Population differentiation and selective constraints in Pelargonium line pattern virus

Aurora Castaño, Leticia Ruiz, Santiago F. Elena, Carmen Hernández*

Instituto de Biología Molecular y Celular de Plantas (Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia). CPI, Ed. 8E. Camino de Vera s/n, 46022 Valencia, Spain

* Corresponding author. Mailing address: Carmen Hernández, Instituto de Biología Molecular y Celular de Plantas, (Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia). CPI, Ed. 8E. Camino de Vera s/n, 46022 Valencia, Spain

Phone: +34 963 877 869 Fax: +34 963 877 859 E-mail: cahernan@ibmcp.upv.es

ABSTRACT

The genomic structure of *Pelargonium line pattern virus* (PLPV), a tentative member of a proposed new genus within the family *Tombusviridae*, has been recently determined. However, little is known about the genetic variability and population structure of this pathogen. Here, we have investigated the heterogeneity of PLPV isolates from different origins by sequence analysis of a 1817 nt fragment encompassing the movement (p7 and p9.7) and coat protein genes as well as flanking segments including the complete 3′ untranslated region. We have evaluated the selective pressures operating on both viral proteins and RNA genome in order to asses the relative functional and/or structural relevance of different amino acid or nucleotide sites. The results of the study have revealed that distinct protein domains are under different selective constraints and that maintenance of certain primary and/or secondary structures in RNA regulatory sequences might be an important factor limiting viral heterogeneity. We have also performed covariation analyses to uncover potential dependencies among amino acid sites of the same protein or of different proteins. The detection of linked amino acid substitutions has permitted to draw a putative network of intra- and interprotein interactions that are likely required to accomplish the different steps of the infection cycle. Finally, we have obtained phylogenetic trees that support geographical segregation of PLPV sequences.

Keywords: Pelargonium line pattern virus, family *Tombusviridae*, genetic variability, selective constraints, covariation analysis

1. **Introduction**

In general, RNA viruses are known to generate high levels of genetic variation that allow them to evolve rapidly facilitating their successful adaptation to new environments. The low fidelity of the viral encoded RNA dependent-RNA polymerases (RdRps), that may lack proofreading functions, has been proposed as the main underlying source for most variation (Agol, 2006; Castro *et al*., 2005; Holland *et al*., 1982; Sanjuán *et al.* 2010; Steinhauer and Holland, 1986). Nevertheless, the molecular composition of viral populations is not the direct result of the error rate of viral RdRps. In the case of plant RNA viruses, other factors play a major role in structuring genetic diversity of pathogen populations, including selection and genetic bottlenecks as those that occur during both systemic infection (French and Stenger, 2003; Li and Roossinck, 2004; Sacristán *et al*., 2003) and horizontal transmission by vectors (Ali *et al*., 2006; Betancourt *et al*., 2008). These other factors may lead to a considerable genetic stability as it has been reported for many different plant RNA viruses that appear more genetically stable than their animal counterparts (García-Arenal *et al*., 2001, 2003). This could be due to a combination of intrinsically lower rates of mutation, as suggested by recent and more accurate estimates (Malpica *et al*., 2002 Sanjuán *et al*., 2009; Tromas and Elena, 2010), and a reduced fixation rate of advantageous non-synonymous mutations because of weaker immune selection (García-Arenal *et al*., 2001). The identification and manipulation of factors that regulate the composition of the viral populations may offer a new set of tools to predict or control emerging diseases.

Pelargonium line pattern virus (PLPV) is a major geranium (*Pelargonium* spp.) pathogen in Spain, where prevalence rates above 50% have been reported (Alonso and Borja, 2005). The virus has also been detected in distinct European countries and it likely has a worldwide distribution (Bouwen and Maat, 1992; Franck and Loebenstein, 1994;

Stone, 1980). The frequent symptomless condition of PLPV infections (Alonso and Borja, 2005) compromise regulatory inspections and might have contributed to the spread of the virus. The factors that influence the appearance of symptoms, characterized by yellow-green spots and line patterns on the leaves, remain unclear but they are likely a combination of the viral isolate, the environmental conditions and the geranium cultivar.

PLPV virions are isometric in shape and hold a single stranded RNA molecule. Cloning and sequencing of genomic RNA (gRNA) together with reverse genetic experiments have recently allowed determination of the genome organization of PLPV (Castaño and Hernández, 2005, 2007; Castaño *et al*., 2009). The gRNA comprises 3883 nt and contains five open reading frames (ORFs) flanked by an unusually short untranslated region (UTR) of 6 nt at the 5´ end and by a 246 nt long UTR at the 3´ end. The two 5´-proximal ORFs encode proteins essential for replication, p27 and its read-through product p87 (the viral RdRp). Two small overlapping ORFs, located at the middle of the genome, encode proteins involved in viral movement (p7 and p9.7), while the 3´-proximal ORF encodes the coat protein (p37 or CP). The two replication proteins are translated directly from the gRNA whereas the movement and encapsidation proteins are translated from the unique PLPV subgenomic RNA (sgRNA) of 1.7 kb detected in infected tissue (Castaño *et al*., 2009).

PLPV taxonomic status has not been fully clarified yet. It was formerly considered as a tentative member of the genus *Carmovirus* but recent results supported its inclusion into a prospective new genus (*Pelarspovirus*) in the family *Tombusviridae* (Castaño and Hernández, 2005; Castaño *et al*., 2009; Stuart *et al*., 2006). Other tentative species of the prospective genus would be Pelargonium ringspot virus (PelRSV), Pelargonium chlorotic ring pattern virus (PCRPV) and Elderberry latent virus (ELV), that, as PLPV, produce only one sgRNA (Kinard and Jordan, 2002) in contrast with typical carmoviruses that generate two (Lommel *et al*., 2005). The distribution and importance of the two PLPV-

related, pelargonium-infecting viruses, PelRSV and PCRPV, is unknown as detection surveys for these pathogens are lacking. Nevertheless, the presence of PCRPV in at least two European countries has been recorded (Lisa *et al*., 1996; Ruiz *et al*., 2008) and PelRSV was reported to cause serious problems to geraniums in Germany (Lesemann and Adam, 1994).

The extent of PLPV variability remains to be ascertained as, so far, only the complete sequence of a German isolate, that from which the genomic organization of the virus was deduced (Castaño and Hernández, 2005; Castaño *et al*., 2009), and a partial sequence of an American isolate, corresponding to the CP gene (Accession No. AY038067), have been reported. Sequence information for other tentative members of the proposed genus *Pelarspovirus* is even scarcer. Indeed, only the primary structure of PCRPV genome has been fully determined while just the CP sequences of PelRSV and ELV are available.

In this work, we have studied the genetic variability among ten PLPV isolates recovered from naturally infected geranium plants which were collected in four countries at different times. We have obtained data that have allowed inferring selective constraints acting on PLPV genome and/or encoded products and that suggest geographical segregation of PLPV sequences. Additionally, covariation analyses have unveiled potential protein interactions and have pointed to particular amino acids as candidates to be involved in intra- and/or interprotein contacts.

2. Materials and methods

2.1. Viral isolates

Field PLPV isolates were obtained from geranium plants collected at distinct geographical locations over four years (2000-2004). Sap from the original plants was

used to pass the virus into the experimental host *C. quinoa* by mechanical inoculation and the viral population was recovered from this infected material. The isolates were designed with the first letters of the country of origin followed by a number to distinguish isolates from the same country (Table 1). PLPV isolate PV-0193, obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany) and characterized previously (Castaño and Hernández, 2005, 2007), was included in the sequence analyses for comparison purposes.

2.2. Reverse transcription, PCR amplification, cloning and sequencing

Total RNA preparations were obtained from infected *C. quinoa* leaves by phenol extraction and lithium precipitation (Verwoerd *et al.*, 1989) and used as templates for reverse transcription (RT) reactions with Superscript II-RT (Invitrogen) and primer CH60 (5´-CC*GGATCC*CGGGCAGATCAGGGGGGTGGGTTAC-3´), complementary to the 3´ terminus of the viral sequence (nt 3859–3883) with a SmaI site (underlined) and a BamHI site (in italics) at the 5´ terminus. RT products were PCR amplified with the Expand High Fidelity PCR System (Roche) and primers CH60 and CH17 (5´-GAAAATGGCCTTCTACGGGGAC-3´), homologous to nt 2067-2088 of the PLPV genome. After an initial denaturation step at 94 ºC for 2 min, PCR was performed for 35 cycles each of 30 sec at 95ºC, 30 sec at 60ºC and 3 min at 68ºC, followed by an extension step of 10 min at 68 ºC. The resulting RT-PCR products were separated by electrophoresis in 1% agarose gels, eluted and cloned into the pGEM-T easy vector (Promega) or the plasmid pTZ19R (Fermentas). Two clones for each DNA fragment were selected for sequencing with an ABI PRISM DNA sequencer 377 (Perkin-Elmer).

2.3. Sequence analysis

148

149 Multiple sequence alignments were constructed using MUSCLE (Edgard, 2004). For
150 coding regions, translated amino acid sequences were first aligned and used as guide to
151 built protein-coding nucleotide sequence alignments by concatenating codons using
152 PAL2NAL (Suyama *et al*., 2006). The best-fitting model of nucleotide substitution was
153 identified by MODELTEST (Posada and Crandall, 1998) as the general reversible GTR +
154 Γ_4 model, with the frequency of each substitution type and the gamma distribution of
155 among-site rate variation with four rate categories estimated from the empirical data.
156 Recombination was ruled out as a potential confounding factor by using GARD
157 (Kosakovsky Pond *et al.,* 2006) and RDP (Martin *et al.,* 2005) algorithms. A maximum
158 likelihood tree was constructed using the above model of nucleotide substitution and its
159 statistical significance was evaluated by bootstrap (upon 10,000 pseudoreplicates) using
160 PHYML (Guindon and Gascuel, 2003). MEGA4 (Tamura *et al.,* 2007) was used for
161 computing within- and among-population nucleotide diversities (standard errors were
162 computed by the bootstrap method based on 1000 pseudoreplicates) as well as for
163 performing Tajima's relative rates test (Tajima, 1993)

164

165 *2.3.1. Identification of adaptive evolution in PLPV genomes*

166

167 It is generally assumed that synonymous substitutions accumulate neutrally, at rate d_S
168 per synonymous site, because they have no effect on the amino acid composition of
169 proteins and, henceforth, may not affect protein folding and functioning. In contrast,
170 nonsynonymous substitutions, occurring at rate d_N per nonsynonymous site, involve
171 amino acid replacements and are more likely to affect, for bad or for good, the folding
172 and function of proteins. The intensity of selection, ω, can thus be evaluated as the ratio
173 $\omega = d_N/d_S$. Values of $\omega < 1$ indicate purifying (i.e., negative) selection that results in

elimination of detrimental mutations from virus populations. A value of $\omega = 1$ represents selective neutrality, i.e., mutations stay in the population at frequencies which are only governed by genetic drift. Finally, values of $\omega > 1$ are indicative of directional (i.e., positive) selection, resulting in fixation of advantageous mutations (Sharp, 1997). Here, we have used the several maximum likelihood Bayesian methods available in the HYPHY packaged (Kosakovsky Pond and Frost, 2005) as implemented in the www.datamonkey.org server. Each ORF was separately analyzed. Only sites identified by at least half of the six methods available will be reported.

2.3.2. Molecular covariation within and among proteins

Selection not necessarily acts on single amino acids but it may operate on groups of amino acids in a concerted manner. This being the case, changes in one amino acid should appear associated to changes in the other members of the interaction group. Two different methods were used to evaluate the presence of coevolving amino acids within any given protein. First the Bayesian graphical model implemented in SPIDERMONKEY (Poon *et al.,* 2007) and available online in the datamonkey server. Second, the mutual information content (*MIC*) approach described in Codoñer *et al.* (2006). For this second approach, significance *P*-values were computed, based on a million permutations, as the fraction of shuffles with a *MIC* value greater than or equal to the observed value. To minimize the number of false positives, the FDR method was applied (Benjamini and Hochberg, 1995). Only sites predicted to covary by both methods will be reported. Intermolecular covariation was only evaluated using the second methodology.

2.3.3. RNA secondary structure predictions

The secondary structure of selected regions of PFBV genome was predicted using MFOLD version 3.1 (www.bioinfo.rpi.edu/applications/mfold) (Mathews *et al.,* 1999; Zucker, 2003).

3. Results

3.1. Genetic diversity in coding and non-coding regions of PLPV genome

The primary structure of a total of 18 PLPV cDNAs, 1817 nt in length and derived from nine isolates (Table 1), was determined, and the resulting sequences were combined with the two additional ones previously characterized from a German isolate (Castaño and Hernández, 2005, 2007), producing a total dataset of 20 sequences which encompassed part of the RdRp gene, the complete p7, p9.7 and CP genes as well as the 3´ UTR.

Genetic distances between each pair of sequences ranged from 0.001 to 0.081. The maximum values were found between pairs of sequences from isolates sampled during different years at different countries (e.g., pair SPA3-USA1 or SPA3-ITA4), and the minimal values were detected for pairs of sequences of a given isolate, as they could differ by just 1 nt (e.g., sequences from isolate ITA2). The nucleotide diversity for the whole population was 0.055±0.007 (Table 2), which was similar to nucleotide diversity estimates of populations of other plant viruses (García-Arenal *et al.,* 2001). Nucleotide diversity calculated independently for each of the coding and non-coding regions included in the analysis ranged from 0.035±0.008 to 0.062±0.009, with the highest diversity corresponding to the CP gene and the lowest to the 3´ UTR (Table 2). Remarkably, the sequence of two genomic segments were strictly conserved in all

226 isolates, one encompassing nt 3642-3707 and corresponding to a 5′-portion of the 3′
227 UTR, and the other comprising nt 2240-2279 and matching the leader sequence of the
228 PLPV sgRNA plus short flanking stretches (Castaño and Hernández, 2005). Consistent
229 with the principle that transitions are biochemically more likely than transversions,
230 transition mutations were much more frequent than transversions, with the maximum
231 composite likelihood estimate of the overall transitions to transversions rates ratio being
232 3.116. This excess also occurs when purines (5.616) or pyrimidines (7.239) are
233 considered separately. A similar observation has been previously made for many other
234 viruses (Liang *et al*., 2002; Mansky and Temin, 1995; Rico *et al.,* 2006; Schneider and
235 Roossinck, 2001; Tromas and Elena, 2010; Vartanian *et al*., 1997).
236 Nucleotide diversity was also estimated between and within PLPV subpopulations,
237 considering a subpopulation as the group of isolates that were originally collected from a
238 given country (Germany, Spain, USA and Italy). Between subpopulation diversity values
239 considering either the complete 1817 nt region, individual ORFs or the 3′ UTR were
240 greater than within subpopulation diversity values (Table 2), suggesting that there is
241 significant differentiation of population according to the country from which the isolates
242 were sampled.
243 To gain a better insight into the relationships between all PLPV isolates, a maximum
244 likelihood phylogenetic tree was constructed from the nucleotide sequences included in
245 the study. The results revealed two major groups of PLPV sequences: group I included all
246 Spanish sequences and group II embraced sequences from Italy, Germany and USA
247 which were divided into three clusters according with their geographical distributions
248 (Fig. 1). When the analysis was performed with amino acid sequences deduced from any
249 of the individual genes, the same phylogenetic groups were defined though the statistical
250 significance of the internal nodes was, in general, lower than that obtained using the
251 complete nucleotide sequences (data not shown).

252

253 *3.2. Selective constraints on coding regions*

254

255 The direction and intensity of selective constraints operating in each coding region
256 was evaluated using the ω rates ratio statistic. The estimated average ω values were 0.109
257 (95% IC: 0.057-0.187) for the partial RdRp gene, 0.188 (95% IC: 0.112-0.292) for the p7
258 gene, 0.201 (95% IC: 0.131-0.293) for the p9.7 gene, and 0.096 (95% IC:0.073-0.123) for
259 the CP gene. Thus the ω ratio was significantly below one for the PLPV ORFs included
260 in the analysis, indicating that all of them are under purifying selection. We sought next
261 identifying which particular amino acid sites were under purifying or directional selective
262 constraints. To do so, ω was estimated for each position in the alignments. For the partial
263 RdRp, four sites were detected under negative selection (T711, L732, N760, I763). In the
264 case of p7, one site was detected to be under negative selection (S23) and another one
265 under positive selection (S5). Curiously, the latter site overlapped with one of those found
266 under negative selection in the RdRp (N760). Concerning p9.7, 10 amino acid sites were
267 found to be under negative selection and 27 sites were found in the case of CP. Such sites
268 were mainly concentrated in the central and N-terminal region of p9.7 and CP,
269 respectively (Fig. 2 and data not shown).

270

271 *3.3. Structural conservation of potential RNA regulatory sequences*

272

273 The PLPV genomic region under study (nt 2067-3883) contains at least two segments
274 that are presumed to play a key role in regulation of viral replication/transcription: the 3′
275 UTR and the promoter for synthesis of the PLPV sgRNA (in the minus strand). Though
276 the latter one has not been experimentally defined, it is expected to embrace an stretch of
277 about 100 nt preceding the initiation site of PLPV sgRNA (nt 2251; Castaño and

278 Hernández, 2005), in line with that reported for subgenomic promoters of related viruses
279 (Li and Wong, 2006; Wang and Simon, 1997; Wang *et al.,* 1999). Such promoters may
280 fold into hairpin-like structure and this type of conformation seems to be critical for the
281 mechanism of transcription of sgRNAs (Li and Wong, 2006; Wang *et al.,* 1999).
282 Secondary structure predictions showed that the putative PLPV subgenomic promoter
283 might also adopt a hairpin-like conformation with a small lateral branch. Interestingly,
284 the sequence variation detected in this segment when comparing isolates, essentially
285 maintained the predicted folding since most mutations were located in single stranded
286 regions or, when affecting double stranded regions, they were compensatory or located at
287 the base of loops or stems (Fig. 3A).

288 Concerning the 3´ UTR, it is expected to contain structural elements critical for viral
289 replication and, most probably, also for translation on the basis of that found in other
290 members of family *Tombusviridae* (Batten *et al*., 2006; Fabian and White, 2006; Fabian
291 *et al*., 2003; Pogany *et al*., 2003; Sarawaneeyaruk *et al*., 2009; Stupina *et al*., 2008;
292 Turner and Buck, 1999; Wang and Wong, 2004). The key role of the 3´ UTR during the
293 infectious cycle implies that strong constraints may operate on the region to preserve its
294 functionality that will likely depend on certain primary, secondary and/or tertiary RNA
295 structures. In line with this view, the 3´ UTR showed a value of genetic diversity that,
296 remarkably, was lower than those calculated for ORFs and, moreover, the nucleotide
297 sequence of a 5´-proximal segment of this region was strictly conserved in all isolates as
298 indicated above. *In silico* analysis showed that the 3´ UTR may fold into a series of stem-
299 loops that was basically conserved in the different variants (Fig. 3B). Remarkably, a 5´-
300 proximal stem-loop was formed by the conserved segment whereas the 3´-adjacent stem-
301 loop concentrated most of the heterogeneity found in the non-coding region (Fig. 3B).
302 Collectively, the results suggested that conservation of specific conformations in
303 regulatory sequences confer selective advantages to the viral RNA.

3.4. Variability in PLPV proteins

The sequence heterogeneity was unevenly distributed in the PLPV proteins. The C-terminal portion of the RdRp inferred from the amplified genomic region was not taken in consideration to analyze variability distribution as it represented only ~1/6 of the complete replication molecule. In the case of p7, 6 out of the 10 polymorphic positions detected were located in the first third of the protein (N-terminal 20 amino acids) despite the corresponding coding sequence overlapped in part with that of RdRp gene (data not shown). An amino acid replacement mapped at the putative RNA binding domain of p7 (V30I) but it did not affect basic residues which have been found essential for RNA binding in other carmoviruses (Marcos *et al*., 1999; Navarro *et al*., 2006). The variability of p9.7 also concentrated at the N-proximal half of the molecule as 12 out of the 17 polymorphic positions were detected within the N-terminal 40 amino acid residues of the protein (Fig. 2A), although most of the corresponding coding sequence overlaps with p7 gene. As reported for other members of the family *Tombusviridae* (Lommel *et al.,* 2005), three different structural domains can be distinguished in the PLPV CP: (i) R, the N-terminal internal domain which contains many positively charged residues and must interact 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. Analysis of the distribution of the heterogeneity in PLPV CP revealed that the percentage of polymorphic positions in the P domain (18.09%) was higher than in the R (10.95%) or S (13.12 %) domains. A stretch within the R domain was absolutely conserved in all isolates (from A14 to N46) probably because structural and/or functional constraints. Supporting the existence of such constraints, a high proportion of the sites predicted to be under negative selection were located in this stretch (Fig. 2B).

Next, as an additional test for the effect of selection, we analyzed the possible existence of covariation groups within and between proteins (Fig. 4). Firstly, we focused on covariations within-proteins. Regarding the partial RdRp, three amino acid residues showed significant covariation, S704N-E713A-S745A. In the case of p7, two covariation groups were detected S8T-V30I and S11I-L43I, whereas three covariation groups were observed for p9.7, Y3C-V6A, S16L-S39L and N24S-G88R (Fig. 4). The covariation that affected amino acids at positions 16 and 39 distinguished sequences from German isolate PV-0193, which showed the combination S16, S39, from those of the remaining isolates, that exhibited L residues at both positions. Up to eight covariation groups were detected for the CP, prominent among which was I10L-T64M, that differentiated the Spanish sequences (bearing the combination L10, M64) from those with other geographical origins (combination I10, T64). The analysis was extended to detect covarying positions between proteins. Remarkably, amino acid residues of p7 covaried with amino acid residues of the other three proteins included in the study. Thus, the p7 covariation S11I-L43I was significantly linked to RdRp substitution N760S and this linkage distinguished Italian isolates (with the combination RdRp S760, p7 I11, I43) from the remaining ones (with the combination RdRp N760, p7 S11, L43). In addition, the amino acid replacement V30I in p7 was linked to the amino acid replacement D32S in p9.7 and to the covarying group I10L-T64M in CP, and the corresponding combinations (p7 I30, p9.7 S32, CP L10, M64 *versus* p7 V30, p9.7 D32, CP I10, T64) segregated the Spanish isolate from the rest, further highlighting geographical distinctions between isolates. Finally, the covarying group S16L-S39L of p9.7 was linked to the amino acid substitution T236N of CP according to the programs employed, though visual inspection of alignments allowed to detect a strict association also with S213N/K (Fig. 2B). Indeed, the combination of S16, S39 in p9.7 and S213, T236 in CP was specific for the PV-0193

355 sequences whereas the remaining isolates showed the combination L16, L39 in p9.7 and
356 N/K213, N236 in CP (Fig. 2B).
357
358 **4. Discussion**
359
360 In this work, the extent and structure of genetic diversity in PLPV have been explored
361 by sequence analysis of an 1817 nt fragment (representing about 50% of the complete
362 viral genome) of ten viral isolates sampled from four distinct geographical areas.
363 Nucleotide diversity of the whole population was relatively low but in the range of those
364 estimated for other plant viruses (Fraile *et al*., 1996, 1997; García-Arenal *et al*., 2001;
365 Moya *et al.,* 1993; Rodríguez-Cerezo *et al*., 1991; Rubio *et al*., 2001; Font *et al.,* 2007).
366 Diversity values among isolates collected in the same country were lower than those
367 found among isolates of different countries suggesting a significant clustering of isolates
368 by country of origin, a postulate supported by phylogenetic analysis. As the sample size
369 of the present study is relatively small, characterization of new PLPV isolates is required
370 to confirm whether genotype distribution into phylogroups related with geographical
371 areas certainly reflects the structure of the viral population at global scale.
372 The relatively low genetic diversity found for PLPV suggested that negative selection
373 is restricting the number of molecular variants. Consistently, ω values estimated for
374 coding regions pointed to purifying selection as the predominant evolutionary pressure
375 operating on such genome segments likely to preserve the encoded amino acid sequences.
376 The smallest ω value was recorded for the CP, though it was similar to those of other
377 plant viruses (García-Arenal *et al*., 2001; Font *et al*., 2007). CP in related viruses has
378 been involved in other functions besides genome encapsidation, such as virus movement
379 and suppression of RNA silencing (Genovés *et al*., 2006; Martínez-Turiño and
380 Hernández, 2009; Thomas *et al*., 2003; Turina *et al.,* 2000), which could explain the

strong negative selection observed. Examination of selective constraints on particular amino acids showed, as expected, a strong bias among the number of sites under negative and positive selection (42 *versus* 1). Only sites under purifying selection were detected in the portion of the RdRp included in the analysis, despite such portion did not comprise any of the eight motifs conserved in the RdRps (Koonin and Dolja, 1993). Regarding p7, the unique negatively selected site was located in the putative RNA-binding motif of the protein but it did not correspond to any of the basic residues that are presumably critical for RNA-binding capability (Marcos *et al*., 1999; Navarro *et al*., 2006). In the case of p9.7, 9 out of the 10 negatively selected sites concentrated in the central part of the molecule (codons 29 to 54) which essentially matched the region that connects the two hydrophobic domains that, according to that reported for related proteins (Navarro *et al*., 2006; Saurí *et al.,* 2005; Vilar *et al*., 2002), must be involved in membrane association. Concerning the CP, almost 2/3 of the negatively selected sites (17/27) were located within the N-terminal 66 residues that constitute the R domain; however, only two of them corresponded to basic residues which are likely critical for RNA-binding capability suggesting that, as likely occurs in the case of p7, selection is acting on the preservation of the proper conformation of the RNA-binding motif. Moreover, the S and P domains showed identical number of negatively selected sites (5 each) despite the general trend to conservation of the former one in family *Tombusviridae* (Lommel *et al*., 2005).

On the other hand, the variability patterns found in regulatory sequences of the viral RNA, such as the putative subgenomic promoter or the 3´ UTR, support the existence of structural constraints that prevent the loss of their functionality. In the case of the subgenomic promoter, besides its predicted role in transcriptional regulation, it completely overlaps the 3´-portion of the RdRp gene (in the minus strand) and thus the same stretch is expected to have a dual function as coding and as regulatory sequence, which should considerably restrict heterogeneity. Accordingly, the mean of nucleotide

diversity in this region was lower than in other coding regions (Table 2) and, moreover, the nucleotide substitutions did not disrupt the predicted hairpin-like structure that is presumably required for the promoter function (Fig. 3A), suggesting that conservation of this conformation significantly influences the profile of naturally occurring mutations. Regarding the 3´ UTR, the variability data support that maintenance of a specific folding composed by a series of stem-loops might limit its sequence heterogeneity. Different members of family *Tombusviridae* have been reported to contain in this region *cis* elements critical for viral replication, such as promoters and repressors for minus strand synthesis (Na and White, 2006; Pogany *et al*., 2003; Stupina and Simon, 1997; Zhang *et al.,* 2004a, 2004b), and others relevant for gene expression, such as translational enhancers that promote cap independent translation, since the viruses of this family are characterized by non-blocked 5´-termini (reviewed by Kneller *et al*., 2006). An element of this type has been proposed to be present in the 3´ UTR of PLPV RNAs (Fabian and White, 2006) which, remarkably, would be embedded in the 5´-proximal stem-loop that is strictly conserved in all isolates (Fig. 3B) providing indirect evidence for its functional significance.

Population diversity studies with other members of the family *Tombusviridae* have also highlighted conservation of structural motifs in both regulatory RNA sequences and encoded proteins. Specifically, analysis of the genetic heterogeneity of *Carnation mottle virus* (CarMV) and *Pelargonium flower break virus* (PFBV), two members of the genus *Carmovirus*, has revealed that the pattern of natural variability preserves the conformation of putative replication *cis*-acting signals (Cañizares *et al*., 2001; Rico *et al*., 2006). Regarding proteins, different degrees of variation have been found when comparing equivalent products of PFBV, CarMV and PLPV though some common tendencies are noticeable. Among them, it is noteworthy mentioning the high conservation of the RNA binding motif of the small movement protein (or of the protein

itself in the case of PFBV) or the considerable sequence flexibility in the N-terminal region of the large movement protein.

Correlated amino acid mutation analysis has been widely used to infer functional interactions between different sites in a protein or between distinct proteins (e.g.: Altschuh *et al.*, 1987; Codoñer *et al*., 2006; Hoffman *et al*., 2003; Larson *et al.*, 2000; Thomas *et al*., 1996). The study of PLPV genetic variability has allowed identification of groups of amino acids that covary both within and between PLPV proteins, revealing a putative network of interactions that is likely needed for maintenance of proper protein folding and for driving the viral RNA from replication to cell-to-cell/systemic translocation. Though the exact mechanism that account for inter-cellular transport of carmo-like viruses is not yet known, it is not unlikely to require a physical interaction among CP and movement proteins as reported for other plant viruses (Akamatsu *et al*., 2007; Kim *et al.*, 2004; Liu *et al*., 2001; Sánchez-Navarro and Bol, 2001; Sánchez-Navarro *et al*., 2006; Takeda *et al*., 2004). Consistently with this view, we have detected covariations between the PLPV movement proteins, p7 and p9.7, and the CP. More intriguing is the covariation found between the RdRp and the p7. Evidence for interaction between the viral polymerase and the movement protein has been obtained for *Cucumber mosaic virus* which has led to the suggestion that both proteins cooperate in regulating the intercellular movement of progeny viral RNA by an unknown mechanism (Hwang *et al*., 2005), a possibility that could also apply to PLPV. Future work will be aimed at corroborating the inferred interactions and at assessing whether the covarying amino acids identified in this work are actually involved in protein-protein contact interfaces as suggested from the present results.

Acknowledgements

We are indebted to Dr. Jan van der Meij (Ball Flora Plant, Chicago) for providing PLPV isolate from USA, and to Dr. Marisé Borja (Fundación Promiva, Madrid) for the Spanish isolates and for valuable comments in the course of this work. We thank to Dr. Selma Gago (IBMCP, Valencia) for critical reading of the manuscript. We are also grateful to Dolores Arocas and Isabella Avellaneda for excellent technical assistance. This research was supported by grants BFU2006-11230 and BFU2009-11699 from Ministerio de Ciencia e Innovación (MICINN, Spain), ACOM09/040 from Generalitat Valenciana (to C.H.), and by grant BFU2009-06993 (MICINN) (to S.F.E.). A.C. was recipient of predoctoral fellowships from the Generalitat Valenciana and from CSIC-Fundación Bancaja and L.R. received a postdoctoral contract from the Juan de la Cierva program of MEC.

References

Agol, V.I., 2006. Molecular mechanisms of poliovirus variation and evolution, in: Domingo, E. (Ed.), Quasispecies: concepts and implications for virology. Current topics in microbiology and immunology, vol. 299. Springer, Heidelberg, Germany, pp. 211–259.

Akamatsu, N., Takeda, A., Kishimoto, M., Kaido, M., Okuno, T., Mise, K., 2007. Phosphorylation and interaction of the movement and coat proteins of *Brome mosaic virus* in infected barley protoplasts. Arch. Virol. 152, 2087-2093.

Ali, A., Li, H., Schneider, M.L., Sherman, D.J., Grey, S., Smith, D., Roossinck, M.J., 2006. Analysis of genetic bottlenecks during horizontal transmission of *Cucumber mosaic virus*. J. Virol. 80, 8345-8350.

Altschuh, D., Lesk, A.M., Bloomer, A.C., Klug, A., 1987. Correlation of co-ordinated amino acid substitutions with function in viruses related to *Tobacco mosaic virus*. J. Mol. Biol. 193, 693–707.

Alonso, M., Borja, M., 2005. High incidence of Pelargonium line pattern virus infecting asymptomatic *Pelargonium* spp. in Spain. Eur. J. Plant Pathol. 112, 95-100.

Batten, J.S., Desvoyes, B., Yamamura, Y., Scholthof, K.B., 2006. A translational enhancer element on the 3'-proximal end of the *Panicum mosaic virus* genome. FEBS Lett. 580, 2591-2597.

Benjamini Y, Hochberg Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B 57, 289-300.

Betancourt, M., Fereres, A., Fraile, A., García-Arenal, F., 2008. Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. J. Virol. 82, 12416-12421.

Bouwen, I., Maat, D.Z., 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.

Cañizares, M.C., Marcos, J.F., Pallás, V., 2001. Molecular variability of twenty-one geographically distinct isolates of *Carnation mottle virus* (CarMV) and phylogenetic relationship within the *Tombusviridae* family. Arch. Virol. 146, 2039-2051

Castaño, A., Hernández, C., 2005. Complete nucleotide sequence and genome organization of Pelargonium line pattern virus and its relationship with the family *Tombusviridae.* Arch. Virol. 150, 949-965.

Castaño, A., Hernández, C., 2007. Biological activity of transcripts from cDNA of Pelargonium line pattern virus. Acta Virol. 51, 271-274.

Castaño, A., Ruiz, L., Hernández, C., 2009. Insights into the translational regulation of biologically active open reading frames of Pelargonium line pattern virus. Virology 386, 417–426.

Castro, C., Arnold, J.J., Cameron, C.E., 2005. Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. Virus Res. 107, 141–149.

Codoñer, F.M., Fares, M.A., Elena, S.F., 2006. Adaptative covariation between the coat and the movement proteins of *Prunus necrotic ringspot virus*. J. Virol. 80, 5833-5840.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

Fabian, M.R, Na, H., Ray, D., White, K.A., 2003. 3'-Terminal RNA secondary structures are important for accumulation of *Tomato bushy stunt virus* DI RNAs. Virology 313, 567-580.

Fabian, M.R., White, K.A., 2006. Analysis of a 3´-translation enhancer in a tombusvirus: a dynamic model for RNA-RNA interactions of mRNA termini. RNA 12, 1304-1314.

Font, M.I., Rubio, L., Martínez-Culebras, V., Jordá, C., 2007. Genetic structure and evolution of natural populations of viruses causing the tomato yellow leaf curl disease in Spain. Virus Res. 128: 43-51

545 Fraile, A., Malpica, J.M., Aranda, M.A., Rodríguez-Cerezo, E., GarcÍa-Arenal, F., 1996.
546 Genetic diversity in tobacco mild green mosaic tobamovirus infecting the wild plant
547 *Nicotiana glauca*. Virology 223, 148–155.
548
549 Fraile, A., Escriu, F., Aranda, M.A., Malpica, J.M., Gibbs, A.J., García-Arenal, F., 1997.
550 A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*.
551 J. Virol. 71, 8316–8320
552
553 Franck, A., Loebenstein, G., 1994. Virus and virus-like diseases of pelargonium in Israel.
554 Acta Hortic. 377, 31-39.
555
556 French, R., Stenger, D.C., 2003. Evolution of *Wheat streak mosaic virus*: dynamics of
557 population growth within plants may explain limited variation. Annu. Rev.
558 Phytopathol. 41, 199-214.
559
560 García-Arenal, F., Fraile, A., Malpica, J.M., 2001. Variability and genetic structure of
561 plant virus populations. Annu. Rev. Phytopathol. 39, 157–186.
562
563 García-Arenal, F., Fraile, A., Malpica, J.M., 2003. Variation and evolution of plant virus
564 populations. Int. Microbiol. 6, 225–232.
565
566 Genovés, A., Navarro, J.A., Pallás, V., 2006. Functional analysis of the five melon
567 necrotic spot virus genome-encoded proteins. J. Gen. Virol. 87, 2371-2380.
568
569 Guindon, S., Gascuel, O., 2003. A simple, fast and accurate algorithm to estimate large
570 phylogenies by maximum likelihood. Syst. Biol. 52, 696-704.

571

572 Hoffman, N.G., Schiffer, C.A., Swanstrom, R., 2003. Covariation of amino acid positions
573 in HIV-1 protease. Virology 314, 536–548.

574

575 Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S, Van de Pol, S., 1982.
576 Rapid evolution of RNA genomes. Science 215, 1577–1585.

577

578 Hwang, M.S., Kim S.H., Lee J.H., Bae, J.M., Paek, K.H., Park, Y.I., 2005. Evidence for
579 interaction between the 2a polymerase protein and the 3a movement protein of
580 *Cucumber mosaic virus*. J. Gen. Virol. 86, 3171-3177.

581

582 Kim, S.H., Kalinina, N.O., Andreev, I., Ryabov, E.V., Fitzgerald, A.G., Taliansky, M.E.,
583 Palukaitis, P., 2004. The C-terminal 33 amino acids of the *Cucumber mosaic virus* 3a
584 protein affect virus movement, RNA binding and inhibition of infection and
585 translation. J. Gen. Virol. 85, 221–230.

586

587 Kinard, G.R., Jordan, R., 2002. Genome organization of Pelargonium chlorotic ring
588 pattern virus: further implications for *Tombusviridae* taxonomy. Acta Hortic. 568, 17-
589 27.

590

591 Kneller, E.L., Rakotondrafara, A.M., Miller, W.A., 2006. Cap-independent translation of
592 plant viral RNAs. Virus Res. 119, 63-75.

593

594 Koonin, E.V., Dolja V.V., 1993. Evolution and taxonomy of positive-strand RNA
595 viruses: implications of comparative analysis of amino acid sequences. Crit. Rev.
596 Biochem. Mol. Biol. 28, 375-430.

597

598 Kosakovsky Pond, S.L., Frost S.D.W., 2005. Not so different after all: a comparison of
599 methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208-
600 1222.

601

602 Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006.
603 Automated phylogenetic detection of recombination using genetic algorithm. Mol.
604 Biol. Evol. 23, 1891-1901.

605

606 Larson, S.M., Di Nardo, A.A., Davidson, A.R., 2000. Analysis of covariation in an SH3
607 domain sequence alignment: applications in tertiary contact prediction and the design
608 of compensating hydrophobic core substitutions. J. Mol. Biol. 303, 433–446.

609

610 Lesemann, D.E., Adam, G., 1994. Electron microscopical and serological studies on four
611 isometrical Pelargonium viruses. Acta Hortic. 377, 41-48

612

613 Li, H., Roossinck, M., 2004. Genetic bottlenecks reduce population variation in an
614 experimental RNA virus population. J. Virol. 78, 10582–10587.

615

616 Li, W., Wong, S.M., 2006. Analyses of subgenomic promoters of *Hibiscus chlorotic*
617 *ringspot virus* and demonstration of 5'-untranslated region and 3'-terminal sequences
618 functioning as subgenomic promoters. J. Virol. 80, 3395-3405.

619

620 Liang, X.Z., Lee, B.T.K., Wong, S.M., 2002. Covariation in the capsid protein of
621 *Hibiscus chlorotic ringspot virus* induced by serial passing in a host that restricts
622 movement leads to avirulence in its systemic host. J. Virol. 76, 12320-12324.

Lisa, V., Vaira, A.M., Dellavalle, G., Masenga, V., Milne, R.G., 1996. Viruses of pelargonium in Italy. Acta Hortic. 432, 108-117.

Liu, H., Boulton, M.I., Oparka, K.J., Davies, J.W., 2001. Interaction of the movement and coat proteins of *Maize streak virus*: implications for the transport of viral DNA. J. Gen. Virol. 82, 35-44.

Lommel, S.A., Martelli G.P., Rubino L., Russo, M., 2005. Family *Tombusviridae,* in: Fauquet C.M., Mayo M.A., Maniloff J., Desselberger U., Ball, L.A. (Eds.). Eighth Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, USA, pp. 907-936.

Malpica, J.M., Fraile, A., Moreno, I., Obies, C.I., Drake, J.W., García-Arenal, F., 2002. The rate and character of spontaneous mutation in an RNA virus. Genetics 162, 1505-1511.

Mansky, L.M., Temin, H.M., 1995. Lower *in vivo* mutation rate of *Human immunodeficiency virus* type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. 69, 5087-5094.

Marcos, J.F, 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.

649 Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and
650 analysis from sequence alignments. Bioinformatics 21, 260-262.
651
652 Martínez-Turiño, S., Hernández, C., 2009. Inhibition of RNA silencing by the coat
653 protein of *Pelargonium flower break virus*: distinctions from closely related
654 suppressors. J. Gen. Virol. 90, 519-525.
655
656 Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H., 1999. Expanded sequence
657 dependence of thermodynamic parameters improves prediction of RNA secondary
658 structure. J. Mol. Biol. 288, 911-940.
659
660 Moya, A., Rodríguez-Cerezo, E., García-Arenal, F., 1993. Genetic structure of natural
661 populations of the plant RNA virus tobacco mild green mosaic virus. Mol. Biol. Evol.
662 10, 449–456.
663
664 Na, H., White, K.A., 2006. Structure and prevalence of replication silencer-3´ terminus
665 RNA interactions in *Tombusviridae*. Virology 345, 305-316.
666
667 Navarro, J.A., Genovés, A., Climent, J., Saurí, A., Martínez-Gil, L., Mingarro, I., Pallás,
668 V., 2006. RNA-binding properties and membrane insertion of *Melon necrotic spot*
669 *virus* (MNSV) double gene block movement proteins. Virology 356, 57–67.
670
671 Pogany, J., Fabian, M.R., White, K.A., Nagy, P.D., 2003. A replication silencer element
672 in a plus-strand RNA virus. EMBO J. 22, 5602-5611.
673

674 Posada, D., Crandall, K., 1998. MODELTEST: testing the model of DNA substitution.
675 Bioinformatics 14, 817-818.
676
677 Poon, A.F.Y., Lewis, F.I., Kosakowsky Pond, S.L., Frost, S.D.W., 2007. An
678 evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-
679 1 envelope. PLoS Comp. Biol. 3, e231.
680
681 Rico, P., Ivars, P., Elena, S.F., Hernández, C., 2006. Insights into the selective pressures
682 restricting *Pelargonium flower break virus* genome variability and evidence of host
683 adaptation. J. Virol. 80, 8124-8132.
684
685 Rodríguez-Cerezo, E., Elena, S.F., Moya, A., García-Arenal, F., 1991. High genetic
686 stability in natural populations of the plant RNA virus tobacco mild green mosaic
687 virus. J. Mol. Evol. 32, 328–332.
688
689 Rubio, L., Abou-Jawdah, Y., Lin, H.X., Falk, B.W., 2001. Geographically distant isolates
690 of the *Crinivirus*, *cucurbit yellow stunting disorder virus* (CYSDV), show very low
691 genetic diversity in the coat protein gene. J. Gen. Virol. 82, 929–933.
692
693 Ruiz, L., Castaño, A., Borja, M., Hernández, C., 2008. Pelargonium chlorotic ring pattern
694 virus: first report in Spain. Plant Pathol. 57, 396.
695
696 Sacristán, S., Malpica, J.M., Fraile, A., García-Arenal, F., 2003. Estimation of population
697 bottlenecks during systemic movement of *Tobacco mosaic virus* in tobacco plants. J.
698 Virol. 77, 9906–9911.
699

Sánchez-Navarro, J.A., Bol, J. F., 2001. Role of the *Alfalfa mosaic virus* movement protein and coat protein in virus transport. Mol. Plant Microbe Interact. 14, 1051-1062.

Sánchez-Navarro, J.A., Herranz, M.C., Pallás, V., 2006. Cell-to-cell movement of *Alfalfa mosaic virus* can be mediated by the movement proteins of ilar-, bromo-, cucumo-, tobamo- and comoviruses and does not require virion formation. Virology 346, 66–73.

Sanjuán, R., Agudelo-Romero, P., Elena, S.F., 2009. Upper-limit mutation rate estimation for a plant RNA virus. Biol. Lett. 5, 394-396.

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belshaw, R. 2010. Viral mutation rates. J. Virol. 84, 9733-9748.

Sarawaneeyaruk, S., Iwakawa, H.O., Mizumoto, H., Murakami, H., Kaido, M., Mise, K., Okuno, T., 2009. Host-dependent roles of the viral 5' untranslated region (UTR) in RNA stabilization and cap-independent translational enhancement mediated by the 3' UTR of *Red clover necrotic mosaic virus* RNA1. Virology 391, 107-118.

Saurí, A., Saksena, S., Salgado, J., Jhonson, A.E., Mingarro, I., 2005. Double-spanning plant viral movement protein integration into the endoplasmic reticulum membrane is signal recognition particle-dependent, translocon-mediated, and concerted. J. Biol. Chem. 280, 25907-25912.

Schneider, W.L., Roossinck, M.J., 2001. Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. J. Virol. 75, 6566-6571.

726 Sharp, P.M., 1997. In search of molecular Darwinism. Nature 385, 111-112.
727
728 Steinhauer, D.A., Holland, J.J., 1986. Direct method for quantification of extreme
729 polymerase error frequencies at selected single base sites in viral RNA. J. Virol. 57,
730 219–228.
731
732 Stone, O.M., 1980. Nine viruses isolated from pelargonium in the United Kingdom. Acta
733 Hortic. 110, 177-182.
734
735 Stuart, G.W., Moffett, P.K., Bozarth, R.F., 2006. A comprehensive open reading frame
736 phylogenetic analysis of isometric positive strand ssRNA plant viruses. Arch. Virol.
737 151, 1159-1177.
738
739 Stupina, V., Simon, A.E., 1997. Analysis *in vivo* of *Turnip crinkle virus* satellite RNA C
740 variants with mutations in the 3′-terminal minus-strand promoter. Virology 238, 470-
741 477.
742
743 Stupina, V.A., Meskauskas, A., McCormack, J.C., Yingling, Y.G., Shapiro, B.A., Dinman,
744 J.D., Simon, A.E., 2008. The 3′ proximal translational enhancer of *Turnip crinkle virus*
745 binds to 60S ribosomal subunits. RNA 14, 2379–2393.
746
747 Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein
748 sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34,
749 W609-W612.
750

751 Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis.
752 Genetics 135, 599-607.
753
754 Takeda, A., Kaido, M., Okuno, T., Mise, K., 2004. The C-terminus of the movement
755 protein of *Brome mosaic virus* controls the requirement for coat protein in cell-to-cell
756 movement and plays a role in long distance movement. J. Gen. Virol. 85, 1751–1761.
757
758 Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary
759 genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.
760
761 Thomas, D.J., Casari, G., Sander, C., 1996. The prediction of protein contacts from
762 multiple sequence alignments. Protein Eng. 9, 941–948.
763
764 Thomas, C.L., Leh, V., Lederer, C., Maule, A.J., 2003. Turnip crinkle virus coat protein
765 mediates suppression of RNA silencing in *Nicotiana benthamiana*. Virology 306, 33–
766 41.
767
768 Tromas, N., Elena, S.F., 2010. The rate and spectrum of spontaneous mutations in a plant
769 RNA virus. Genetics 185, 983-989.
770
771 Turina, M., Desvoyes, B., Scholthof, K.B., 2000. A gene cluster encoded by *Panicum*
772 *mosaic virus* is associated with virus movement. Virology 266, 120-128.
773

774 Turner, R.L, Buck, K.W., 1999. Mutational analysis of *cis*-acting sequences in the 3'- and
775 5'-untranslated regions of RNA 2 of *Red clover necrotic mosaic virus*. Virology 253,
776 115-124.
777
778 Vartanian, J.P., Plikat, U., Henry, M., Mahieux, R., Guillemot, L., Meyerhans, A., Wain-
779 Hobson, S., 1997. HIV genetic variation is directed and restricted by DNA precursor
780 availability. J. Mol. Biol. 270, 139-151.
781
782 Verwoerd, T.C., Dekker, B.M.M., Hoekema, A., 1989. A small-scale procedure for the
783 rapid isolation of plant RNAs. Nucleic Acids Res. 17, 2362.
784
785 Vilar, M., Sauri, A., Monné, M., Marcos, J.F., von Heijne, G., Pérez-Payá, E., Mingarro, I.,
786 2002. Insertion and topology of a plant viral movement protein in the endoplasmic
787 reticulum membrane. J. Biol. Chem. 277, 23447-23452.
788
789 Wang, S., Carpenter, C.D., Simon, A.E., 1999. Minimal sequence and structural
790 requeriments of a subgenomic RNA promoter for *Turnip crinkle virus*. Virology 253,
791 327-336.
792
793 Wang, J., Simon, A.E., 1997. Analysis of the two subgenomic RNA promoters for *Turnip*
794 *crinkle virus in vivo* and *in vitro*. Virology 232, 174-186.
795
796 Wang, H.H., Wong, S.M., 2004. Significance of the 3'-terminal region in minus strand
797 RNA synthesis of *Hibiscus chlorotic ringspot virus*. J. Gen. Virol. 85, 1763-1776.
798

799 Zhang, G., Zhang, J., Simon, A.E., 2004a. Repression and derepression of minus-strand
800 synthesis in a plus-strand RNA virus replicon. J. Virol. 78, 7619-7633.
801
802 Zhang, J., Stuntz, R.M., Simon, A.E., 2004b. Analysis of a viral replication repressor:
803 sequence requirements for a large symmetrical internal loop. Virology 15, 90-102.
804
805 Zucker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction.
806 Nucleic Acids Res. **31,** 3406-3415.

807

808 **FIGURE LEGENDS**

809

810 **Fig. 1.** Minimum evolution unrooted phylogenetic tree inferred from nucleotide
811 sequences derived from 10 isolates of PLPV (see Table 1). Phylogenetic analysis was
812 conducted with programs included in the MEGA4 package. The numbers at the nodes are
813 bootstrap support values based on 10,000 pseudoreplicates; only values >50% are shown.

814

815 **Fig. 2.** Alignment of partial amino acid sequences of p9.7 (A) and CP (B) from different
816 isolates of PLPV. The reference sequence derived from isolate PV-0193 (Castaño and
817 Hernández, 2007) is shown at the top of the each alignment. Numbers above the
818 reference sequence correspond to positions in the complete protein. Those residues
819 conserved in all isolates are indicated by dots. The two amino acid sequences inferred
820 from two cDNA clones selected from each isolate are shown. Underlined residues are
821 under negative selection. In (A), the arrow demarcates the residues whose coding
822 sequence overlaps with that of p7 gene. The two predicted hydrophobic regions of the
823 protein are depicted within brackets at the bottom. In (B), the domains of the CP at which
824 the residues belong are indicated at the top.

825

826 **Fig. 3.** MFOLD-predicted RNA secondary structures of the putative subgenomic
827 promoter (A) and the 3´ UTR (B) of PLPV. The distribution of polymorphic positions is
828 indicated on the most stable folding of the reference sequence corresponding to isolate
829 PV-0193 (Castaño and Hernández, 2007). Numbers denote positions in the PLPV gRNA.
830 The minus strand is shown in (A).

831

832 **Fig. 4.** Covariations within and between p7, p9.7, CP and the partial RdRp. The residues
833 covarying within a given protein are connected by solid lines and those covarying among
834 proteins are connected by dashed lines.

835
836
837

Table 1. PLPV isolates used in this work

Isolate[a]	Country	Year(s)	Original Host	Accesion number
SPA0	Spain	2000	*P. zonale*	EU852912, EU852913
SPA1	Spain	2000	*P. zonale*	EU852914, EU852915
SPA3	Spain	2000	*P. zonale*	EU852916, EU852917
SPA4	Spain	2000	*P. zonale*	EU849616, EU852918
SPA7	Spain	2000	*P. zonale*	EU852919, EU852920
SPA8	Spain	2004	*P. zonale*	EU852921, EU852922
USA1	USA	2006	*P. peltatum*	EU852923, EU852924
ITA2	Italy	2004	*P. zonale*	EU852925, EU852926
ITA4	Italy	2004	*P. zonale*	EU852927, EU852928
PV-0193	Germany	1990s	*P. peltatum*	AY613852, EU835946

[a] PV-0193 has been previously characterized by Castaño and Hernández (2005; 2007)

Table 2. Nucleotide diversity ± SEM (based on 1000 bootstrap replicates). Isolates from Spain, Italy, USA and Germany have been considered as subpopulations

	Entire population π_T	**Mean within subpopulations** π_S	**Mean among subpopulations** $\pi_{ST} = \pi_T - \pi_S$	**Coefficient of differentation** $N_{ST} = \delta_{ST} / \pi_T$
RdRp	0.041 ± 0.008	0.010 ± 0.003	0.031 ± 0.007	0.757 ± 0.052
p7	0.054 ± 0.010	0.010 ± 0.003	0.044 ± 0.009	0.813 ± 0.036
p9.7	0.052 ± 0.008	0.011 ± 0.002	0.041 ± 0.007	0.789 ± 0.026
CP	0.062 ± 0.009	0.015 ± 0.002	0.047 ± 0.007	0.765 ± 0.017
3' UTR	0.035 ± 0.008	0.009 ± 0.002	0.026 ± 0.006	0.736 ± 0.046
Complete	0.055 ± 0.007	0.013 ± 0.002	0.042 ± 0.005	0.769 ± 0.012

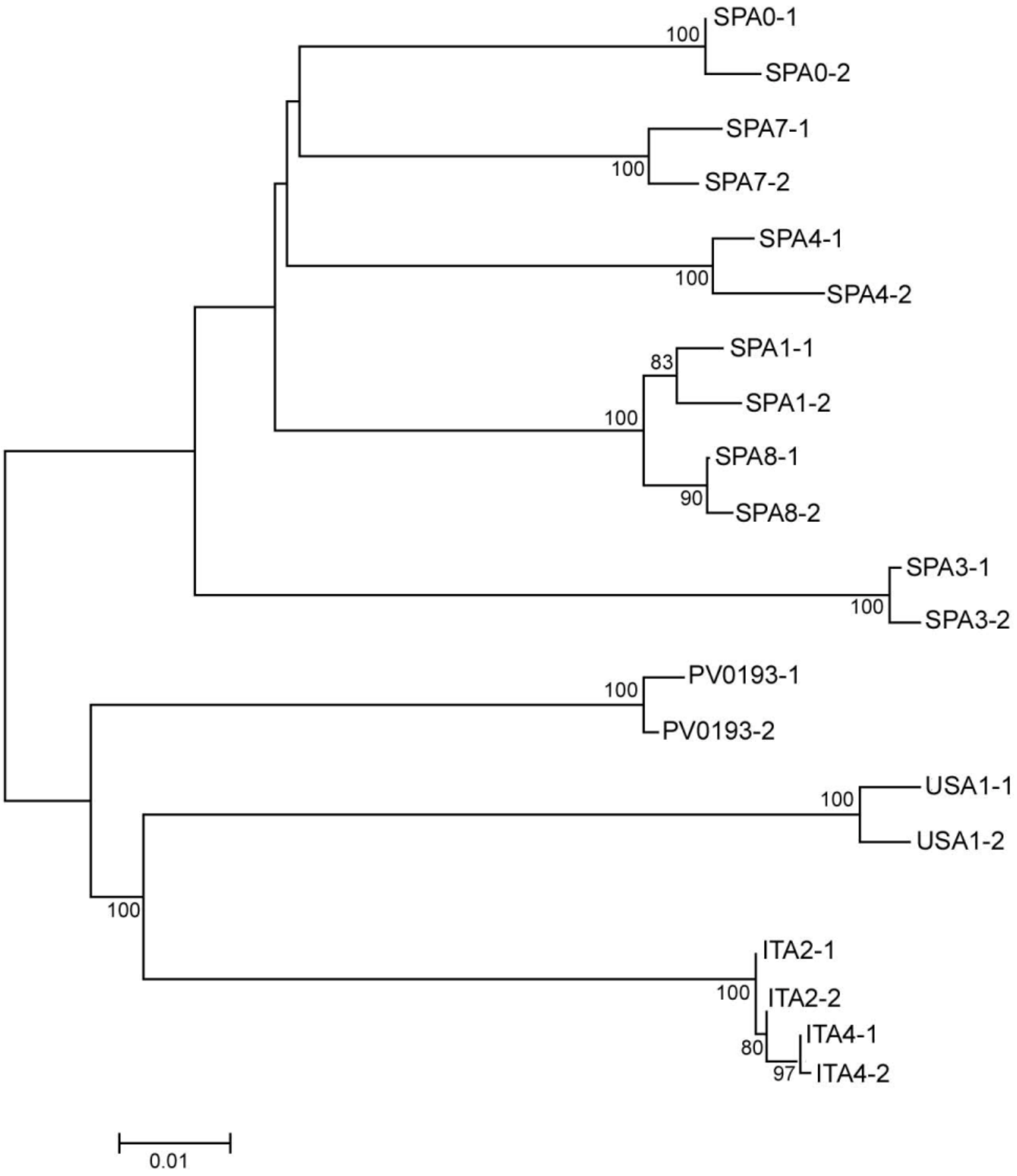
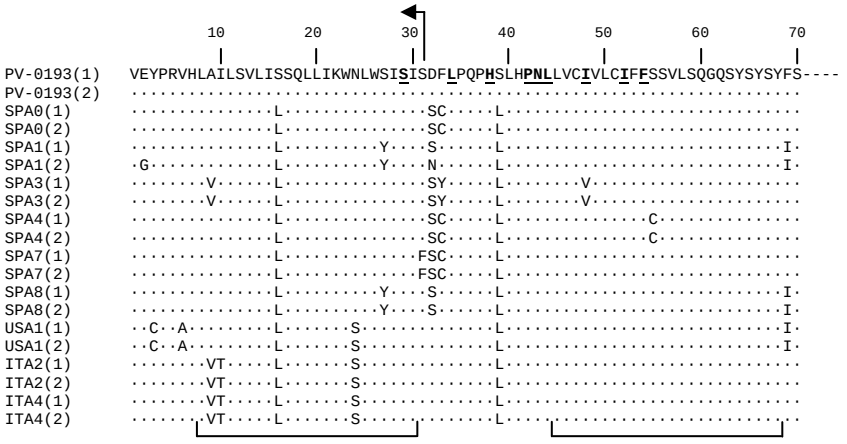

SPA0-1
100
SPA0-2
SPA7-1
100
SPA7-2
SPA4-1
100
SPA4-2
SPA1-1
83
SPA1-2
100
SPA8-1
90
SPA8-2
SPA3-1
100
SPA3-2
PV0193-1
100
PV0193-2
USA1-1
100
USA1-2
100
ITA2-1
100
ITA2-2
ITA4-1
80
97
ITA4-2
0.01

Fig. 1

(A)


10 20 30 40 50 60 70
PV-0193(1) VEYPRVHLAILSVLISSQLLIKWNLWSISISDFLPQPHSLHPNLLVCIVLCIFFSSVLSQGQSYSYSYFS----
PV-0193(2)
SPA0(1)
SPA0(2)
SPA1(1)
SPA1(2)
SPA3(1)
SPA3(2)
SPA4(1)
SPA4(2)
SPA7(1)
SPA7(2)
SPA8(1)
SPA8(2)
USA1(1)
USA1(2)
ITA2(1)
ITA2(2)
ITA4(1)
ITA4(2)

(B)

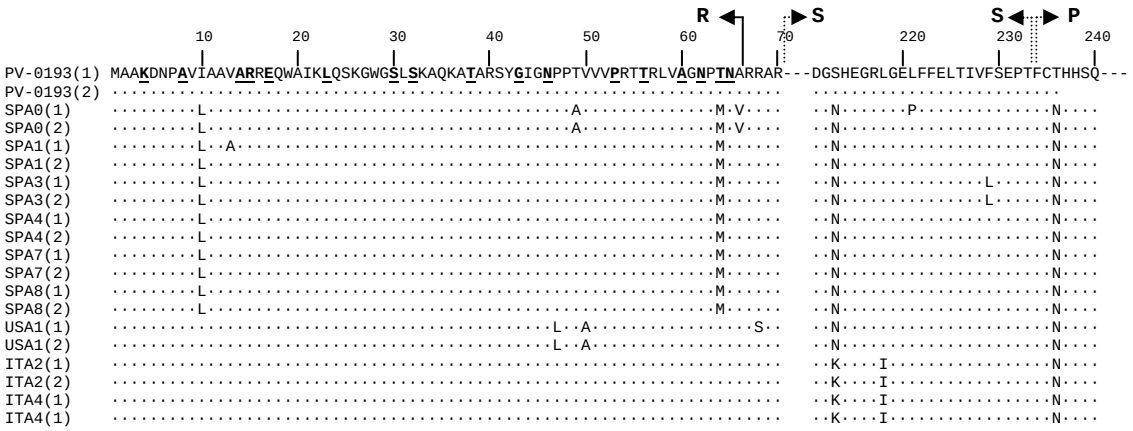

R
S
S
P
10 20 30 40 50 60 70 220 230 240
PV-0193(1) MAAKDNPAVIAAVARREQWAIKLQSKGWGSLSKAQKATARSYGIGNPPTVVVPRTTRLVAGNPTNARRAR---DGSHEGRLGELFFELTIVFSEPTFCTHHSQ---
PV-0193(2)
SPA0(1)
SPA0(2)
SPA1(1)
SPA1(2)
SPA3(1)
SPA3(2)
SPA4(1)
SPA4(2)
SPA7(1)
SPA7(2)
SPA8(1)
SPA8(2)
USA1(1)
USA1(2)
ITA2(1)
ITA2(2)
ITA4(1)
ITA4(2)

Fig. 2

(A)

(B)

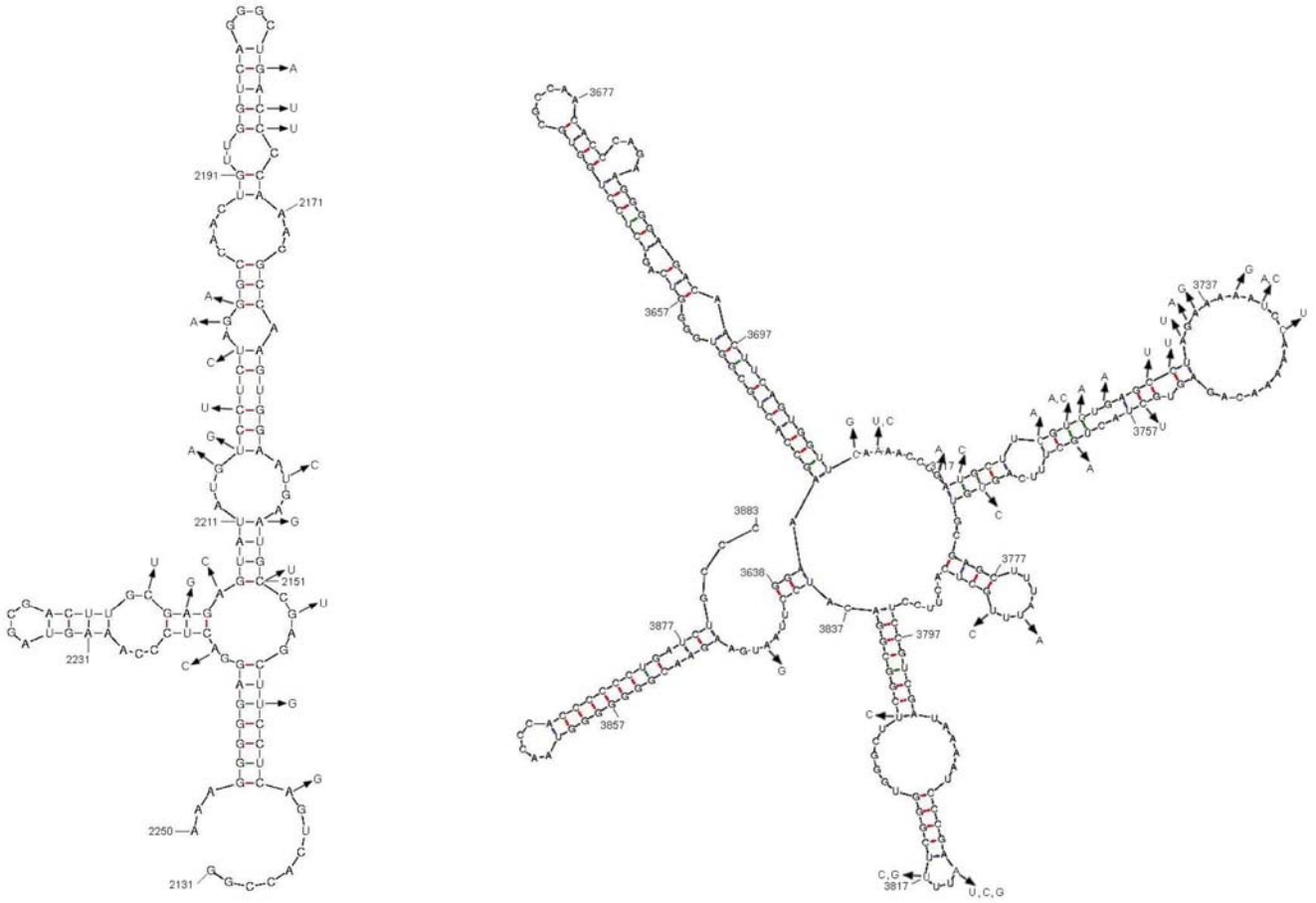

2191
2171
2211
2151
2231
2250
2131
3677
3657
3697
3737
3757
3883
3638
3777
3877
3837
3797
3857
3817
C,G
U,C,G

Fig. 3

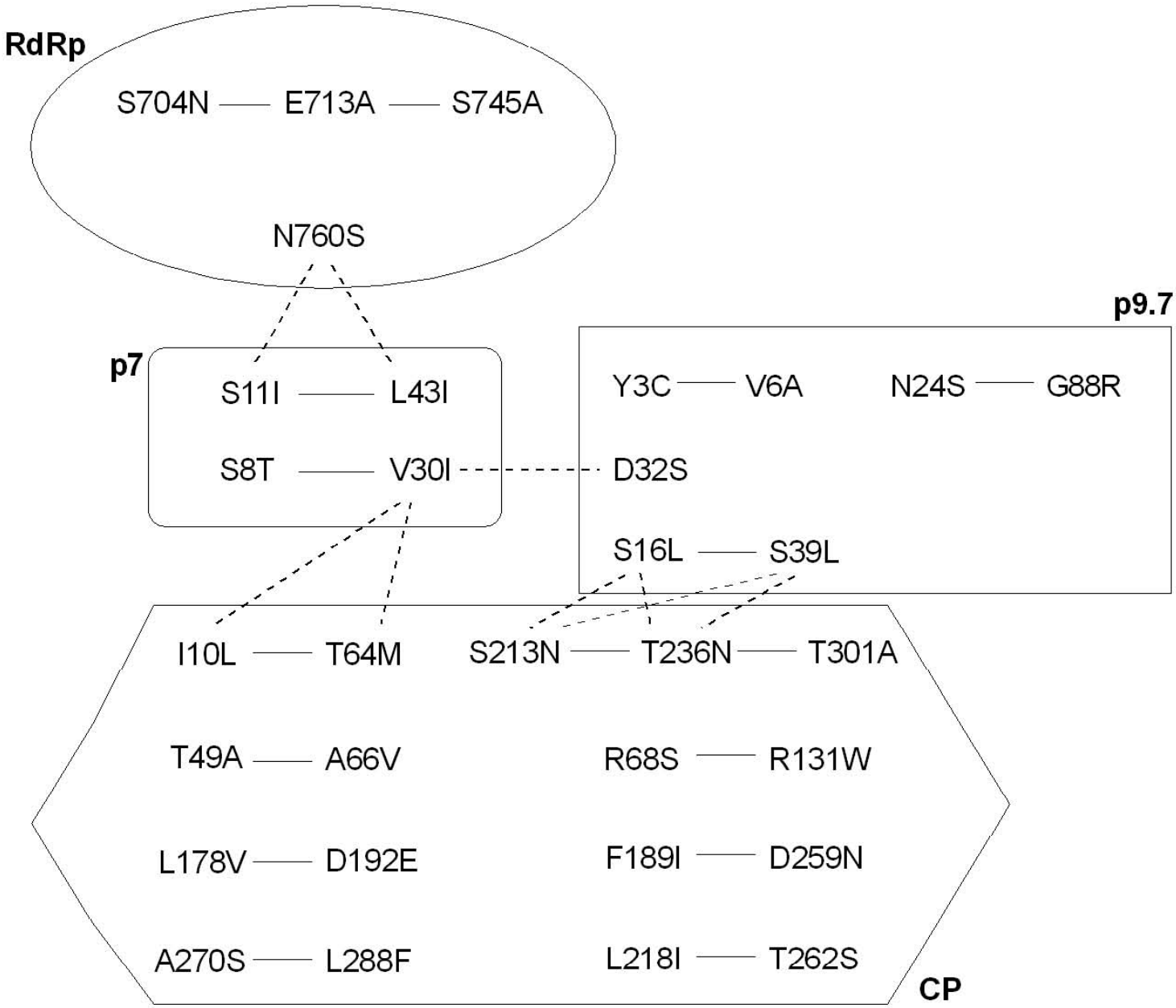

RdRp
S704N
E713A
S745A
N760S
p9.7
p7
S11I
L43I
S8T
V30I
Y3C
V6A
N24S
G88R
D32S
S16L
S39L
I10L
T64M
S213N
T236N
T301A
T49A
A66V
R68S
R131W
L178V
D192E
F189I
D259N
A270S
L288F
L218I
T262S
CP

Fig. 4