Arch Virol (2005) 150: 949–965
DOI 10.1007/s00705-004-0464-y

Archives of Virology
Printed in Austria

Complete nucleotide sequence and genome organization of Pelargonium line pattern virus and its relationship with the family *Tombusviridae*

A. Castaño and **C. Hernández**

Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC),
Universidad Politécnica de Valencia, Valencia, Spain

Received July 9, 2004; accepted November 10, 2004
Published online January 13, 2005

Summary. The complete nucleotide sequence of Pelargonium line pattern virus (PLPV) has been determined. The PLPV genomic RNA comprises 3884 nt and contains six open reading frames (ORFs) potentially encoding proteins of 27 (p27), 13 (p13), 87 (p87), 7 (p7), 6 (p6), and 37 kDa (p37), respectively. The arrangement of these ORFs on the PLPV genome closely resembles that of members of the genus *Carmovirus* in the family *Tombusviridae* and, moreover, most of the putative PLPV gene products showed high identity with proteins of this viral group. However, several striking differences were noticed. Carmoviruses generate two subgenomic RNAs whereas PLPV produces a single one. In addition, only p7 showed similarity with movement proteins of carmoviruses whereas p6 (as p13) has no viral (or other) homologs. This protein might be expressed from a non-canonical start codon or, alternatively, through a −1 frameshift (FS) mechanism. Both, the production of one subgenomic RNA and the likely involvement of a −1 FS for expression of an internal ORF parallel the translation strategies reported for the unique species of the genus *Panicovirus*, belonging also to the family *Tombusviridae*. Overall, the results support the placement of PLPV in this family although its peculiar characteristics preclude its direct assignment to any of the current genera.

The nucleotide sequence reported has been deposited in the GenBank databases under the accession number AY613852.

Introduction

Pelargonium line pattern virus (PLPV) has a broad experimental host range including more than 30 plant species from nine families, although reports on PLPV infections in nature are limited to geranium (*Pelargonium* ssp.). This virus may induce the appearance of yellow/green spots and line patterns on the leaves of the affected plants leading to significant quality deterioration [42, 43]. In Spain, a survey done on approximately one thousand geranium samples collected from different regions of the country revealed the presence of PLPV in more than 50% of the tested scions (M. Borja, personal communication). This situation is probably similar to that existing in other countries where this ornamental crop has been largely cultivated, as studies made in the nineties showed a high incidence of PLPV in the Netherlands and Israel [4, 11]. PLPV is transmitted by vegetative propagation and mechanical inoculation, and no vector has been identified.

PLPV virions are icosahedral in shape, about 30 nm in diameter and contain a linear positive sense single stranded RNA of approximately 4 kb [35, 44]. The taxonomic position of PLPV is unclear. It was considered as a tentative member of the genus *Carmovirus* (family *Tombusviridae*) on the basis of its morphological and physico-chemical properties [32], but the last report of the ICTV does not include PLPV within any established viral group or in the list of unassigned viruses [51]. This is most likely due to the poor characterization of PLPV and of other small isometric viruses infecting geranium which, as a consequence, has led to some controversies about their identity and to a confusing nomenclature [4, 18, 19, 27, 41, 43]. This situation underlines the need for additional information about these pathogens to facilitate their discrimination and to shed light on their phylogenetic relationships. A partial sequence corresponding to the PLPV coat protein gene has only recently become available at the sequence databases (Accession No. AY038067). In this work we report the entire primary structure of the PLPV genomic RNA and compare the genomic organization and the deduced amino acid sequences from its ORFs with those of related viruses. In addition, we have characterized the single PLPV subgenomic RNA detected in infected plants to gain an insight into the expression strategies of this virus. The results support the inclusion of PLPV in the family *Tombusviridae* but its assignment to genus level is still not unequivocal, and is discussed.

Material and methods

Virus source

PLPV isolate PV-0193 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany) and maintained by mechanical sap inoculation onto the experimental hosts *Chenopodium quinoa* and *Nicotiana clevelandii*.

RNA extraction and analysis

Total RNA preparations of healthy and infected *C. quinoa* and/or *N. clevelandii* leaves were obtained by phenol extraction and lithium precipitation [52]. Double stranded (ds) RNAs were purified from infected tissue by total nucleic acid extraction and subsequent partitioning

on nonionic cellulose (CF11, Whatman) with STE (50 mM Tris-HCl pH 7.2, 100 mM NaCl, 1 mM EDTA) containing 16% ethanol [33]. The dsRNAs were separated in non-denaturing 5% polyacrylamide gels and after ethidium bromide staining, the bands corresponding in size to the PLPV genomic and subgenomic RNAs were eluted and recovered by ethanol precipitation.

For Northern blot analysis, 4 μg of total RNA from healthy and PLPV infected leaves were denatured by glyoxal-dimethyl sulfoxide treatment, electrophoresed in 1% agarose gels, blotted to nylon membranes (Scheicher & Schuell), and cross-linked by UV irradiation. The membranes were hybridized with ^{32}P-radioactive DNA probes at 42 °C for at least 12 h in the presence of 50% formamide, 5 × SSC and 5% sodium dodecyl sulfate (SDS). The probes, prepared with Ready to go DNA labelling beads (Amersham Biosciences), corresponded to selected PLPV genome regions: a 5′ terminal probe which included from nt 1 to nt 181 of the PLPV genomic RNA, and a 3′ terminal probe which comprised from nt 3095 to 3884. After hybridization, membranes were washed for 15 min, twice at room temperature in 2 × SSC plus 0.1% SDS and once at 60 °C in 0.1 × SSC plus 0.1% SDS.

cDNA synthesis, cloning and sequencing

A partial PLPV cDNA clone (PL-1030) containing an uncharacterized PLPV cDNA ligated to the pUC9 vector, was provided by the DSMZ plant virus collection. Sequencing of the insert termini allowed the design of two PLPV specific oligonucleotides, CH17 (5′-GAAAATGGCCTTCTACGGGGAC-3′), homologous to nt 2067–2088 of the PLPV genomic RNA, and CH18 (5′-GCCTCCCTGCTGGCATAAACTAATAC-3′), complementary to nt 3578–3603. These primers were used for RT-PCR reactions with Titan One Tube RT-PCR system (Roche) employing total RNA preparations from PLPV infected leaves as templates. The amplified cDNA product of 1536 bp was inserted into the pGEM-T easy vector (Promega) and after determination of its sequence, new oligonucleotides were designed to fully characterize the primary structure of PLPV. Oligonucleotide CH32 (5′-CTGACCCCAAACGCCAAGTGGAATG-3′), complementary to nt 2156–2180, was used together with a degenerate primer, able to anneal to carmovirus related polymerase sequences (31), in RT-PCR reactions that yielded a cDNA of approximately 1000 bp. To obtain clones containing the 5′ and 3′ termini of PLPV genomic RNA, the corresponding dsRNA was denatured with 10 mM methylmercuric hydroxide and polyadenylated using yeast poly(A) polymerase (U.S. Biochemicals). After phenol–chloroform extraction and ethanol precipitation, the polyadenylated RNAs were subjected to reverse transcription (RT) with Superscript II-RT (Invitrogen) using primer GeneRacer Oligo dT (Invitrogen), which has a dT tail of 18 nt at its 3′ end and contains the priming sites for the GeneRacer 3′ and GeneRacer 3′ Nested oligonucleotides (Invitrogen) at 5′ end. RT products were PCR amplified with the Expand High Fidelity PCR System (Roche) using primers GeneRacer 3′ and GeneRacer 3′ Nested in combination with others derived from internal regions: CH43 (5′-TAGCTACTGGTTTCTGGAATTGGAG-3′) or CH46 (5′-TGGCCAACCCATGGA CGCCAATCCG-3′), complementary to nt 1378–1402 and nt 158–181, respectively, for the 5′ terminus, and CH36 (5′-ATAGAGGCGGAGCAACCAGGTAGTGC-3′) or CH53 (5′-CCTTAATGAAGAACGGGGGGGTAACC-3′), homologous to nt 3548–3573 and nt 3840–3865, respectively, for the 3′ terminus.

Complete cDNA clones of the PLPV subgenomic RNA were obtained using the single-primer method for cloning dsRNAs with unknown sequences [17, 54] with minor modifications. Briefly, the purified dsRNA was polyadenylated and reverse transcribed with primer GeneRacer Oligo dT. After treatment with RNase H (Roche), the cDNA complementary strands were annealed for 10 min at 80 °C, 16 h at 65 °C and 3 h at 30 °C and the resulting

hybrid was filled in with T4 DNA polymerase (Roche) and PCR amplified with Expand High Fidelity PCR System (Roche) and primer GeneRacer Oligo dT.

PCR products were separated by electrophoresis in 1% agarose gels or in non-denaturing 5% polyacrilamide gels, depending on the expected size. The DNA fragments were eluted and cloned into the pGEM-T easy vector (Promega) or the plasmid pTZ19R (Fermentas). At least two clones were selected for each DNA fragment and the nucleotide sequence of the inserts was determined with an ABI PRISM DNA sequencer 377 (Perkin-Elmer).

Primer extension assays

The 5′ terminal sequence of PLPV genomic RNA and the transcriptional start site of the subgenomic RNA were corroborated by primer extension assays. The substrates for the RT reactions were total RNA preparations from PLPV infected tissue in conjunction with primer CH46 (to map the 5′ end of the genomic RNA) or primer CH50 (5′-GCACCGA CAAGATCGCGAGATGTAC-3′), complementary to nt 2386–2410 (to map the 5′ end of the subgenomic RNA). RT reactions were performed with Superscript-RT II in the presence of (α-^{32})P(dCTP). The RT products were run in sequencing gels together with sequencing ladders obtained with the corresponding primer and proper recombinant plasmids with PLPV derived DNA inserts.

Sequence analysis

Analysis of the coding capacity of the nucleotide sequences was carried out with the TRANSLATE program (available at the ExPASy proteomics server; http://www.expasy.org/tools/dna.html). Database searches for nucleotide and protein similarities were done using the programs BLASTN, BLASTP and BLASTX (www.genome.ad.jp, www.ncbi.nlm.nih.gov, www.ebi.ac.uk). Pairwise comparisons of the PLPV nucleotide and amino acid sequences with those of related viruses were performed using the ALIGN program at the Genestream network (http://www2.igh.cnrs.fr/bin/align-guess.cgi) [34]. The RNA secondary structure of the PLPV 3′ region was predicted at 37 °C using MFOLD version 3.1 [30, 56]. Multiple alignments of protein sequences were generated with the CLUSTAL X program [47]. These alignments were used to conduct phylogenetic analysis employing programs of the MEGA 2.1 package (24). Sequences used for comparison were retrieved from the sequence databases and correspond to the following accession numbers: CarMV (*Carnation mottle virus*; X02986), TCV (*Turnip crinkle virus*, M22445), MNSV (*Melon necrotic spot virus*, M29671), CCFV (*Cardamine chlorotic fleck virus*, L16015), CPMoV (*Cowpea mottle virus*, U20976), SCV (*Saguaro cactus virus*, NC001780), GaMV (*Galinsoga mosaic virus*, Y13463), HCRSV (*Hibiscus chlorotic ringspot virus*, X86448), JINRV (*Japanese iris necrotic ring virus*, NC002187), PSNV (*Pea stem necrosis virus*, NC004995), PFBV (*Pelargonium flower break virus*, AJ514833), PoLV (*Pothos latent virus*, X87115), OCSV (*Oat chlorotic stunt virus*, X83964), CRSV (*Carnation ringspot virus*, L18870), RCNMV (*Red clover necrotic mosaic virus*, J04357), SCNMV (*Sweet clover necrotic mosaic virus*, L07884), MCMV (*Maize chlorotic mottle virus*, X14736), TNV-A (*Tobacco necrosis virus-A*, M33002), TNV-D (*Tobacco necrosis virus-D*, D00942), OLV-1 (*Olive latent virus 1*, X85989), LWSV (*Leek white stripe virus*, X94560), PMV (*Panicum mosaic virus*, U55002), TBSV (*Tomato bushy stunt virus*, M21958) AMCV (*Artichoke mottled crinkle virus*, X62493), CIRV (*Carnation Italian ringspot virus*, X85215), CuNV (*Cucumber necrosis virus,* M25270), CymRSV (*Cymbidium ringspot virus*, X15511), and PNSV (Pelargonium necrotic spot virus, AJ607402). The coat protein sequences of PelRSV (Pelargonium ringspot virus, AY038068), PCRPV (Pelargonium chlorotic ring pattern virus, AY038069), and ELV (Elderberry latent virus, AY038066) were also used for phylogenetic analysis.

Results

Nucleotide sequence and coding regions

The complete nucleotide sequence of PLPV and its deduced amino acid sequences have been determined. The PLPV genomic RNA consists of 3884 nt and contains six open reading frames (ORFs) (Fig. 1). The first ORF potentially encodes a protein of 27 kDa (p27). This ORF(p27) begins at AUG (nt 7–9) in a favourable translation context (CAAAAUGGCAC), with an A at position −3 and a G at position +4 (23, 28), and terminates at an UAG codon (nt 727–729). If this amber codon is read-through, the ORF extends to a downstream UAA (nt 2302–2304) giving rise to a protein of 87 kDa. Both p27 and p87 present homology with proteins of members of the family *Tombusviridae* (Table 1) which have been shown

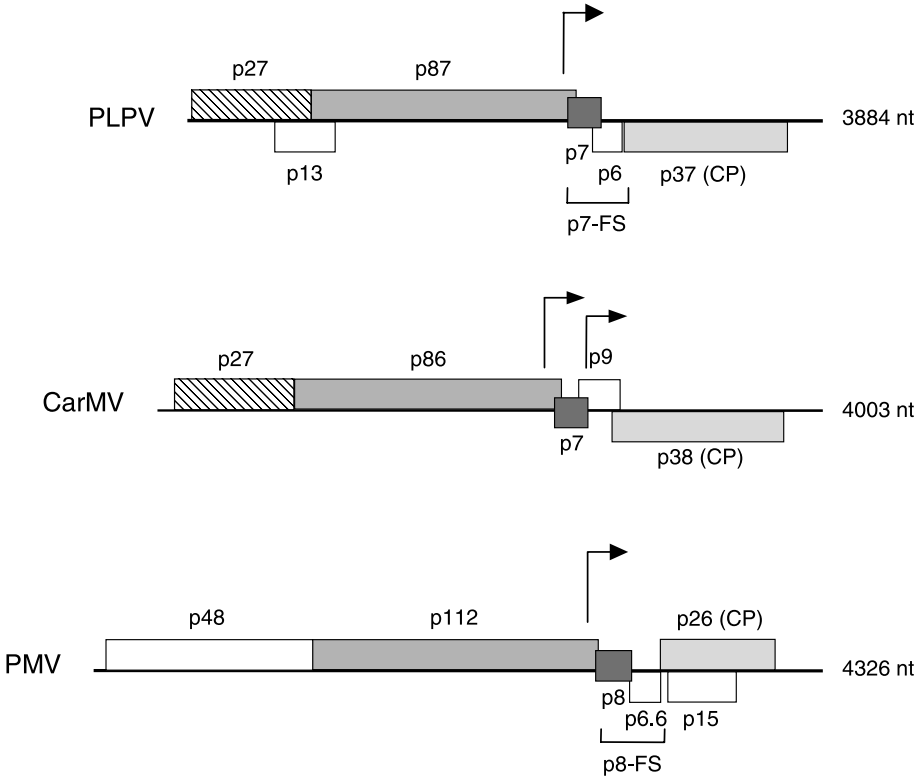

Fig. 1. Comparison between the genome organization of PLPV and those of the type species of the genera *Carmovirus* (*Carnation mottle virus*, CarMV) and *Panicovirus* (*Panicum mosaic virus*, PMV) in the family *Tombusviridae*. The ORFs are indicated by boxes which are similarly shaded when the encoded proteins share significant amino acid identity. The transcription start sites of the corresponding subgenomic RNAs are marked by arrows. Both p6 and p6.6 of PLPV and PMV, respectively, might be translated from non-canonical initiation codons. A −1 frameshift (FS) event could be required for the generation of the putative p7-FS and p8-FS proteins of PLPV and PMV, respectively. These potential frameshift products would have an N-terminus corresponding to p7 (for PLPV) or p8 (for PMV), and a C-terminus corresponding to almost the complete p6 (for PLPV) or the complete p6.6 (for PMV)

Table 1. Sequence identity (%) between the predicted PLPV proteins and those of viruses from the different genera of the family *Tombusviridae*[a]

Virus	Proteins				
	p27	p86	RT[b]	p7	CP
Carmoviruses					
CarMV	25.7	43.2	51.8	37.9	31.0
TCV	28.2	41.8	49.2	34.7	26.5
MNSV	24.5	39.8	47.6	27.5	24.6
CCFV	26.8	42.7	50.5	39.4	26.0
CPMoV	24.4	38.9	45.8	27.5	25.9
SCV	28.2	42.3	49.1	46.0	34.9
GaMV	21.5	35.8	42.6	23.2	23.1
HCRSV	28.2	44.0	51.3	38.6	27.6
JINRV	27.9	40.2	46.1	27.6	27.7
PSNV	25.8	39.8	46.2	25.4	20.5
PFBV	32.8	43.4	48.4	35.7	33.5
Aureusviruses					
PoLV	20.0	33.8	40.7	–[c]	24.3
Avenaviruses					
OCSV	21.9	32.0	36.5	18.9	20.6
Dianthoviruses					
CRSV	21.7	26.9	28.8	–	21.6
RCNMV	19.8	28.7	34.2	–	21.4
SCNMV	18.5	28.5	32.1	–	21.3
Machlomoviruses					
MCMV	18.0	33.4	46.2	22.1	18.8
Necroviruses					
TNV-A	26.0	38.0	44.2	19.4	17.5
TNV-D	21.3	33.7	40.0	25.4	18.2
OLV-1	26.4	38.5	44.4	26.7	17.2
LWSV	15.4	31.4	38.9	20.8	16.8
Panicoviruses					
PMV	17.2	32.5	44.0	27.0	16.7
Tombusviruses					
TBSV	19.2	33.3	41.2	–	24.1
AMCV	17.9	32.9	41.6	–	21.3
CIRV	19.0	32.4	41.1	–	18.9
CuNV	20.6	33.9	42.0	–	22.8
CymRSV	19.9	33.0	40.9	–	26.2
PNSV	20.2	33.2	41.2	–	24.5

[a]See the Methods section for viral abreviations and accession numbers of the sequences used for comparison

[b]Read-through portion of the RdRp

[c]Homologue is absent

to be involved in replication [1, 14, 55] and, moreover, p87 contains the motifs characteristic of the viral RNA dependent RNA polymerases (RdRp) including the highly conserved GDD box [22]. Another ORF, starting at AUG (nt 530–532, in the context CCGGAUGAAAA) and ending at UAA (nt 911–913), is nested between ORFs (p27) and (p87) and potentially encodes a protein of 13 kDa (p13) that is unique to this virus.

A small centrally located ORF begins at AUG (nt 2274–2276, in the context UCGAAUGGAUA) and stops at UGA (nt 2463–2465), and may encode a protein of about 7 kDa (p7) which exhibits high identity with movement proteins from species of the family *Tombusviridae*, especially with those of the genus *Carmovirus* (Table 1). In this PLPV putative product, a domain adopting a characteristic α-helix, similar to that able to bind RNA in the p7 of CarMV (29), can be predicted (data not shown). An additional small ORF, overlapping ORF(p7), potentially encodes a protein of 6 kDa (p6) which does not show clear similarity with any viral (or other) protein. This ORF might be expressed from a non-canonical start codon, AUU (nt 2441–2443, in the context UGGAAUUUGUG), and/or by a −1 frameshift (FS) mechanism which would lead to the production of a 12 kDa fusion protein (p12 or p7-FS). Supporting the latter possibility, an heptanucleotide is found immediately upstream of the stop codon of ORF(p7) whose sequence (Fig. 2) fits the canonical frameshifting motifs (5, 6). Moreover, a stem-loop is predicted downstream of the putative shifty site (data not shown), a type of structure which seems to favour frameshifting (reviewed in [12]).

Finally, the 3′ proximal ORF of PLPV, starting at AUG (nt 2621–2623, in the context GGAAAUGGCGG) and ending at UGA (nt 3635–3637), is predicted to encode a 37 kDa coat protein (p37 or CP), as indicates the high amino acid identity found between this product and proteins involved in encapsidation in the genus *Carmovirus* and in other genera of the family *Tombusviridae* (Table 1). The three different structural domains described in this type of proteins (39) can be also distinguished in the PLPV CP: (i) R, the N-terminal internal domain which

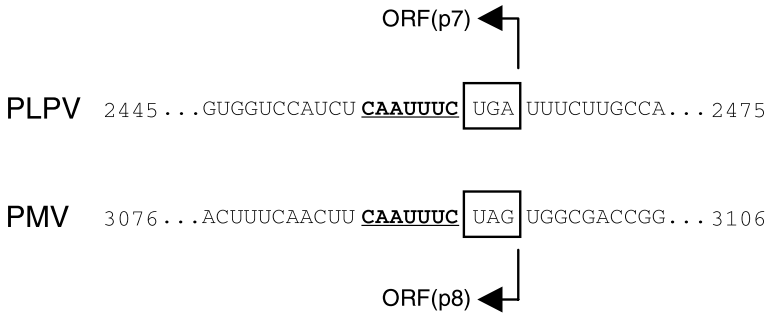

Fig. 2. Comparison of the nucleotide sequences of equivalent genome regions of PLPV and PMV. The heptanucleotide which fits the canonical frameshifting motifs [5] is shown in bold underlined letters. The stop codons of the PLPV ORF(p7) and PMV ORF(p8) are boxed

contains many positively charged residues and interacts with RNA, (ii) S, the shell domain which forms a barrel structure made up of β strands and constitutes the capsid backbone and, (iii) P, the protruding C-terminal domain.

When overlapping cDNA clones were compared, eighteen nucleotide substitutions were detected within coding regions. Only eight of these point mutations resulted in amino acid changes, five affecting the p27 and/or the p87 (two of them non-conservative, substitutions P → S and R → W at positions 40 and 336, respectively, of p87), two the p13 (both non-conservative, substitutions S → W and A → D at positions 64 and 86, respectively), and one the CP.

Comparison of nucleotide sequence of the CP gene of the PLPV isolate characterized in this work with that available at the databases showed an overall sequence identity of 93.7%. The corresponding protein sequences differed at nine amino acid positions including two non-conservative changes affecting the S and P domains (substitutions E → G and L → Q at positions 140 and 296, respectively).

5′ and 3′ untranslated regions of PLPV genomic RNA

The 5′ and 3′ terminal sequences of the PLPV genomic RNA were determined by sequencing of cDNA clones derived from these regions as indicated in the Methods section. This approach was complemented by primer extension assays (Fig. 3A) which confirmed that the 5′ untranslated region (UTR) of PLPV RNA comprises only 6 nt, to our knowledge, the shortest 5′ UTR described so far for the genomic RNA of a plant virus. The closest examples in this respect correspond to subgenomic RNAs encoding viral coat proteins as that of *Brome mosaic virus* (BMV) which has a 5′ UTR composed of only 9 nt [3]. It is also worth mentioning that the primer extension assays suggested that the PLPV genomic RNA lacks a 5′ terminal cap structure since the cDNA doublet normally associated with a capped RNA was absent [2, 13].

The 3′ UTR of PLPV RNA contains 247 nt. As observed in most viral species of the family *Tombusviridae*, the 3′ terminal sequence may fold into a stem-loop which is particularly stable in the case of PLPV because of the presence of six contiguous G–C pairs in the stem (Fig. 4). Studies with several representatives of the family *Tombusviridae* have shown that this type of structure is required for the efficient *in vivo* replication of the viral RNAs [7, 9, 10, 46, 50]. In addition, a midrange base-pairing interaction between the 5 nt in the 3′ terminus and a sequence located about 50 nt upstream, similar to that described recently as a replication silencer in *Tomato bushy stunt virus* and related viruses [36], can also be predicted for PLPV (Fig. 4).

Identification and characterization of a PLPV subgenomic RNA

As stated above, analysis of the coding capacity of PLPV predicts the synthesis of at least three proteins from downstream ORFs. To determine if subgenomic RNAs are generated during PLPV infection which could serve as mRNAs for the production of these proteins, total RNA from *C. quinoa* leaves infected by PLPV

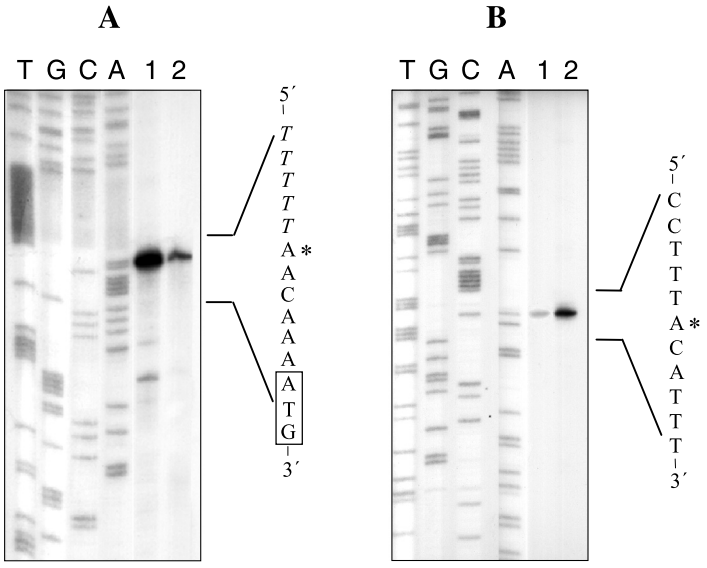

Fig. 3. Determination of the 5′ termini of the PLPV genomic (**A**) and subgenomic (**B**) RNAs by primer extension assays. Reverse transcriptions (RT) were performed using oligonucleotides complementary to nt 158–181 (**A**) or nt 2386–2410 (**B**) of the PLPV genome as primers and two distinct total RNA preparations (*1* and *2*) from *C. quinoa* PLPV infected leaves as templates. The RT products were run in sequencing gels together with sequencing ladders obtained with the same primers employed in the RT reactions and proper PLPV cDNA clones. The T tail derived from the cDNA cloning procedure is shown in italics. The 5′ terminal nucleotides of the genomic (**A**) and subgenomic (**B**) RNAs are indicated by asterisks, and the initiation codon of the ORF(p27) is boxed

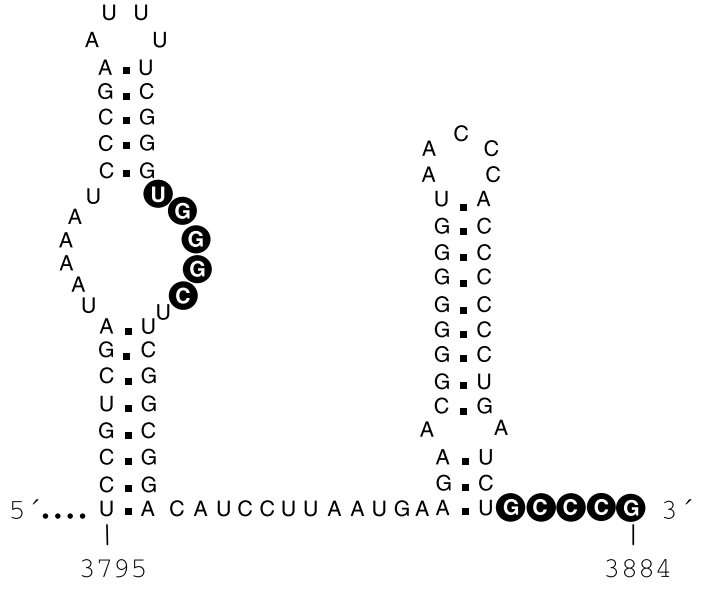

Fig. 4. MFOLD-predicted RNA secondary structure of the 3′-terminal region of PLPV. The nucleotides that could participate in a silencer 3′-terminus interaction are white in black circles. Numbers below correspond to positions in the genomic RNA

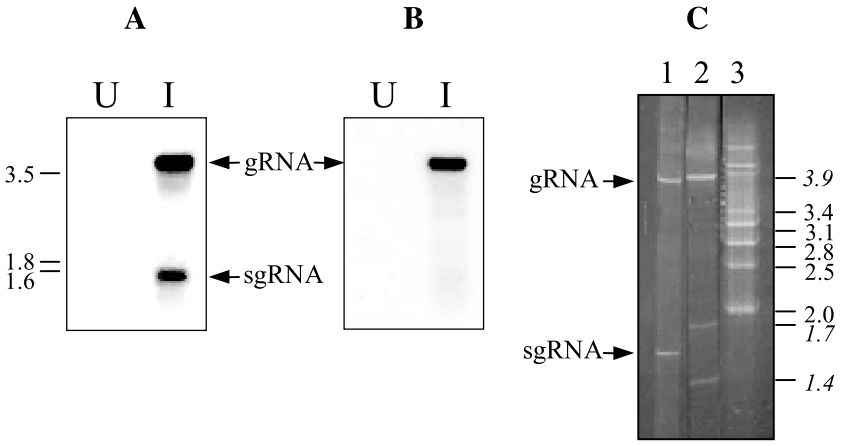

Fig. 5. Analysis of PLPV RNAs. **A** and **B**. Northern blot hybridizations were performed on total RNA preparations from uninfected (*U*) and PLPV-infected (*I*) *C. quinoa* leaves using ^{32}P-labeled probes derived from the 3′ (**A**) or 5′ (**B**) terminus of the PLPV genomic RNA. The positions of single stranded RNA molecular weight markers (in kb), corresponding to *in vitro* transcripts, are indicated at the left margin in **A**. **C** PAGE analysis in 5% non-denaturing polyacrilamide gels of PLPV dsRNAs (*1*). The dsRNAs of *Pelargonium flower break virus* (PFBV) and the mycoviral-like ones associated to Cherry chlorotic rusty spot disease (Covelli et al. and Coutts et al., in press) have been included for comparison (*2* and *3*, respectively) with the sizes (in kbp) of some of them shown at the right margin (in italics those corresponding to the dsRNAs of PFBV). In all panels, the bands corresponding to the PLPV genomic (*g*) and subgenomic (*sg*) RNAs are indicated with arrows

were analyzed by Northern blot hybridization with cDNA probes derived from different regions of the genomic RNA. When a 3′ terminal probe was used, a strongly hybridizing RNA species of about 1.6 kb was detected in infected tissue in addition to the genomic size RNA (Fig. 5A). Such 1.6 kb RNA corresponded very likely to a subgenomic RNA since a 5′ specific probe did not hybridize with it (Fig. 5B).

Analysis of dsRNA extracts from PLPV infected tissue revealed the presence of two dsRNA species of about 4.0 kbp and 1.6 kbp, in agreement with the results of the Northern blot (Fig. 5C). To determine the precise transcription start site of the 1.6 kb subgenomic RNA, the corresponding dsRNA was eluted from the gel, reverse-transcribed and cloned using the single-primer approach described in the Methods section. Sequencing of some of the clones showed that the 5′ terminus of the PLPV subgenomic RNA maps 23 nt upstream (position 2251 of the genomic RNA) from the initiation codon of ORF(p7). Primer extension analysis confirmed this result and suggested that such RNA lacks a 5′ terminal cap structure as seems to be the case for the genomic RNA (Fig. 3B). The molecular structure of the PLPV subgenomic RNA suggests that is generated for the translation of the small internal ORFs and the CP ORF having, therefore, the potential to encode at least three proteins.

Phylogenetic analysis

Multiple alignments with the sequences of different proteins of PLPV and those from members of the family *Tombusviridae* were generated and used for phylogenetic analysis. Comparisons based on the p27 and p7 homologs yielded tentative phylogenetic trees in which the distinction by genera was not clearly discernible since the viruses showed a tendency to form a continuum rather than to cluster (data not shown) in agreement with recent reports [8, 45]. When the p87 homologs were used to perform the analysis, the resulting tree exhibited a better correlation between grouping and genera (with the same exceptions reported in 45) and revealed a close association of PLPV with carmoviruses (Fig. 6A). In the analysis based on the CP sequences, those corresponding to three additional isometric viruses infecting *Pelargonium* spp., PelRSV, PCRPV and ELV, were included. The CP tree showed PLPV grouped with most carmoviruses but, remarkably, appeared tightly associated to PelRSV, PCRPV and ELV, three viral entities whose taxonomic position remains to be defined (see below).

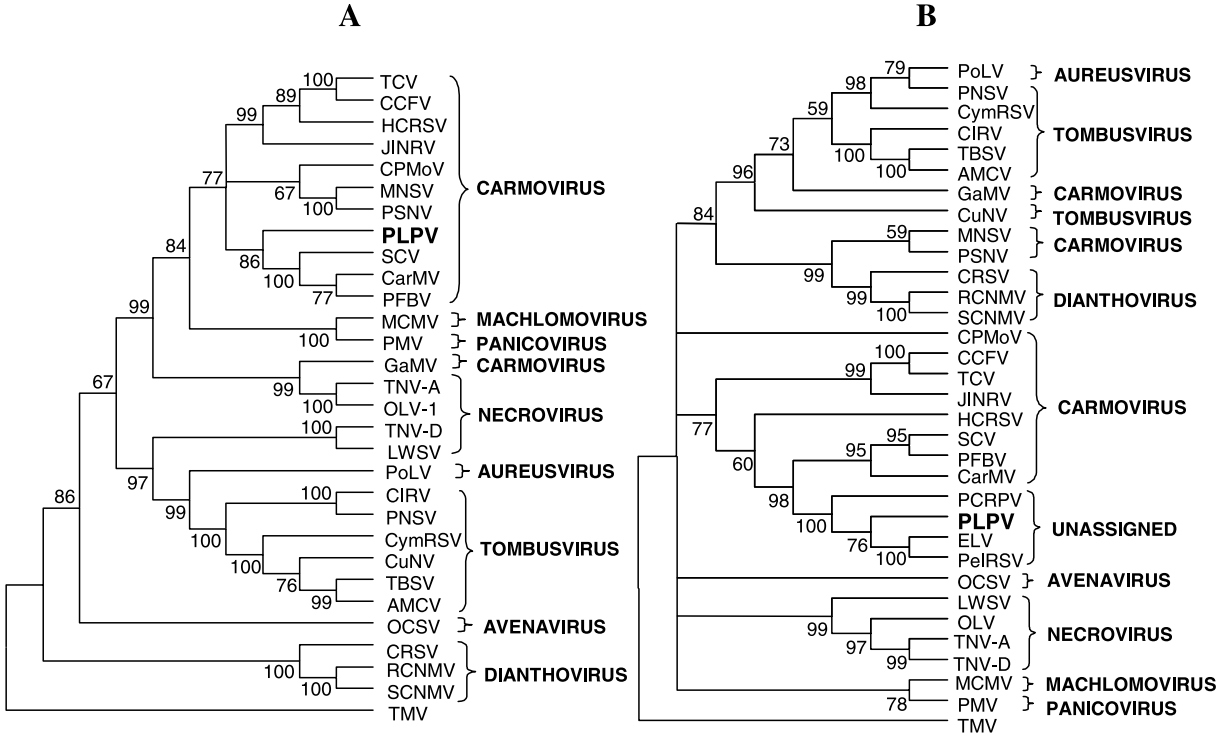

Fig. 6. Phylogenetic trees inferred from RNA dependent RNA polymerase (**A**) and coat protein (**B**) sequences of PLPV and members of the family *Tombusviridae.* Virus name abbreviations and accession numbers of the sequences used to construct the tree are indicated in the Methods section. Phylogenetic analysis was conducted with programs included into the MEGA 2.1 package. The numbers at the nodes indicate the number of times out of 10000 trees that this grouping occurred after bootstrapping the data; only values >50% are shown. The unrelated sequences of the corresponding proteins of *Tobacco mosaic virus* (TMV, accession number D38444) were used as outgroups

Discussion

We have determined the complete 3884 nt sequence of the genomic RNA of PLPV. The arrangement of ORFs in this virus resembles that of members of the genus *Carmovirus* and, moreover, phylogenetic analyses have shown that PLPV is closely related to this viral group. However, PLPV presents some distinctive features that prevent its direct assignment to this genus. The central region of the genomic RNA of carmoviruses contains two small out-of-frame ORFs whose translation gives rise to two proteins, p7 (or p8) and p9, involved in cell-to-cell movement [14, 25]. PLPV genome is not predicted to code for a product homologue of the second movement protein of carmoviruses and presents, instead, a small ORF which potentially encodes a polypeptide, p6, with no clear similarity to any viral protein. In addition, only one subgenomic RNA seems to be produced for translation of downstream PLPV ORFs, whereas carmoviruses generate two subgenomic RNAs for expression of the movement proteins and the CP, respectively [39].

The ORF(p13) is also unique to PLPV although its location on the viral genome is similar to that of the novel ORF(p23) of *Hibiscus chlorotic ringspot carmovirus* which encodes a protein indispensable for host specific replication [26]. Database searches have not demonstrated similarities between the putative PLPV p13 and any known protein and, moreover, computer-aided sequence analysis has not revealed any motif in p13 which could provide an insight about its function. Hence, the biological significance of p13, if produced in infected plants, is yet to be determined.

Caution is required when the possible translation events on RNA are deduced only from sequence data but should all proteins potentially encoded by PLPV be produced *in vivo*, the virus must use a variety of mechanisms for expression of its genes. Translation of PLPV ORFs (p27) and (p87) will probably occur from the genomic RNA involving a read-through mechanism in the case of ORF(p87), as demonstrated for related ORFs of members of the family *Tombusviridae* [16, 40, 48, 55]. Indeed, the sequence surrounding the amber stop codon of ORF(p27) is AAA-UAG-GGA, which is consistent with the consensus sequence required for efficient read-through AA(A/G)-UAG-G(G/U)(G/A) [16]. In the case of ORF(p13), an internal ribosomal entry is likely for its translation because the optimal sequence context of the initiation codon of ORF(p27) makes its expression via a leaky scanning mechanism unlikely [12]. The remaining downstream ORFs may be translated from the abundant 1.6 kb subgenomic RNA detected in PLPV infected plants. ORF(p7) would be expressed by conventional ribosomal scanning whereas ORF(p6) might be translated from a non-canonical start codon by leaky scanning and/or by a -1 FS mechanism giving rise to a p12 fusion protein (p7-FS). This also distinguishes PLPV from carmoviruses, whose small internal ORFs present canonical initiation codons and are presumably translated by a leaky scanning mechanism, with the exception of *Melon necrotic spot virus* for which a read-through event has been predicted [38]. While we cannot exclude the production of a subgenomic RNA accumulating to very low amounts for the CP gene's expression,

over exposure of the Northern blots did not reveal such potential RNA species. Thus, a leaky scanning mechanism or an internal ribosomal entry may account for its translation.

Interestingly, the arrangement and predicted expression strategies of the ORFs on the PLPV subgenomic RNA parallel those proposed for the unique subgenomic RNA of *Panicum mosaic virus* (PMV), the sole member of the genus *Panicovirus* in the family *Tombusviridae*, with the exception of an additional ORF overlapping the CP gene within PMV (Fig. 1) [48]. Both subgenomic RNAs have a region upstream of the CP gene devoid of AUG codons except for the one opening the 5′ proximal ORF which, moreover, has not an optimal translation context. Furthermore, the proposed "shifty" heptanucleotide of PLPV is identical in sequence and in an equivalent position to one found immediately upstream of the stop codon of the PMV ORF(p8) (Fig. 2), thus reinforcing the hypothesis of the production of a frameshift fusion protein. Attempts to detect the putative PMV p8-FS *in vivo* have failed however [49]. In contrast, *in vitro* translation assays have demonstrated that the non-canonical start codon of PMV ORF(p6.6) is active despite its suboptimal initiation context (CUAGUGG). Subsequent mutagenesis experiments have shown this ORF is required for efficient infection in plants [49]. Despite not share clear sequence similarity, the PMV p6.6 and the potential PLPV p6, also presumably translated from a non-canonical codon in a suboptimal context, do have similar hydrophobicity profiles and, moreover, resemble those of the p9 movement proteins of carmoviruses (data not shown). Only one membrane-spanning domain however is predicted in the PMV p6.6 and the PLPV p6 whereas the carmovirus proteins contain two transmembrane segments [53]. Bioassays with PMV mutants have suggested that the p6.6 plays a role in virus spread [49] and, considering the analogies found between the two proteins, an equivalent function can be anticipated for the PLPV p6.

A series of small isometric viruses naturally restricted to geranium has been described including PFBV, *Pelargonium leaf curl virus* (PLCV), PNSV, PelRSV, PCRPV, and PLPV. The characterization of these viruses has been mainly based on their biological/physicochemical properties and serological relationships, raising questions as to their taxonomy and status as distinct virus species [4, 18, 19, 27, 41, 43]. This has been further complicated by designating the same virus with different names, for example PLPV, which was formerly known as Pelargonium ring pattern virus [4]. Information on the genomic organization of these pathogens will undoubtedly shed light on this issue. The primary structures of PFBV, a carmovirus, and of PNSV, a proposed new species of the genus *Tombusvirus*, have been recently reported [15, 37]. In addition, some authors have determined the complete nucleotide sequence of other pelargonium viruses such as PCRPV, PelRSV and also PLPV [20] but have not made the data public. Recently, the genomic organization of PCRPV which potentially encodes five proteins very similar to those involved in replication (p27 and p87), movement (p7 and p9) and encapsidation (p37) in the genus *Carmovirus* has been described [21]. Interestingly, PCRPV resembles PLPV in producing a single subgenomic RNA

for expression of downstream ORFs and in the likely involvement of a −1 FS mechanism for translation of the second small internal ORF. It has been suggested that such traits are also shared by PelRSV and ELV [21] and, based on these common features and sequence similarities, it has been proposed to place the four viruses in a new genus within the family *Tombusviridae* for which the name *Pelarspovirus* has been advanced [21]. Only the CP sequences of PCRPV, PelRSV and ELV are available and show high identity with that of PLPV (44, 48 and 50%, respectively). Consistently, the phylogenetic tree based on CP sequences grouped the four pathogens together (Fig. 6). However, the limited sequence information has not allowed us to further analyze similarities and differences between these viruses and, therefore, commenting on the taxonomic proposal done.

To conclude, the results of this study support the classification of PLPV as a member of the family *Tombusviridae*. Although the PLPV genomic organization and the characteristics of its putative proteins indicate a close association with members of the genus *Carmovirus*, the uniqueness in the arrangement of some ORFs and proposed expression strategies hinder the direct assignment of the virus to this genus. The detailed description of the genomes of PCRPV, PelRSV and ELV, which resemble PLPV, will allow the establishment of more accurate relationships and the performance of phylogenetic analysis based on whole genome sequences as described recently [45]. We are currently attempting to elucidate the PLPV translation mechanisms and to ascertain the role of the novel ORFs found through the viral genome.

Acknowledgements

We thank Drs. R. Flores and V. Pallás for critical reading of the manuscript and suggestions, and Dr. M. Borja for valuable comments in the course of this work. This research was supported by grants AGL2000-0942 and AGL2003-04249 from the Ministerio de Ciencia y Tecnología de España and by grant Grupos 03/064 from the Generalitat Valenciana (to C. H.). A. C. was the recipient of predoctoral fellowships from the Generalitat Valenciana and from CSIC-Fundación Bancaja.

References

1. Andriessen M, Meulewaeter F, Cornelissen M (1995) Expression of *Tobacco necrosis virus* open reading frames 1 and 2 is sufficient for the replication of satellite *Tobacco necrosis virus*. Virology 212: 222–224
2. Ahlquist P, Janda M (1984) cDNA cloning and *in vitro* transcription of the complete *Brome mosaic virus* genome. Mol Cell Biol 4: 2876–2882
3. Ahlquist P, Luckow V, Kaesberg P (1981) Complete nucleotide sequence of *Brome mosaic virus* RNA 3. J Mol Biol 153: 23–38
4. Bouwen I, Maat DZ (1992) *Pelargonium flower-break* and Pelargonium line pattern viruses in the Netherlands; purification, antiserum preparation, serological identification, and detection in pelargonium by ELISA. Neth Plant Pathol 98: 141–156
5. Brierley I (1995) Ribosomal frameshifting on viral RNAs. J Gen Virol 76: 1885–1892
6. Brierley I, Jenner AJ, Inglis SC (1992) Mutational analysis of the "slippery sequence" component of a coronavirus ribosomal frameshifting signal. J Mol Biol 227: 463–479

7. Bringloe DH, Pleij CW, Coutts RH (1999) Mutation analysis of cis-elements in the 3′- and 5′-untranslated regions of satellite *Tobacco necrosis virus* strain C RNA. Virology 264: 76–84
8. Cañizares MC, Marcos JF, Pallás V (2001) Molecular variability of twenty-one geographically distinct isolates of *Carnation mottle virus* (CarMV) and phylogenetic relationships within the *Tombusviridae* family. Arch Virol 146: 2039–2051
9. Carpenter CD, Simon AE (1998) Analysis of sequences and predicted structures required for viral satellite RNA accumulation by *in vivo* genetic selection. Nucleic Acids Res 26: 2426–2432
10. Fabian MR, Na H, Ray D, White KA (2003) 3′-Terminal RNA secondary structures are important for accumulation of Tomato bushy stunt DI RNAs. Virology 313: 567–580
11. Franck A, Loebenstein G (1994) Virus and virus-like diseases of pelargonium in Israel. Acta Hortic 377: 31–39
12. Futterer J, Hohn T (1996) Translation in plants-rules and exceptions. Plant Mol Biol 32: 159–189
13. Goldberg K-B, Modrell B, Hillman BI, Heaton LA, Choi TJ, Jackson AO (1991) Structure of the glycoprotein gene of *Sonchus yellow net virus*, a plant rhabdovirus. Virology 185: 32–38
14. Hacker DL, Petty ITD, Wei N, Morris TJ (1992) *Turnip crinkle virus* genes required for RNA replication and virus movement. Virology 186: 1–8
15. Heinze C, Wobbe V, Lesemann DE, Zhang DY, Willingmann P, Adam G (2004) Pelargonium necrotic spot virus: a new member of the genus *Tombusvirus*. Arch Virol 149: 1527–1539
16. Huang M, Koh D, Weng L, Chang M, Yap Y, Zhang L, Wong S (2000) Complete nucleotide sequence and genome organization of *Hibiscus chlorotic ringspot virus*, a new member of the genus *Carmovirus*: evidence for the presence and expression of two novel open reading frames. J Virol 74: 3149–3155
17. Isogai M, Uyeda I, Hataya T (1998) An efficient cloning strategy for viral double-stranded RNAs with unknown sequences. Ann Phytopathol Soc Jpn 64: 244–248
18. Jones AT, McGavin WJ, Brunt AA, Phillips S (2000) Elderberry latent virus: its relationship to Pelargonium ringspot virus and its identification as a distinct member of the genus *Carmovirus*, family *Tombusviridae*. Ann Appl Biol 136: 147–152
19. Kinard GR, Jordan RL, Hurtt SS (1996) Partial characterization of Pelargonium line pattern and Pelargonium ringspot viruses. Acta Hortic 432: 148–155
20. Kinard GR, Jordan RL (1998) Genome organization of Pelargonium ringspot and Elderberry latent viruses. Phytopathology 88: S48
21. Kinard GR, Jordan R (2002) Genome organization of Pelargonium chlorotic ring pattern virus: further implications for *Tombusviridae* taxonomy. Acta Hortic 568: 17–27
22. Koonin EV, Dolja VV (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
23. Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283–292
24. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA 2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245
25. Li WZ, Qu F, Morris TJ (1998) Cell-to-cell movement of *Turnip crinkle virus* is controlled by two small open reading frames that function *in trans*. Virology 244: 405–416
26. Liang XZ, Lucy AP, Ding SW, Wong SM (2002) The p23 protein of *Hibiscus chlorotic ringspot virus* is indispensable for host-specific replication. J Virol 76: 12312–12319

27. Lisa V, Vaira AM, Dellavalle G, Masenga V, Milne RG (1996) Viruses of pelargonium in Italy. Acta Hortic 432: 108–117
28. Lütke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG initiation codons differs in plants and animals. EMBO J 6: 43–48
29. Marcos JF, Vilar M, Pérez-Payá E, Pallás V (1999) *In vivo* detection, RNA-binding properties and characterization of the RNA-binding domain of the p7 putative movement protein from *Carnation mottle carmovirus* (CarMV). Virology 255: 354–365
30. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288: 911–940
31. Morozov S, Ryabov E, Leiser R, Zavriev S (1995) Use of highly conserved motifs in plant virus polymerases as the tags for specific detection of carmovirus-related RNA-dependent RNA-polymerase genes. Virology 207: 312–315
32. Morris T, Carrington J (1988) *Carnation mottle virus* and viruses with similar properties. In: Koenig R (ed). The plant viruses, vol 3: Polyhedral virions with monopartite RNA genomes. Plenum Press, New York, pp 73–112
33. Morris TJ, Dodds JA (1979) Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. Phytopathology 69: 854–858
34. Pearson WR, Wood T, Zhang Z, Miller W (1997) Comparison of DNA sequences with protein sequences. Genomics 46: 24–36
35. Plese N, Stefanac Z (1980). Some properties of a distinctive isometric virus from pelargonium. Acta Hortic 110: 183–190
36. Pogany J, Fabian MR, White KA, Nagy PD (2003) A replication silencer element in a plus-strand RNA virus. EMBO J 22: 5602–5611
37. Rico P, Hernández C (2004) Complete nucleotide sequence and genome organization of *Pelargonium flower break virus*. Arch Virol 149: 641–651
38. Riviere CJ, Rochon DM (1990) Nucleotide sequence and genomic organization of *Melon necrotic spot virus*. J Gen Virol 71: 1887–1896
39. Russo M, Burgyan J, Martelli GP (1994) Molecular Biology of *Tombusviridae*. Adv Virus Res 44: 381–428
40. Scholthof KB, Scholthof HB, Jackson AO (1995) The *Tomato bushy stunt virus* replicase proteins are coordinately expressed and membrane associated. Virology 208: 365–369
41. Stone OM (1980) Nine viruses isolated from pelargonium in the United Kingdom. Acta Hortic 110: 177–182
42. Stone OM, Hollings M (1971) Pelargonium. Rep Glasshouse Crops Res Inst 1970, pp 153–155
43. Stone OM, Hollings M (1976) Pelargonium. Rep Glasshouse Crops Res Inst 1975, pp 119–120
44. Stone OM, Hollings M (1977) Pelargonium. Rep Glasshouse Crops Res Inst 1976, pp 121–122
45. Stuart G, Moffett K, Bozarth RF (2004) A whole genome perspective on the phylogeny of the plant virus family *Tombusviridae*. Arch Virol 149: 1595–1610
46. Stupina V, Simon AE (1997) Analysis *in vivo* of *Turnip crinkle virus* satellite RNA C variants with mutations in the 3′-terminal minus-strand promoter. Virology 238: 470–477
47. Thompson JD, Gibson TJ, Plewniak F, Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882

48. Turina M, Maruoka M, Monis J, Jackson AO, Scholthof KB (1998) Nucleotide sequence and infectivity of a full-length cDNA clone of *Panicum mosaic virus*. Virology 241: 141–155
49. Turina M, Desvoyes B, Scholthof KB (2000) A gene cluster encoded by *Panicum mosaic virus* is associated with virus movement. Virology 266: 120–128
50. Turner RL, Buck KW (1999) Mutational analysis of cis-acting sequences in the 3′- and 5′-untranslated regions of RNA 2 of *Red clover necrotic mosaic virus*. Virology 253: 115–124
51. Van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB (2000) Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses, Academic Press, San Diego
52. Verwoerd TC, Dekker BMM, Hoekema A (1989) A small-scale procedure for the rapid isolation of plant RNAs. Nucleic Acids Res 17: 2362
53. Vilar M, Sauri A, Monne M, Marcos JF, von Heijne G, Pérez-Payá E, Mingarro I (2002) Insertion and topology of a plant viral movement protein in the endoplasmic reticulum membrane. J Biol Chem 244: 23447–23452
54. Vreede FT, Cloete M, Napier GB, van Dijk AA, Viljoen GJ (1998) Sequence-independent amplification and cloning of large dsRNA virus genome segments by poly(dA)-oligonucleotide ligation. J Virol Methods 72: 243–247
55. White KA, Skuzeski JM, Li W, Wei N, Morris TJ (1995) Inmunodetection, expression strategy and complementation of *Turnip crinkle virus* p28 and p88 replication components. Virology 211: 525–534
56. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415

Author's address: Carmen Hernández, Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Universidad Politécnica de Valencia, Avenida de los Naranjos s/n, 46022 Valencia, Spain; e-mail: cahernan@ibmcp.upv.es