAP and PILEUP programs (version 8.1; University of Madison Genetics Computer Group, Madison, USA) (Devereaux *et al.*, 1984). Phylogenies were estimated using the neighbour-joining algorithm and amino acid sequences were aligned using CLUSTAL_X (Thompson *et al.*, 1997). Phylograms were drawn using TREEVIEW version 1.6.1 freeware (Page, 1996). RNA secondary structures were predicted using the STAR algorithm (Gultyaev *et al.*, 1995).

RESULTS AND DISCUSSION

Electrophoretic analysis of dsRNA preparations from Peruvian PYVV-infected plants consistently revealed the presence of at least five dsRNA species (Salazar *et al.*, 2000) that were absent from healthy plant extracts. The sizes of these dsRNAs (denoted dsRNA 1, dsRNA 2, dsRNA 3, x and y) were estimated to be approximately 8, 5·3, 3·8, 2·0 and 1·8 kbp, respectively (Fig. 1b). The three largest dsRNAs were tentatively assigned as double-stranded versions of single-stranded PYVV genomic RNA.

RNA 1

Analysis of the determined sequence of PYVV RNA 1 (8035 nt) revealed the existence of three ORFs: the replication module, which consisted of ORF 1a and b and contained the conserved L-Pro, MTR, HEL and RdRp domains, and p7, a small hydrophobic protein (Fig. 1a). The calculated molecular mass of the polypeptide encoded by ORF 1a was 226 kDa and contained at the N terminus a putative L-Pro with a conserved papain-like domain, similar to other viruses in the family *Closteroviridae* (Karasev,

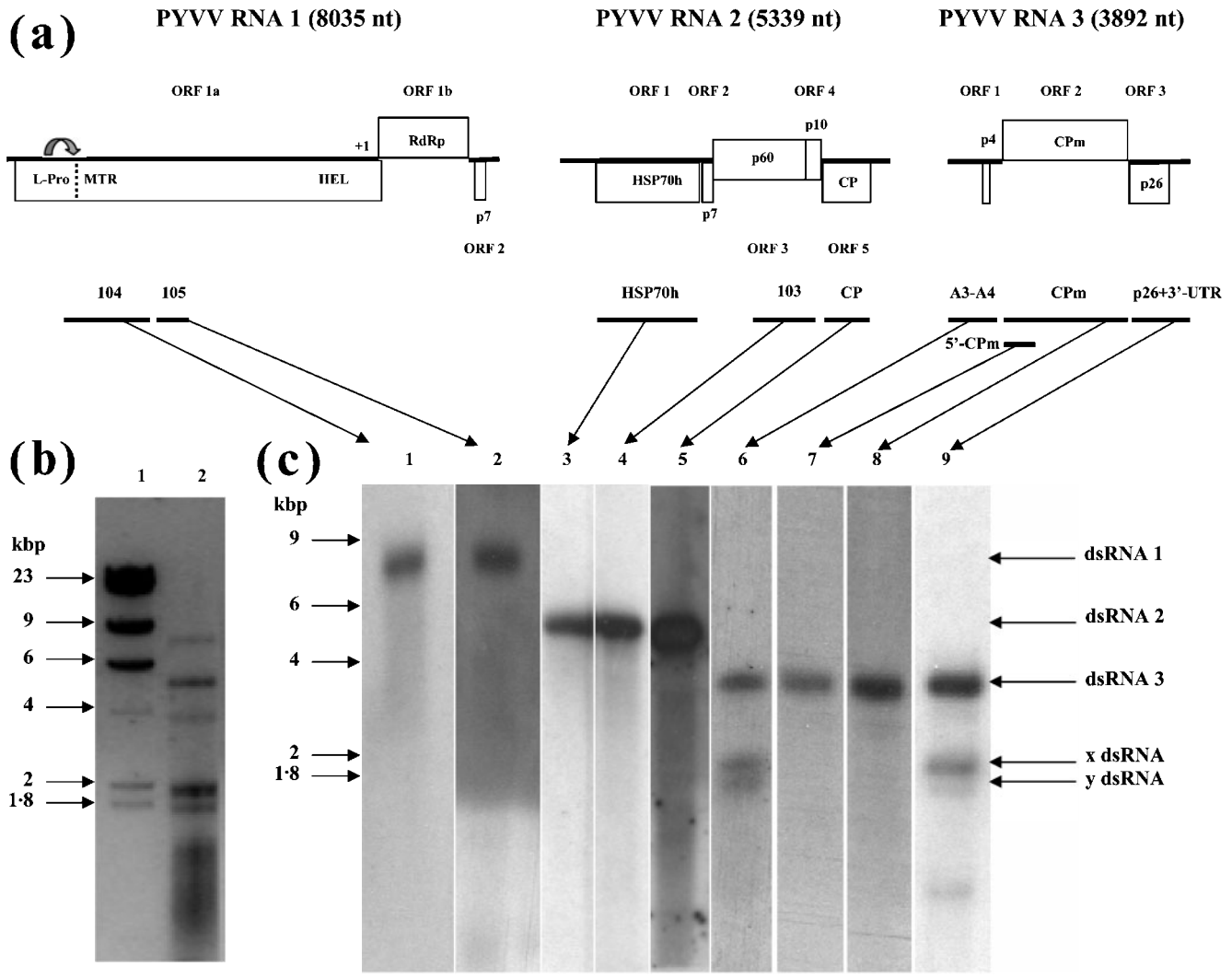

Fig. 1. (a) Schematic representation of the genome organization of PYVV. Lines represent PYVV RNAs 1, 2 and 3, and ORFs are shown above, below or on the line, indicating which reading frame they are found in. On PYVV RNA 1, the arrow indicates the predicted autocatalytic cleavage site and +1 designates the putative ribosomal +1 frameshift site. (b) Agarose gel electrophoresis of a dsRNA preparation from PYVV-infected plants (lane 2). Lane 1 contains *Hin*dIII-digested lambda DNA molecular mass markers. (c) Northern blot hybridization analysis of PYVV dsRNAs. Positive-sense riboprobes, generated by *in vitro* transcription of the linearized clones shown in (a), were used to probe individually blotted dsRNAs shown in lanes 1–9.

2000). Autoproteolysis at the predicted cleavage site between Gly-468 and Met-469 by the catalytic residue His-450 would yield a putative L-Pro of 48 kDa and a larger protein of 178 kDa. Upstream of the catalytic Cys-400, the leader proteins of PYVV RNA 1 and other viruses in the family *Closteroviridae* showed limited sequence similarity. In BYV, L-Pro is multifunctional where, apart from its primary role in polyprotein processing and long-distance virus transport, the 54 N-terminal amino acids of the protein also function in RNA amplification (Peng *et al.*, 2001, 2003). Downstream of the L-Pro domain were found the conserved MTR (aa 556–730) and HEL (aa 1728–1996) domains. As with other criniviruses, the region between the MTR and HEL domains contained no significant similarity to any known proteins. The ORF 1b encoded a conserved RdRp domain that may be expressed by a +1 frameshift as proposed for other members of the *Closteroviridae* (Agranovsky *et al.*, 1994). The frameshifting in PYVV RNA 1 would yield a fusion protein of 285 kDa identical in size to those predicted for the equivalent proteins of CYSDV RNA 1 (Aguilar *et al.*, 2003; Coutts & Livieratos, 2003a) and SPCSV RNA 1 (Kreuze *et al.*, 2002).

ORF2 putatively encoded a small hydrophobic protein of 6·9 kDa (Fig. 1a) that contained a potential transmembrane helix between amino acids 31 and 51. Hydrophobic proteins of similar size but with little sequence similarity are encoded by all members of the *Closteroviridae* and are

thought to possess membrane-binding properties. However, the number, size, genomic location and nature of these small proteins vary in all sequenced criniviruses. In BYV, a subgenomic mRNA encoding a small hydrophobic protein has been identified (Peremyslov & Dolja, 2002), whose gene product is essential for virus cell-to-cell movement (Alzhanova *et al.*, 2000).

RNA 2

The RNA 2 (5339 nt) sequence revealed five predicted ORFs: Hsp70h, p7, p60, p10 and CP (Fig. 1a), which is part of the closterovirus gene array. PYVV RNA2 ORF 1 could be translated to produce a 62·3 kDa protein homologous to members of the Hsp70 family of molecular chaperones and closely related to those found in other closteroviruses. The putative PYVV Hsp70h protein was most closely related to the CuYV, SPCSV, CYSDV and *Tomato chlorosis crinivirus* (ToCV) *Hsp70h* genes. The Hsp70h proteins of BYV (Napuli *et al.*, 2000, 2003) and LIYV (Tian *et al.*, 1999) are physically associated with virus particles and are also essential for correct virion assembly and virus movement in both BYV (Alzhanova *et al.*, 2001) and CTV (Satyanarayana *et al.*, 2000).

ORF 2 encoded a putative protein of 6·8 kDa that was similar in size, genomic position and sequence (30 % identity) to CYSDV p6 (Aguilar *et al.*, 2003), which was not described by Livieratos & Coutts (2002), and a putative 6·2 kDa protein predicted, but not described, from the sequence of CuYV RNA 1 (Hartono *et al.*, 2003). Whether these putative proteins are expressed in plants is not yet known but all of the proteins contained a central conserved stretch of amino acids.

ORF 3 putatively encoded a protein of 60 kDa, which was homologous to proteins of other clostero- and criniviruses encoded at the corresponding genomic position (Karasev, 2000) with the corresponding CuYV p59 being the most similar (56 % identity) amongst criniviruses. Recently, it has been demonstrated that BYV p64 is the fourth integral component of BYV virions together with CPm and Hsp70h (Napuli *et al.*, 2003). Napuli *et al.* (2003) also illustrated that ~60 kDa proteins of viruses from the genera *Closterovirus* and *Ampelovirus* contain CP-like domains and that, by analogy with BYV, these proteins may have a dual function in virion tail assembly and motility. We extended this multiple alignment of CPs, CPms and CP-like C-terminal domains of the ~60 kDa proteins to include PYVV and all sequenced criniviruses (Fig. 2). Three amino acid residues, asparagine (N), arginine (R) and aspartic acid (D), were conserved in all the crinivirus alignments and representative members of the genera *Ampelovirus* (GLRaV-3) and *Closterovirus* (CTV) (Fig. 2). These alignments suggested that the crinivirus ~60 kDa proteins may fulfil a similar role in the virus replication cycle to the BYV p64 protein and the observation that the LIYV Hsp70h and the p59 proteins are associated with virions (Tian *et al.*, 1999) is not at variance with this notion.

```
                                  10        20        30        40        50
CYSDVCPm      372   : CNIERAMLRMYSSQIFKLLKDGVLVPGWS--HVKRRNFKEECAYMACDFYD :  38
SPCSVCPm      597   : CNVERTLLRNRSDKIFRLLKAKVLVLPLK--HARVRGIKPEMAHMACDFMD :  38
CuYVCP        161   : DNPIRQYLRGNENQVAILRATGKLKSDGH--LAAKHGTTTQFWDSTSDFTN :  42
PYVVCP        164   : SNPLRQYMRLCENQIEMLKSSGKIVSDGH--LAAKHGTTSQYSSSVGDYVN :  42
CuYVp59       429   : INVLRQWANKRSYRALTLFRSANFDPG------FFSNVPGILPYMRFDFYK :  43
PYVVp60       429   : MNVLRLWANSRSSRAMRLFRSLNFNPG------LFSYCPGILDYMRFDFYK :  47
CYSDVp59      429   : DNIIRQWANHRGNRAMRLFKVSGFKPG------LFSTVPNILPWMRFDFFK :  44
SPCSVp60      430   : NNLIRQWANNRGDRARAMFKICEYHPG------LFGSIPRIDNHMRFDFFK :  43
CuYVCPm       570   : CNVERLLLRSRSLLIFNLLKRKQLEWPHL--HAQRRGLRPEFAYMACDFYD :  38
CYSDVCP       158   : PNPFRQYARAHEHEIEILKATGKVTVDHH--LQAKHGVLPQFWNVPADYVN :  44
PYVVCPm       587   : RNVERIYLRSRSAEILEKLRNRVLIWPKT--HANKRGLLPEYAYLACDFFK :  38
SPCSVCP       164   : KNAARKWARAHANDYKVLVGLGIVKPDYH--LQAKQGVLPEYWHLATDFMR :  45
LIYVp59       426   : SNIVRLWCNKRSNLALGYFKSRNIQLY------LYSKYPRLLNYMRFDYFK :  43
LIYVCPm       365   : HNLERVLLRHYAQEILILLRSKVLEWPRK--LARNKGIFEQYAYMACDFFD :  38
LIYVCP        155   : PNVVRQYARAMEKTINNIRSAGIINSNGV--LAAKHGVLASYRNSYSDFAV :  45
LChV-1p60     428   : RNYIRLFCSKRANLAIRVNEKFKFQPT------VMDKCPKVLPYMKIDFYK :  44
LChV-1CPm     581   : INLFRRYLRSCTVEVLELLRIGTLSPSFG--RAIMLGIPKQFAFLACDFWN :  34
LChV-1CP      308   : ENAERRFAKKEYNAIQNALAAAGYVSSER--LAAKWGAAPNKRGKVSDATP :  47
CTVp61        406   : INVRRQFMGRHSEVALRLYRNLGLRFPPISSVRLPAHHGYLYV----DFYK :  82
GLRaV-3p55    400   : INAQRVLCRYYSDLTCLARRHYGIRRNNWKTLSYVDGTL---AYDTADCIT :  35
GLRaV-3CP     221   : ENPVRQYLAYFTPTFITATLNGKLVMNEK--VMAQHGVPPKFFPYTIDCVR :  43
GLRaV-3CPm    387   : TNTVRSFMRYFAHTTITLLIEKKIQPACT--ALAKHGVPKRFTPYCFDFAL :  41
CTVCPm        147   : PNALRKFACTFEELHLCMARLRPDLYENK--RTTRAGTPHLKGYLSADFLS :  44
CTVCP         132   : TNALRVWGRTNDALYLAFCRQNRNLSYGG--RPLDAGIPAGYHYLCADFLT :  42
                       N  R                                          D
```

Fig. 2. Multiple alignment of the CPs, CPms and CP-like C-terminal domains of the ~60 kDa proteins of all sequenced criniviruses, LChV-1 and a representative member of the genus *Ampelovirus* (GLRaV-3) and the genus *Closterovirus* (CTV). The alignment was constructed using the CLUSTAL_X algorithm (Thompson *et al.*, 1997). The two signature residues that are conserved in the CPs of elongated plant viruses and a conserved N residue are shown in white against a black background. The ranges of the aligned regions in the corresponding protein sequences are indicated at the beginning and end of each sequence. The GenBank accession numbers are as follows: CTV, U16304; CuYV, AB085612 and AB085613; CYSDV, AJ537493 and AJ439690; GLRaV-3, AF039204; LChV-1, X93351; LIYV, U15440 and U15441; PYVV, AJ557128, AJ557129 and AJ508757; SPCSV, AJ428554 and AJ428555. Each virus is shown by its name followed by the protein.

ORF 4 encoded a putative protein of 9·8 kDa. This gene showed no significant similarities with any other sequences in the database, and pairwise alignment with putative proteins encoded by similar ORFs in identical positions in all sequenced criniviruses revealed moderate amino acid sequence identity. These equivalent proteins appear to be unique to the genus *Crinivirus*, since no similar ORFs have been detected in closteroviruses (Karasev, 2000).

ORF 5 encoded the putative CP of 28·2 kDa. Based on size and homology with the CPs of all criniviruses and the possession of the invariant consensus of amino acid residues serine (S), R, glycine (G) and D (R and D are shown in Fig. 2), as found in most filamentous virus CPs (Alzhanova *et al.*, 2001; Dolja *et al.*, 1991).

RNA 3

The RNA 3 (3892 nt) sequence revealed the existence of three potential ORFs: p4, CPm and p26, which represent the remainder of the closterovirus gene array (Fig. 1a). ORF1 encoded a putative protein of 4 kDa that had no counterpart with any other protein described for the *Closteroviridae* and no significant homology with any other sequence in the databases.

ORF2 encoded a putative protein of 77·5 kDa. The predicted protein was identified as CPm based on significant homology with other characterized closterovirus CPm proteins, especially those from criniviruses (Livieratos *et al.*, 2002). Studies of clostero- and crinivirus particles have shown that CPm is assembled at one end of the virus (Agranovsky *et al.*, 1995; Tian *et al.*, 1999; Zinovkin *et al.*, 1999) and is essential for virion assembly and cell-to-cell movement (Alzhanova *et al.*, 2000, 2001; Napuli *et al.*, 2003). In LIYV, CPm has been implicated in virus transmission by the whitefly vector (Tian *et al.*, 1999).

ORF 3 encoded a putative protein of 26·4 kDa. Putative proteins of a similar size and genomic position are found in LIYV RNA 2 (Klaassen *et al.*, 1995), CYSDV RNA 2 (Aguilar *et al.*, 2003; Livieratos & Coutts, 2002), SPCSV RNA 2 (Kreuze *et al.*, 2002) and CuYV RNA 1 (Hartono *et al.*, 2003) and appear to be unique to the genus *Crinivirus*. However, PYVV p26 shared only limited identity with its counterpart proteins in the criniviruses (maximum 36 % identity with CuYV p26) and no significant homology with any other sequence in the databases.

Northern hybridization analysis of PYVV dsRNAs

In agreement with the single-strand conformation polymorphism analysis (Offei *et al.*, 2004), electrophoretic and Northern hybridization analysis illustrated that the dsRNA profiles of the three largest dsRNA species from numerous field-infected Colombian PYVV isolates were virtually indistinguishable from the Peruvian isolate used in this study (Fig. 3). To distinguish between the different dsRNA species present in the PYVV dsRNA preparations, radiolabelled probes were constructed from clones of each dsRNA for Northern blot hybridization. Riboprobes generated from clones 104 and 105 only hybridized with RNA 1 (Fig. 1c, lanes 1 and 2, respectively). Riboprobes generated from clones HSP70h, 103 and CP only hybridized with RNA 2 (Fig. 1c, lanes 3–5, respectively), while riboprobes generated from clones A3-A4, 5′-CPm, CPm and p26 + 3′-UTR only hybridized with RNA 3 (lanes 6–9, respectively). Riboprobes A3-A4 and p26 + 3′-UTR also hybridized with the x and y dsRNA species (Fig. 1c, lanes 6 and 9, respectively).

Using probes constructed from cloned RT-PCR amplification products (detailed in Fig. 1a), Northern hybridization analysis confirmed that the PYVV dsRNAs contained the typical crinivirus Hsp70h, p60, p9, CP, CPm and p26 ORFs distributed on two dsRNA molecules, RNA 2 and RNA 3. In addition, the two smallest PYVV dsRNA species (designated x and y) contained sequences derived from the 5′ and 3′ termini of PYVV dsRNA 3 with extensive internal deletions. These observations suggest that these species are unlikely

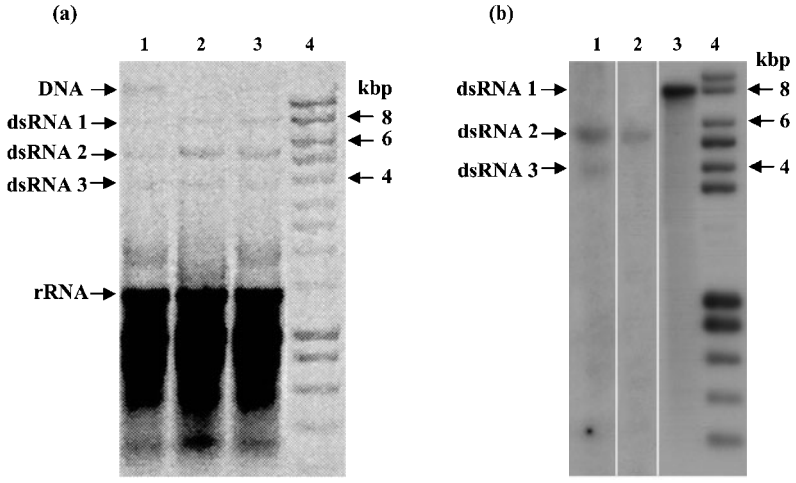

Fig. 3. (a) Agarose gel electrophoresis (1 %) of dsRNA preparations from three different PYVV Colombian field isolates (lanes 1–3). (b) Northern blot hybridization analysis of dsRNA extracts from a Colombian field PYVV isolate. The blots were hybridized with positive-sense riboprobes derived from the HSP70h and p26 ORFs (lane1), the HSP70h ORF (lane 2) and the HEL motif (lane 3). A molecular mass marker (Hyperladder I; Bioline) was included in lane 4 of both analyses and plant rRNA is indicated with an arrow.

to represent 3′ co-terminal subgenomic dsRNAs, as they do not contain contiguous sequences and are more likely to be defective (D) dsRNAs. D RNAs are a common feature of closterovirus infections and in the most extensively studied crinivirus example, LIYV, the D RNAs were derived exclusively from LIYV RNA 2 and included both the 5′ and 3′ termini of LIYV RNA 2, but had one extensive internal deletion. In LIYV-infected plants, these D RNAs exist as a heterogeneous population, differing in size and deletion location (Rubio *et al.*, 2000).

Phylogeny

Phylogenetic analyses were performed on the amino acid sequences of the conserved proteins RdRp, Hsp70h, p60 and CP of all sequenced members of the family *Closteroviridae*. The data revealed that PYVV sequences consistently cluster together with the known members of the genus *Crinivirus* and the unassigned *Little cherry virus* 1 (LChV-1), whose vector is not known (Fig. 4). PYVV is distinct from the members of the aphid-transmissible genus *Closterovirus* and the mealybug-transmissible genus *Ampelovirus*, and is most closely related to CuYV, a *T. vaporariorum*-transmitted crinivirus denoted as a strain of *Beet pseudo-yellows virus* (Hartono *et al.*, 2003).

UTRs

As with the first 5, 8, 10 and 6 nt of the respective 5′-UTRs of LIYV, SPCSV, CuYV and CYSDV RNA 1 and 2, which are identical for each virus (Klaassen *et al.*, 1995; Kreuze *et al.*, 2002; Hartono *et al.*, 2003; Aguilar *et al.*, 2003), the first 5 nt of all three PYVV RNAs were also identical. The putative 5′-UTRs of PYVV RNAs 1, 2 and 3, which were 182, 620 and 626 nt in length, respectively, were analysed for secondary structure using the STAR program. Interestingly, RNA 1, having by far the shortest 5′-UTR, revealed a repeated sequence (nt 61–87 and nt 95–121) just upstream of a long (31 nt) single-stranded region of C and A residues only. These repeats were folded into a hairpin with a conserved symmetrical internal loop having two conserved C–U mismatches, while the bottom stem was supported by one covariation. The only other difference was a single substitution in the hairpin loop (Fig. 5a). Interestingly, an identical hairpin was present in RNA 3 (nt 77–101) as part or top of a longer hairpin, which in turn was also immediately adjacent to a downstream A/C-rich region (Fig. 5a). In this case, three covariations strongly supported the existence of the top part of this hairpin. However, no equivalent repeated structure was found upstream in RNA 3, except for possibly two similar C residues in an internal loop (nt 18–19, Fig. 5a). For RNA 2, two hairpins having a stem region in common (nt 25–57 and nt 76–120) were also proposed. Again, both hairpins were followed downstream by a C/A-rich region. The occurrence of two single-stranded C and U residues, respectively, at the bottom of the repeated stem may be reminiscent of the situation in the two repeated hairpins of RNA 1. So far, no common features have been detected for PYVV RNAs 2 and 3 for the regions downstream of the C/A-rich regions, nor when compared with the sequences of the other crinivirus RNAs. Also, no (repeated) hairpins have been found so far in either RNA 1 or RNA 2 of the other criniviruses.

The putative 3′-UTRs of PYVV RNAs 1, 2 and 3 are 241, 291 and 232 nt in length, respectively. Using the STAR algorithm (Gultyaev *et al.*, 1995) and sequence comparisons, four stem–loop structures, hpI–hpIV, were identified in the 3′-terminal 180 nt of RNAs 1, 2 and 3 (Fig. 5b; Fig. 6). HpI and hpII were flanked by a putative interaction, designated A1. Of these stem–loops, hpII, hpIII and hpIV were supported by covariations in RNAs 1 and 2 (Fig. 5b). HpI was strongly conserved both in primary and secondary structure and showed the possibility of pseudoknot formation with the G/C-rich sequence near the 3′ end (not predicted by STAR). To obtain phylogenetic support for hpI and the pseudoknot PK1, the 3′-UTRs of RNAs from other criniviruses were folded. HpII, III and IV could also be identified in RNAs from SPCSV, CYSDV, LIYV and CuYV, yielding many covariations that supported their existence (Fig. 5b). HpI was conserved in RNAs 1 and 2 of SPCSV, CYSDV and CuYV. In LIYV RNA 1, hpI was missing stem region B2 but was otherwise almost identical to hpI of the SPCSV RNAs. In LIYV RNA 2, whose 3′-UTR shares only 30 % sequence identity with that of RNA 1, the equivalents of hpI and PK1 could not be formed. Interestingly, the possibility of pseudoknot formation is conserved in the other crinivirus RNAs and is supported by covariations in SPCSV and CYSDV RNAs (Fig. 7). This type of pseudoknot is quite unusual in that it has a small stem–loop inside loop L1 (Pleij *et al.*, 1985). Comparison with other members of the family *Closteroviridae* showed that the pseudoknot is conserved in the genus *Closterovirus* as well (Fig. 5b; Fig. 7). It is noteworthy that the 3′-UTR of LChV-1 RNA, which is an unassigned species in the family *Closteroviridae* (Martelli *et al.*, 2002), shared many features with the 3′-UTR of criniviruses, including the formation of PK1 (Fig. 5b; Fig. 7).

Conclusions

Due to the lack of a successful protocol for the purification of PYVV virions (Salazar *et al.*, 2000) and, subsequently, genomic ssRNA, we used dsRNA preparations to clone and sequence the PYVV genome of 17 kb. Both our sequence and Northern hybridization analysis showed that RNA 1 encodes the replication module, while two distinct RNA molecules contain the remainder of the closterovirus hallmark gene array. These data suggest that PYVV could possess a tripartite genome and are supported by the observation that all three PYVV RNA 3′-UTRs are very similar but different from each other. To confirm this, an analysis of virion ssRNA and the detection of subgenomic RNAs in infected plants are ultimately required, bearing in mind the well-documented formation of D RNAs for several closteroviruses (Che *et al.*, 2003; Rubio *et al.*, 2000). Irrespective of this, crinivirus genomes appear to be particularly diverse immediately downstream of the RdRp

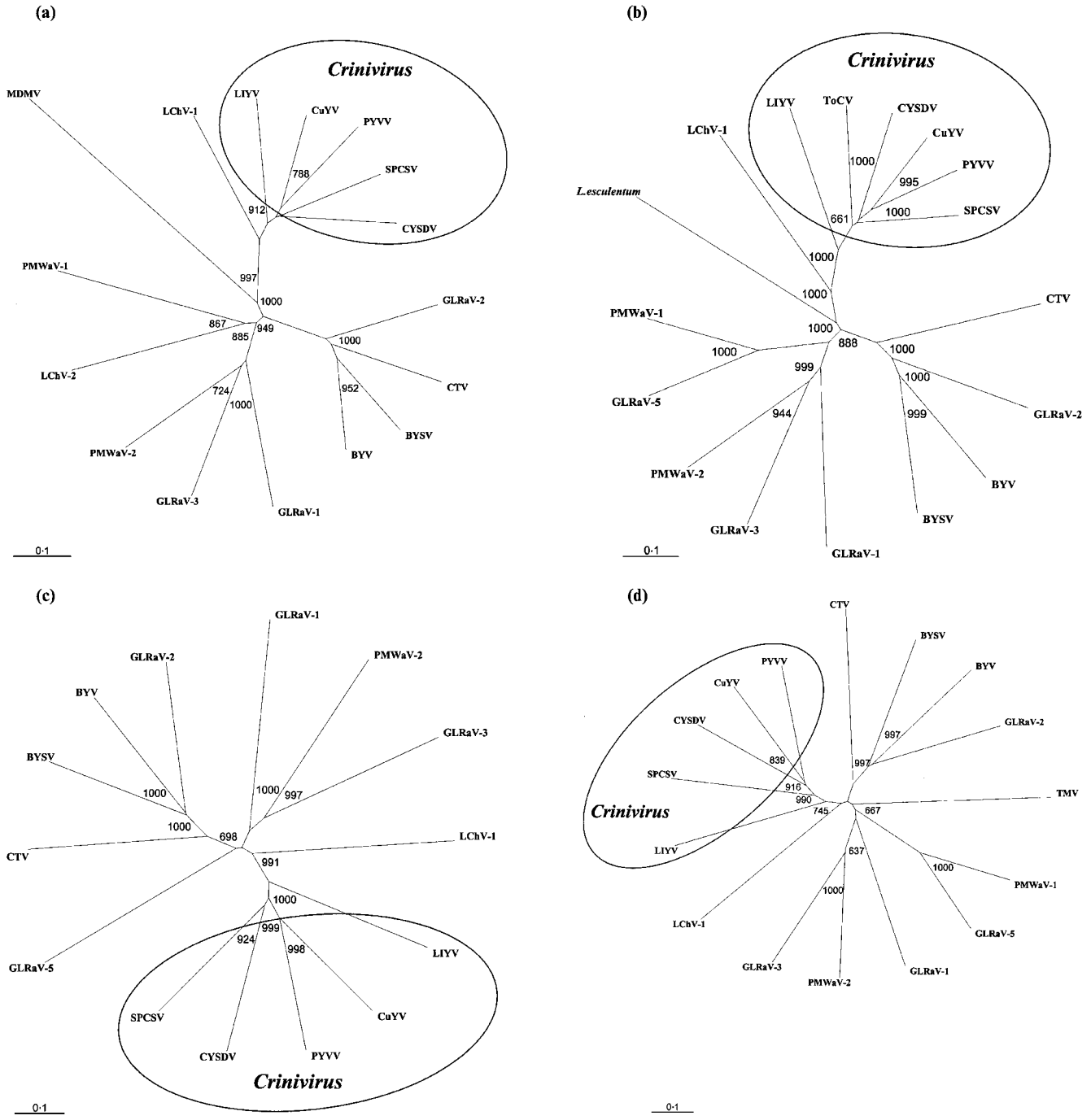

Fig. 4. Dendrograms showing phylogenetic trees of closteroviruses as determined from homologous putative genes. These relationships were determined from the amino acid sequences deduced for the RdRp (a), Hsp70h (b), p60 and equivalent proteins (c), and CP (d) of the respective closteroviruses. For RdRp analysis, the corresponding domain of *Maize dwarf mosaic virus* (MDMV; GenBank accession no. NP569138) was used as an outgroup. For Hsp70h analysis, the corresponding domain of the tomato Hsp70h (*Lycopersicon esculentum*; GenBank accession no. P34935) was used as an outgroup. For CP analysis, the corresponding sequence from *Tobacco mosaic virus* (TMV; GenBank accession no. P03576) was used as an outgroup. The phylograms were generated by the CLUSTAL_X program (Thompson *et al*., 1997) as described in the text. The scale box represents distances scaled as substitutions per amino acid residue. For each internal node, bootstrap numbers (runs out of a thousand in which the given node appeared) are included. The closterovirus GenBank accession numbers are as shown in Fig. 2, plus: BYV, AF190581; *Beet yellow stunt virus* (BYSV), U51931; GLRaV-1, -2, -3 and -5, AF195822, AF039204, AF037268 and AF039552, respectively; LChV-2, AF333237; Pineapple mealybug wilt-associated virus 1 (PMWaV-1), AF414119; PMWaV-2, AF28313; ToCV, AF024630.

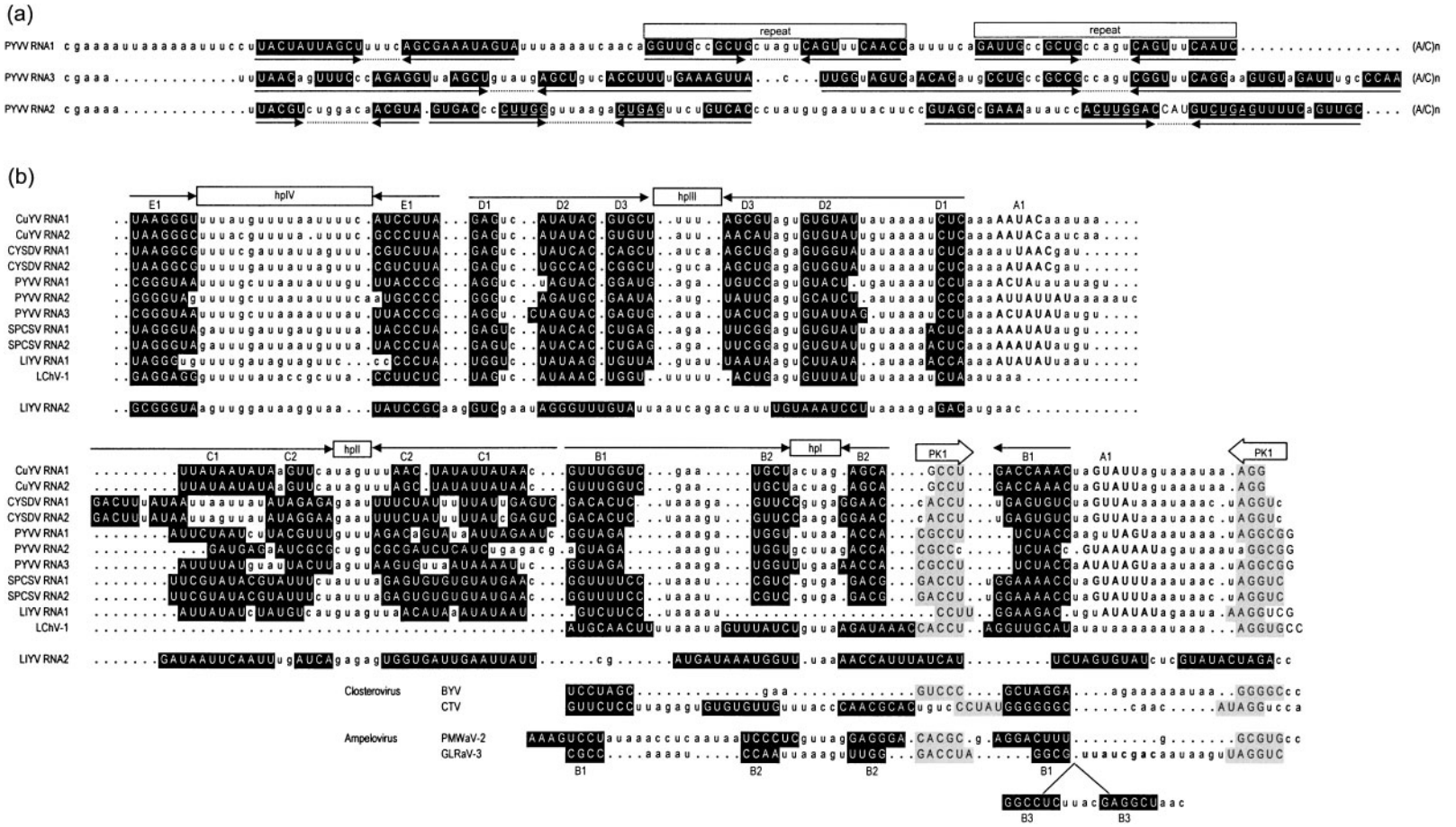

Fig. 5. (a) Structure alignment of the 5′ part of the 5′-UTR of PYVV RNAs 1, 2 and 3. Pairs of rightward- and leftward-pointing arrows below the sequence indicate stem–loop structures. Nucleotides involved in base pairing are shown in highlighted upper case; unpaired bases are shown in lower case. $(A/C)_n$ denotes the A/C-rich region in the 5′-UTRs (see text). The underlined regions in PYVV RNA 2 represent a repeated stem region. Dots have been introduced to maximize the alignment. (b) Structure alignment of the 3′-terminal region of the 3′-UTRs of RNAs of the genus *Crinivirus*. For comparison, the 3′ termini of two representative members each of the genera *Closterovirus* and *Ampelovirus* are shown in the bottom panel. Note that LIYV RNA 2 is set apart to emphasize its divergent structure with respect to hpI and PK1. Pairs of rightward- and leftward-pointing arrows above the sequence indicate stem–loop structures. The putative stem region A1 is shown in bold upper case. See (a) for definition of stem regions. Abbreviations of closteroviruses are as shown in Figs 2 and 4.

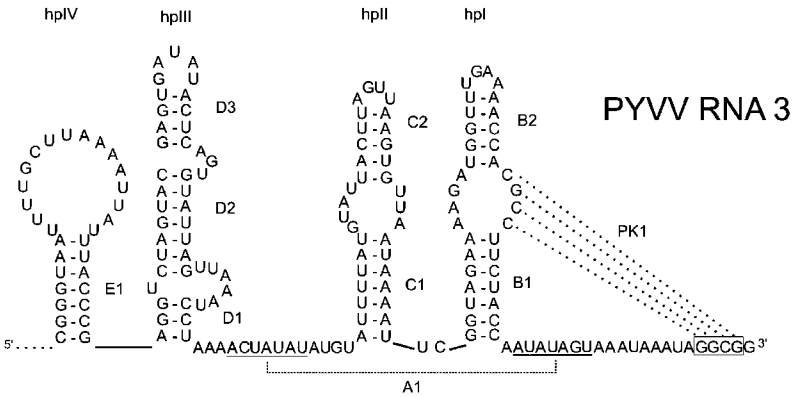

Fig. 6. Secondary structure model for the 3′-UTR of PYVV RNAs (the sequence of RNA 3 is shown). Stem–loop structures are denoted as hpI, hpII, hpIII and hpIV, and are composed of stem elements B1 and B2, C1 and C2, D1–D3, and E1, respectively. Sequences involved in the putative interaction A1 are underlined. Dotted lines indicate the pseudoknot interaction PK1. Lines between E1 and D1 and between C1 and B1 have been introduced for the sake of clarity.

ORF and the presence of a sole ORF (p7) in PYVV RNA 1 is in agreement with this notion. Also, the p7 ORF, which is positioned between the Hsp70h and p60 ORFs in PYVV (this study), CuYV (Hartono *et al.*, 2003) and CYSDV (Aguilar *et al.*, 2003), is absent from LIYV and SPCSV RNA 2 (Klaassen *et al.*, 1995; Kreuze *et al.*, 2002). The 3′-UTRs of all crinivirus and some closterovirus RNAs are predicted to contain a very similar secondary structure that includes four stem–loops followed by a pseudoknot structure. One of the major functions of the 3′ *cis* elements is the initiation of negative-strand synthesis, which involves higher-order structures concerned with motif presentation or promoter replicase binding functions (Buck, 1996). In addition, 3′ structures can be involved in RNA stability, cellular targeting, packaging and regulation of RNA synthesis and translation (Dreher, 1999). Studies on the importance of 3′ structures in closterovirus RNA replication have been described for CTV (Satyanarayana *et al.*, 2002) but similar investigations with criniviruses have yet to be addressed.

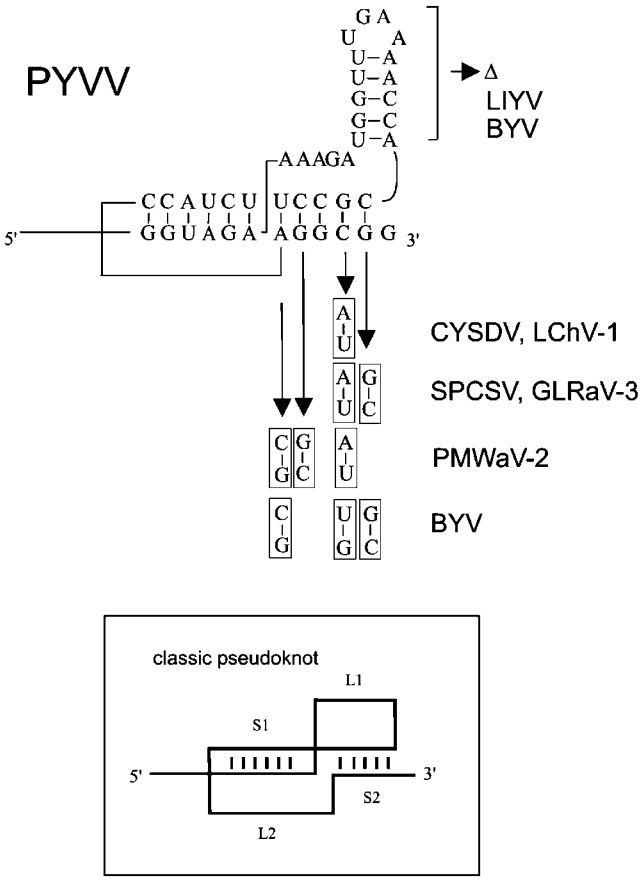

Fig. 7. Schematic representation of the pseudoknot in the three PYVV RNAs. Only the relevant bases are shown. The inset shows a schematic diagram of a classic H-type pseudoknot. S1 and S2 denote stems 1 and 2; L1 and L2 denote loops 1 and 2. Covariations occurring in S2 of other viral RNAs from the family *Closteroviridae* are indicated by the boxed base pairs below S2. Δ denotes the absence of the hairpin in L1 from LIYV RNA 1 and BYV RNA.

ACKNOWLEDGEMENTS

We thank The Central Research Fund of the University of London and the Department for International Development (DFID-UK) for financial support.

REFERENCES

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

Agranovsky, A. A., Lesemann, D. E., Maiss, E. M., Hull, R. & Atabekov, J. G. (1995). "Rattlesnake" structure of a filamentous plant RNA virus built of two capsid proteins. *Proc Natl Acad Sci U S A* **92**, 2470–2473.

Aguilar, J. M., Franco, M., Marco, C. F., Berdiales, B., Rodriguez-Cerezo, E., Truniger, V. & Aranda, M. A. (2003). Further variability within the genus *Crinivirus*, as revealed by determination of the complete RNA genome sequence of *Cucurbit yellow stunting disorder virus*. *J Gen Virol* **84**, 2555–2564.

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

Alzhanova, D. V., Hagiwara, Y., Peremyslov, V. V. & Dolja, V. V. (2000). Genetic analysis of the cell-to-cell movement of beet yellows closterovirus. *Virology* **268**, 192–200.

Alzhanova, D. V., Napuli, A. J., Creamer, R. & Dolja, V. V. (2001). Cell-to-cell movement and assembly of a plant closterovirus: roles for the capsid proteins and Hsp homolog. *EMBO J* **20**, 6997–7007.

Bringloe, D. H., Gultyaev, A. P., Pelpel, M., Pleij, C. W. A. & Coutts, R. H. A. (1998). The nucleotide sequence of satellite tobacco necrosis virus strain C and helper-assisted replication of wild-type and mutant clones of the virus. *J Gen Virol* **79**, 1539–1546.

Buck, K. W. (1996). Comparison of the replication of positive-stranded RNA viruses of plants and animals. *Adv Virus Res* **47**, 159–251.

Che, X., Dawson, W. O. & Bar-Joseph, M. (2003). Defective RNAs of *Citrus tristeza virus* analogous to *Crinivirus* genomic RNAs. *Virology* **310**, 298–309.

Coutts, R. H. A. & Livieratos, I. C. (2003a). Nucleotide sequence and genome organization of *Cucurbit yellow stunting disorder virus* RNA1. *Arch Virol* **148**, 2055–2062.

Coutts, R. H. A. & Livieratos, I. C. (2003b). A rapid method for sequencing the 5′- and 3′-termini of dsRNA viral templates using RLM-RACE. *J Phytopathol* **151**, 525–527.

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

Dolja, V. V., Boyko, V. P., Agranovsky, A. A. & 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.

Dreher, T. W. (1999). Functions of the 3′-untranslated regions of positive strand RNA viral genomes. *Annu Rev Phytopathol* **37**, 151–174.

Gultyaev, A. P., van Batenburg, F. H. D. & Pleij, C. W. A. (1995). The computer simulation of RNA folding pathways using a genetic algorithm. *J Mol Biol* **250**, 37–51.

Hartono, S., Natsuaki, T., Genda, Y. & Okuda, S. (2003). Nucleotide sequence and genome organization of *Cucumber yellows virus*, a member of the genus *Crinivirus*. *J Gen Virol* **84**, 1007–1012.

Karasev, A. V. (2000). Genetic diversity and evolution of closteroviruses. *Annu Rev Phytopathol* **38**, 293–324.

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.

Klaassen, V. A., Mayhew, D., Fisher, D. & Falk, B. W. (1996). *In vitro* transcripts from cloned cDNAs of the lettuce infectious yellows closterovirus bipartite genomic RNAs are competent for replication in *Nicotiana benthamiana* protoplasts. *Virology* **222**, 169–175.

Kreuze, J. F., Savenkov, E. I. & Valkonen, J. P. T. (2002). Complete genome sequence and analysis of the subgenomic RNAs of *Sweet potato chlorotic stunt virus* reveal several new features for the genus *Crinivirus*. *J Virol* **76**, 9260–9270.

Liu, X. & Gorovsky, M. A. (1993). Mapping the 5′ and 3′ ends of *Tetrahymena thermophila* mRNAs using RNA ligase-mediated amplification of cDNA ends. *Nucleic Acids Res* **21**, 4954–4960.

Livieratos, I. C. & Coutts, R. H. A. (2002). Nucleotide sequence and phylogenetic analysis of *Cucurbit yellow stunting disorder virus* RNA2. *Virus Genes* **24**, 225–230.

Livieratos, I. C., Katis, N. & Coutts, R. H. A. (1998). Differentiation between cucurbit yellow stunting disorder virus and beet pseudo-yellows virus by a reverse transcription-polymerase chain reaction assay. *Plant Pathol* **47**, 362–369.

Livieratos, I. C., Avgelis, A. D. & Coutts, R. H. A. (1999). Molecular characterization of the cucurbit yellow stunting disorder virus coat protein gene. *Phytopathology* **89**, 1030–1035.

Livieratos, I. C., Müller, G., Salazar, L. F., Eliasco, E. & Coutts, R. H. A. (2002). Identification and sequence analysis of *Potato yellow mosaic virus* capsid protein minor gene. *Virus Genes* **25**, 317–322.

Martelli, G. P., Agranovsky, A. A., Bar-Joseph, M. & 13 other authors (2002). The family *Closteroviridae* revised. *Arch Virol* **147**, 2039–2044.

Napuli, A. J., Falk, B. W. & Dolja, V. V. (2000). Interaction between HSP70-homolog and filamentous virions of beet yellows virus. *Virology* **274**, 232–239.

Napuli, A. J., Alzhanova, D. V., Doneanu, C. E., Barofsy, D. F., Koonin, E. V. & Dolja, V. V. (2003). The 64-kilodalton capsid protein homolog of *Beet yellows virus* is required for assembly of virion tails. *J Virol* **77**, 2377–2384.

Offei, S. K., Arciniegas, N., Müller, G., Guzmán, M., Salazar, L. F. & Coutts, R. H. A. (2004). Molecular variation of *Potato yellow vein virus* isolates. *Arch Virol* **149**, 821–827.

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Peng, C. W., Peremyslov, V. V., Mushegian, A. R., Dawson, W. O. & Dolja, V. V. (2001). Functional specialization and evolution of leader proteinases in the family *Closteroviridae*. *J Virol* **75**, 12153–12160.

Peng, C. W., Napuli, A. J. & Dolja, V. V. (2003). Leader proteinase of *Beet yellows virus* functions in long-distance transport. *J Virol* **77**, 2843–2849.

Peremyslov, V. V. & Dolja, V. V. (2002). Identification of the subgenomic mRNAs that encode the 6-kDa movement protein and Hsp70 homolog of *Beet yellows virus*. *Virology* **295**, 299–306.

Pleij, C. W. A., Rietveld, K. & Bosch, L. (1985). A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res* **13**, 1717–1731.

Rubio, L., Yeh, H.-H., Tian, T. & Falk, B. W. (2000). A heterogeneous population of defective RNAs is associated with *Lettuce infectious yellows virus* (LIYV). *Virology* **271**, 205–212.

Salazar, L. F., Müller, G., Querci, M., Zapata, J. L. & Owens, R. A. (2000). Potato yellow vein virus: its host range, distribution in South America, and identification as a crinivirus transmitted by *Trialeurodes vaporariorum*. *Ann Appl Biol* **137**, 7–19.

Satyanarayana, T., Gowda, S., Mawassi, M., Albiach-Marti, M. R., Ayllon, M. A., Robertson, C., Garnsey, S. M. & Dawson, W. O. (2000). Closterovirus encoded HSP70 homolog and p61 in addition to both coat proteins function in efficient virion assembly. *Virology* **278**, 253–265.

Satyanarayana, T., Gowda, S., Ayllon, M. A., Albiach-Marti, M. R. & Dawson, W. O. (2002). Mutational analysis of the replication signals in the 3′-nontranslated region of *Citrus tristeza virus*. *Virology* **300**, 140–152.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Tian, T., Rubio, L., Yeh, H.-H., Crawford, B. & Falk, B. W. (1999). Lettuce infectious yellows virus: *in vitro* acquisition analysis using partially purified virions and the whitefly *Bemisia tabaci*. *J Gen Virol* **80**, 1111–1117.

Valverde, R., Nameth, S. T. & Jordan, R. L. (1990). Analysis of double-stranded RNA for plant virus diagnosis. *Plant Dis* **74**, 255–258.

Wisler, G. C., Duffus, J. E., Liu, H.-Y. & Li, R. H. (1998). Ecology and epidemiology of whitefly-transmitted closteroviruses. *Plant Dis* **82**, 270–279.

Yeh, H.-H., Tian, T., Rubio, L., Crawford, B. & Falk, B. W. (2000). Asynchronous accumulation of *Lettuce infectious yellows virus* RNAs 1 and 2 and identification of an RNA 1 *trans* enhancer of RNA 2 accumulation. *J Virol* **74**, 5762–5768.

Zinovkin, R. A., Jelkmann, W. & Agranovsky, A. A. (1999). The minor coat protein of beet yellows closterovirus encapsidates the 5′ terminus of RNA in virions. *J Gen Virol* **80**, 269–272.