JOURNAL OF VIROLOGY, June 2006, p. 5833–5840 Vol. 80, No. 12
0022-538X/06/$08.00+0 doi:10.1128/JVI.00122-06

Adaptive Covariation between the Coat and Movement Proteins of Prunus Necrotic Ringspot Virus

Francisco M. Codoñer,[1] Mario A. Fares,[2] and Santiago F. Elena[1]*

Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-UPV, 46022 València, Spain,[1] *and Molecular Evolution and Bioinformatics Laboratory, Department of Biology, National University of Ireland, Maynooth, Ireland*[2]

Received 18 January 2006/Accepted 26 March 2006

The relative functional and/or structural importance of different amino acid sites in a protein can be assessed by evaluating the selective constraints to which they have been subjected during the course of evolution. Here we explore such constraints at the linear and three-dimensional levels for the movement protein (MP) and coat protein (CP) encoded by RNA 3 of prunus necrotic ringspot ilarvirus (PNRSV). By a maximum-parsimony approach, the nucleotide sequences from 46 isolates of PNRSV varying in symptomatology, host tree, and geographic origin have been analyzed and sites under different selective pressures have been identified in both proteins. We have also performed covariation analyses to explore whether changes in certain amino acid sites condition subsequent variation in other sites of the same protein or the other protein. These covariation analyses shed light on which particular amino acids should be involved in the physical and functional interaction between MP and CP. Finally, we discuss these findings in the light of what is already known about the implication of certain sites and domains in structure and protein-protein and RNA-protein interactions.

Prunus necrotic ringspot virus (PNRSV) is a positive-sense RNA plant virus with a tripartite genome that belongs to the genus *Ilarvirus* of the *Bromoviridae* family (14). PNRSV is distributed worldwide, infecting most cultivated *Prunus* spp. and causing symptoms that range from no damage to necrotic spots and shot holes in young leaves and to important rugose mosaic, all causing the plant to lose yield and vigor (51). PNRSV is easily transmittable by pollen grains and seeds and by routine propagation methods (17). Cherry PNRSV isolates were classically classified into three serotypes (CH3, CH9, and CH30) and two pathotypes (rugose mosaic disease and mild virosis) (34), and a correlation between symptom severity and the primary sequence of the gene for MP or CP was observed (19). However, similar studies using PNRSV isolates from other hosts and different geographical origins showed no clear correlation between molecular variability and biological properties (1, 53), indicating that the results derived from the cherry isolates were not necessarily valid for other hosts. Consensus, however, exists as isolates are classified on the basis of their sequence homology into three groups (PE-5, PV-32, and PV-96) (1, 18, 52). Group PV-32 is characterized by the presence of two repeated amino acids in the first third of CP coming from a duplication of two preceding ones (1). Group PE-5 contains three cherry isolates (two of them producing severe symptoms) of the CH30 serotype but also contains isolates from peach that induce mild and latent symptoms in infected trees (18). Attempts to find an association between geographical distribution and the abundance of certain serotypes have failed (1, 2, 18, 52).

Ilarviruses have the same genome organization, encoding products functionally similar to those of alfalfa mosaic virus (AMV) and other members of the *Bromoviridae* family. RNAs 1 and 2 encode the proteins involved in replication (P1 and P2, respectively). RNA 3 is bicistronic and encodes the movement protein (MP) and coat protein (CP) (35), although CP is translated from subgenomic RNA 4. Both MP and CP play important roles at different stages of the infectious cycle of ilarviruses. For example, the binding of CP into stem-loop structures flanked by AUGC sequences located in the 3′ nontranslated region of the inoculum RNAs is required for initiation of infection, a phenomenon known as genome activation (5, 6, 8, 24). PNRSV and AMV are phylogenetically close (10, 11, 42). Most studies on the implication of CP in the viral life cycle have been conducted with AMV, whereas there are relatively few experimental data on PNRSV CP structural properties and biological functions (3). For instance, AMV CP is required for plus-strand RNA accumulation, encapsidation, cell-to-cell movement, and systemic spread of the virus (49). This latter function requires the establishment of a physical interaction between CP and MP. It has also been suggested that, in AMV, MP may form tubular structures that traverse the cell wall through modified plasmodesmata, and these tubules mediate unidirectional transport of viral RNA-MP-CP complexes (41).

Tenllado and Bol (49) showed that it was possible to map the different AMV CP functions into different domains of the protein. The N-terminal arm looks sufficient for genome activation without the concourse of CP dimers, as demonstrated by using synthetic peptides of this protein region in in vivo experiments (4). Mutations in the N-terminal arm also affect cell-to-cell movement. The C-terminal domain, except the last 7 to 14 amino acids, is involved in dimer formation, which is required, for example, for correct virion formation. Amino acids 17 to 20 are critical for plus-strand RNA accumulation. Finally,

* Corresponding author. Mailing address: Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Avenida de los naranjos s/n, 46022 València, Spain. Phone: 34 963 877 895. Fax: 34 963 877 859. E-mail: sfelena@ibmcp.upv.es.

mutations at the N and C termini affect the movement of viral materials through the vascular system (49).

The CP N-terminal domain has been shown to be involved in cell-to-cell movement and in systemic spread in members of the different genera within the family *Bromoviridae* (AMV [41], brome mosaic virus [BMV] [16, 38, 39, 40], and cucumber mosaic virus [CMV] [25]). Furthermore, Sánchez-Navarro and Bol (41) also showed that deletions in the N- or C-terminal arm of AMV MP did not interfere with the ability of the protein to assemble into tubules and allow cell-to-cell and systemic movements. However, simultaneously deleting the 11 N-terminal amino acids and 45 amino acids at the C terminus produced proteins able to form tubules but not proficient in promoting cell-to-cell movement. Furthermore, recently Sánchez-Navarro et al. (43) have shown that the 44 amino acids at the AMV MP C terminus physically interact with CP to elicit movement. Apparently, the C terminus of AMV MP confers specificity on the transport process (41) as it does in the cases of BMV (47) and CMV (26). MPs of PNRSV, AMV, CMV, and BMV have been shown to have an RNA-binding domain located at the N terminus of the protein in the first two cases or at the C terminus in the last two (3, 4, 7, 21). At least in the case of PNRSV, it has been shown that basic amino acids in this domain are required for RNA-binding activity and cell-to-cell movement (21). Finally, it has been reported that CMV MP also interacts with the 2a polymerase gene, although this interaction is not relevant for the formation of the replicase 1a-2a complex (23).

Although it is widely accepted that MP and CP should interact to facilitate viral cell-to-cell movement, direct evidence has not been provided. Here we report results from a molecular evolution study seeking to identify the selective constrains acting upon CP and MP during the evolution of different isolates of PNRSV. Our aim is twofold. At the one side, using a maximum-parsimony approach, we have identified amino acid sites under selection in each protein at different evolutionary time points. At the other side, we have applied statistical methods drawn from information theory to identify groups of amino acids that covary throughout time as a consequence of their functional or structural relationship. The existence of covariation groups was explored both within and between proteins.

TABLE 1. PNRSV isolates included in this study

Isolate	Country	Host	Accession no.	
			MP	CP
1-13	Czech Republic	Cherry		AF170156
21-1	Czech Republic	Cherry		AF170157
30-4	United States	Peach	U57046	
4-8	Czech Republic	Cherry		AF170165
6-54	Czech Republic	Plum		AF170160
7-20	Czech Republic	Cherry		AF170164
AlmIt.cor1	Italy	Almond	AJ133204	AJ133204
AlmIt.pre1	Italy	Almond	AJ133202	AJ133202
AprIt.caf1	Italy	Apricot	AJ133199	AJ133199
AprIt.nap1	Italy	Apricot	AJ133200	AJ133200
AprIt.try1	Italy	Apricot	AJ133201	AJ133201
CH30	United States	Cherry	AF034994	
CH38	United States	Cherry	AF034991	
CH39	United States	Cherry	AF034990	
CH57	United States	Cherry	AF034993	
CH61	United States	Cherry	AF034989	
CH71	United States	Cherry	AF034995	
CH9	United States	Cherry	AF034992	
ChrIt.bla1	Italy	Cherry	AJ133210	AJ133210
ChrIt.lam1	Italy	Cherry	AJ133203	AJ133203
ChrIt.mrs1	Italy	Cherry	AJ133209	AJ133209
E260	India	Rose		CAF04057
Mission	United States	Almond	AF013285	
Nahnutá	Czech Republic	Plum		AF170169
Nahrbu	Czech Republic	Wild plum		AF170170
NctSp.mur1	Spain	Nectarine	AJ133208	AJ133208
PchIt.may1	Italy	Peach	AJ133205	AJ133205
PchIt.mry1	Italy	Peach	AJ133207	AJ133207
PchTu.unk1	Tunisia	Peach	AJ133206	AJ133206
PE-5	United States	Peach	L38823	
PlmAl.unk1	Albania	Plum	AJ133211	AJ133211
PlmIt.clf1	Italy	Plum	AJ133212	AJ133212
PlmIt.mrb1	Italy	Plum	AJ133213	AJ133213
Prune	United States	Plum	AF013286	
PS12-16	Czech Republic	Cherry		AF170158
PS14-22	Czech Republic	Flowering cherry		AF170159
PS7-11	Czech Republic	Sour cherry		AF170161
PS7-12	Czech Republic	Sour cherry		AF170162
PS7-5a	Czech Republic	Sour cherry		AF170166
PV-32	United States	Apple	Y07568	
PV-96	Germany	Mahaleb cherry	S78312	
Sss	Czech Republic	Plum		AF170168
SW6	United States	Cherry	AF013287	
UH1	Czech Republic	Plum		AF170167
UN	Czech Republic	Sour cherry		AF170163
Valtická	Czech Republic	Peach		AF170171

MATERIALS AND METHODS

Sequence alignments and phylogenetic analyses. The MP and CP sequences used in this study were downloaded from GenBank, and the corresponding accession numbers are shown in Table 1. Only full-length sequences have been used. Protein sequences were aligned with CLUSTAL-X version 1.8 (50). Nucleotide sequences were then aligned by concatenating triplets according to the amino acid sequence alignment with the program DAMBE version 4.2.7 (53). Alignments are available upon request. Hereafter, amino acid numbering will be assigned with CH30/USA/Cherry (accession no. AF034994) as the reference sequence for MP and E260/India/Rose (accession no. CAF4057) as the reference sequence for CP. These two isolates were chosen because no gaps were introduced into them during alignment.

Prior to any other evolutionary analyses by RNA alignment, the intensity of the phylogenetic signal contained was evaluated with the likelihood-mapping approach as implemented in the program TREEPUZZLE version 5.0 (45).

According to the likelihood ratio test implemented in MODELTEST version 3.7 (37), the molecular evolution model that best explains the observed pattern of nucleotide diversity for MP alignment was that proposed by Tamura and Nei (48), which was modified to account for the heterogeneity in substitution rates among sites (gamma parameter 0.285). Similarly, the model that best explained the observed sequence variation in CP alignment was that proposed by Kimura (27), with a transition to a transversion rate ratio of 2.291 and also incorporating heterogeneity in substitution rates among sites (gamma parameter 0.265). Phylogenetic reconstructions were obtained by the minimum-evolution method as implemented in the MEGA3 program (31) with the above nucleotide substitution models. The support of the internal nodes of the trees was evaluated by the bootstrap method with 10,000 pseudoreplicates (15). Nodes with bootstrap support of <75% were collapsed to the nearest significant node.

Analysis of selective constraints. To uncover the signature of coevolution between CP and MP, we first explored the selective constraints operating upon each individual protein. The advantage of using parsimony methods to determine codon regions consistently under specific selective constrains relies on its robustness against methodological biases (46). In maximum-parsimony analyses, the comparison of the rates of synonymous (d_S) and nonsynonymous (d_N) substitutions per class of sites is used to describe the evolutionary dynamics of protein-coding genes. The ratio of these two rates ($\omega = d_N/d_S$) provides information on whether the gene has been fixing amino acid replacements in a neutral fashion ($\omega = 1$), amino acid changes have been removed by the action of purifying selection ($\omega < 1$), or changes have been fixed by adaptive evolution ($\omega > 1$).

The particular method chosen is based on the parsimony criterion, which is appropriate here, given the conservation at the codon level of our alignments, and is implemented in the program SWAPSC version 1.0 (12). SWAPSC uses a statistically optimized window size to detect selective constraints in specific codon regions of the alignment at a particular branch of the phylogenetic tree

(13). Briefly, the method estimates the expected distribution of d_S and d_N by Li's method (32) from simulated sequence alignments and assuming a Poisson distribution of substitutions. A statistically optimum window size is then estimated that makes the detection of adaptive evolution independent of the window size. The empirical values of d_S and d_N obtained by using the optimal window size are contrasted with the expected distributions, and several hypotheses regarding the selective constraints acting on codon regions are tested. Simulated sequence alignments were obtained with the program EVOLVER from the PAML package version 3.14 (54) with parameters estimated from the true sequence alignments after running the most appropriate codon-based model in PAML (models M2 and M8 for MP and CP, respectively [55]).

The Bayesian approach implemented in the CODEML program from PAML was also used. The sites determined to be under selection were the same as those found by the above-mentioned maximum-parsimony approach, giving robustness to the conclusion (data not shown).

Testing for coevolution between amino acid sites. The aim of this analysis was to determine whether MP and CP have coevolved during PNRSV phenotypic radiation because of their putative functional link. We also tested whether coevolving amino acids can be used to predict protein-protein contact interfaces. We realize that this type of analysis is not definitive in resolving docking problems, but nonetheless, results could aid in identifying specific amino acid sites responsible for protein interactions and in designing new experiments. We are also interested in a more general definition of protein interaction, which can be functional, physical, or phylogenetic. Significance of coevolution was tested by the mutual information criterion (MIC) approach taken from information theory (28, 29). MIC definition involves the joint probability distribution, $P(x_i, y_j)$, of the occurrence of symbol x at position i and symbol y at position j belonging to the same protein or to two different proteins. MIC values range between 0, indicating independent evolution, and a positive value whose magnitude depends on the strength of covariation between sites. Variable positions included in the analyses were only those parsimony informative (i.e., containing at least two types of amino acids each with a minimum count of two). The statistical power of this test depends both on the sample size and on the level of variation at the sites considered. The significance of MIC values was assessed by means of a randomization test in which columns in the amino acid alignment of parsimony-informative sites were shuffled in place for each protein. The MIC value was recalculated thereafter for each shuffle to generate the expected probability distribution under the null hypothesis of no association. 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 significance level was set to 1%. Only the 15 isolates for which both full-length protein sequences were available in the database were used for these analyses (Table 1).

Threading predictions of protein tertiary structure. Tertiary structures of MP and CP were predicted by threading with the server http://cbsuapps.tc.cornell.edu/loopp.aspx and default options (33). Structure predictions are available upon request (see Fig. 3). So far, the native tertiary structures of PNRSV MP and CP have not been experimentally determined. However, the structure of phylogenetically close AMV CP has been obtained at 4 Å resolution (30). At this resolution, CP presents a globular structure with both the C and the N termini emerging from it. This gives indirect support to our prediction of PNRSV CP folding as a globule with protruding ends. Even less is known about MP structure. The PNRSV MP RNA-binding domain (amino acids 56 to 88) has been shown by circular dichroism to fold into an α-helix (22). Our predicted globular structure for MP reflects this experimental evidence. Nonetheless, given the scarceness of information on the in vivo three-dimensional structure adopted by MP, our threading prediction cannot be contrasted with experimental data and therefore a flag of caution has to be put on any conclusion drawn from the predicted structure.

Sites identified as selectively important or involved in covariation were mapped into the predicted structures (see Fig. 3).

RESULTS

Testing of the phylogenetic signal in MP and CP genes. Likelihood-mapping analyses showed strong phylogenetic signals in the two data sets. For the gene for MP, the tree-likeness (i.e., the percentage of all well-resolved quartets) of the data was 87.85%, whereas for the gene for CP it was 75%; we thus concluded that our data sets contained a significant phylogenetic signal and we proceeded with the molecular evolution study. (Because of limitations of the TREEPUZZLE software, the nucleotide substitution model of Tamura and Nei was used for both genes in this particular analysis.)

Figure 1 shows the minimum-evolution phylogenetic trees inferred for the genes for MP and CP. Regardless of the gene analyzed, internal nodes are generally well supported by the bootstrap analysis. External nodes are generally less well supported (Fig. 1). More importantly, the three main groups previously defined in the literature upon sequence homology (PE-5, PV-32, and PV-96) are highly supported by bootstrap values (Fig. 1). However, the phylogenetic position of isolate ChrIt.mrs1/Italy/Cherry, previously considered a member of the PV-96 group (2), was not well resolved. As shown in Fig. 1A, its splitting node has a bootstrap support value of $<75\%$ and thus has been collapsed. Nonetheless, to keep our classification consistent with previous proposals (2), ChrIt.mrs1/Italy/Cherry has been conservatively included within the PV-96 group (Fig. 1A).

The clusters containing isolates belonging to the PV-32 and PV-96 groups appear more closely related to each other than they are to isolates in group PE-5. This result is consistent with either a more recent common ancestor of the isolates belonging to the PV groups or, alternatively, an accelerated rate of evolution of group PE-5, this being the case for the genes for both MP and CP. The results reported in the following section support the second hypothesis.

Detecting selective constraints in MP and CP. The parsimony-based method found evidence of adaptive evolution in both proteins (Table 2). The optimum window sizes estimated by SWAPSC were five and four codons for MP and CP, respectively. A single region in MP appears to be under adaptive evolution, i.e., positive selection ($\omega > 1$), E252 to E257. This region was selected after the split occurred within the PE-5 clade and produced the ancestor of a set of geographically and host diverse isolates (branch G in Fig. 2A). MP sites under purifying selection ($\omega < 1$) are located in regions L61 to S65 and R212 to L216. The first region has undergone purifying selection in all branches of the phylogenetic tree, whereas the action of purifying selection on the second region has been limited to branches A, B, and G (Table 2 and Fig. 2A), the internal tree branches that gave rise to the different clades. Finally, the method also detected three regions of the protein that show globally accelerated rates of amino acid substitution (Table 2). The hypermutagenesis of these regions took place in the branches leading to the PE-5 cluster (branches G and H in Fig. 2A), supporting the hypothesis of an accelerated rate of evolution during the genesis of this clade. The above-mentioned sites were mapped onto the predicted tertiary structure (Fig. 3).

Regarding CP, the picture is as follows. Two regions contain amino acids under directional positive selection (S49 to G53 and K139 to Q142), four regions contain amino acids under purifying negative selection (Table 2), and one stretch had suffered from hypermutagenesis (Table 2). Interestingly, positive selection acted along the deepest branches of the tree, during the diversification of the lineage leading to PE-5 isolates. Purifying selection acted along several branches of the tree (Table 2 and Fig. 2B), and its action during the separation of lineages PV-32 and PV-96 was noticeable (Table 2 and Fig. 2B). Hypermutagenesis only happened in branch D (Table 2

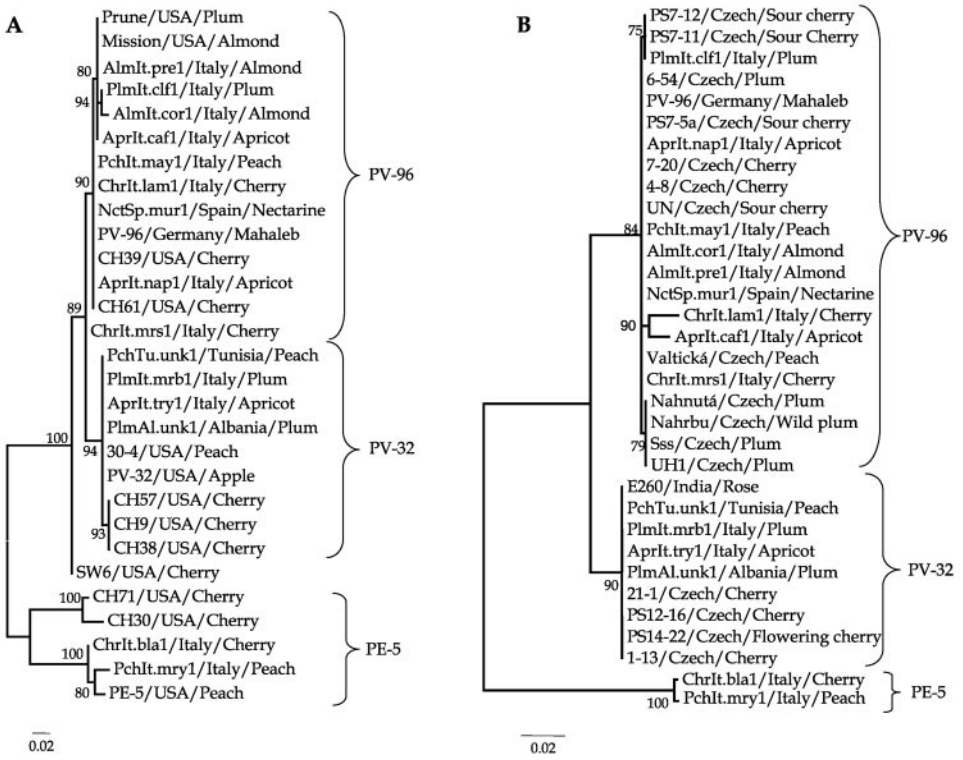

FIG. 1. Phylogenetic analysis of MP (A) and CP (B). Trees were inferred by the minimum-evolution method using the Tamura and Nei corrected nucleotide distances (48). Values at the nodes are bootstrap support values based on 10,000 pseudoreplicates. Nodes with bootstrap support of <75% were collapsed. ChrIt.mrs1/Italy/Cherry has been included in the PV-96 group for consistency with previous proposals (2).

and Fig. 2B). The above sites were mapped onto the predicted tertiary structure (Fig. 3).

The fact that the branch leading to the ancestor of PE-5 isolates was found to be under adaptive evolution in both proteins provides a first indication of coevolution between MP and CP. Interestingly, this putative process of coevolution is in the root of the genesis of a new serotype.

Covariation within and between MP and CP. By the MIC approach, we have detected several covariation groups ($P < 0.01$) within and between proteins (Table 3). Figure 3 illustrates the distribution of covariation sites on the predicted tertiary structures. In a first step, we analyzed within-protein covariation. Regarding MP, three amino acid residues show significant covariation (L253, E257, and I261; Table 3). Here-

TABLE 2. Summary of SWAPSC maximum-parsimony analysis of adaptive evolution in PNRSV CP and MP[a]

Protein and region	d_N	$P(d_N)$	d_S	$P(d_S)$	ω	Constraint	Tree branch(es)
MP							
L61-S65	0	0.044	0.057–0.768	≥0.006	0	NS	A, B, C, D, E, F, G, H
R136-V142	0.342	0.013	0.099	0.237	3.447	AdN	H
R212-L216	0	0.044	0.124–0.371	≥0.030	0	NS	A, B, G
E252-E257	0.306	0.043	0.096	0.116	3.183	PS	G
S263-E267	0.339	0.026	0.138	0.043	2.448	AdN	G
G280-A285	0.359	0.010	0.842	0.304	0.426	AdN	H
CP							
R47-S50	0	0.038	0.353	0.107	0	NS	C
S49-G53	0.637	0.009	0.509	0.204	1.252	PS	D
L114-L117	0	0.038	0.353–1.080	≥0.083	0	NS	B, C
T118-R122	0.546	0.019	0.755	0.138	0.723	AdN	D
R122-S128	0	≥0.049	0.261	≥0.167	0	NS	D
K139-Q142	0.729	0.003	0.411	0.136	1.771	PS	D
A148-K152	0	0.038	0.353–0.178	≥0.054	0	NS	A, C, D

[a] An amino acid region is considered to be under positive selection if ω is >1 and $P(d_S)$ is >0.05, indicating no saturation or selection over synonymous sites. The branches in the phylogenetic trees where the constraint was found are labeled as in Fig. 3. Constraints are indicated as NS (negative selection), PS (positive selection), and AdN (accelerated rate of substitution).

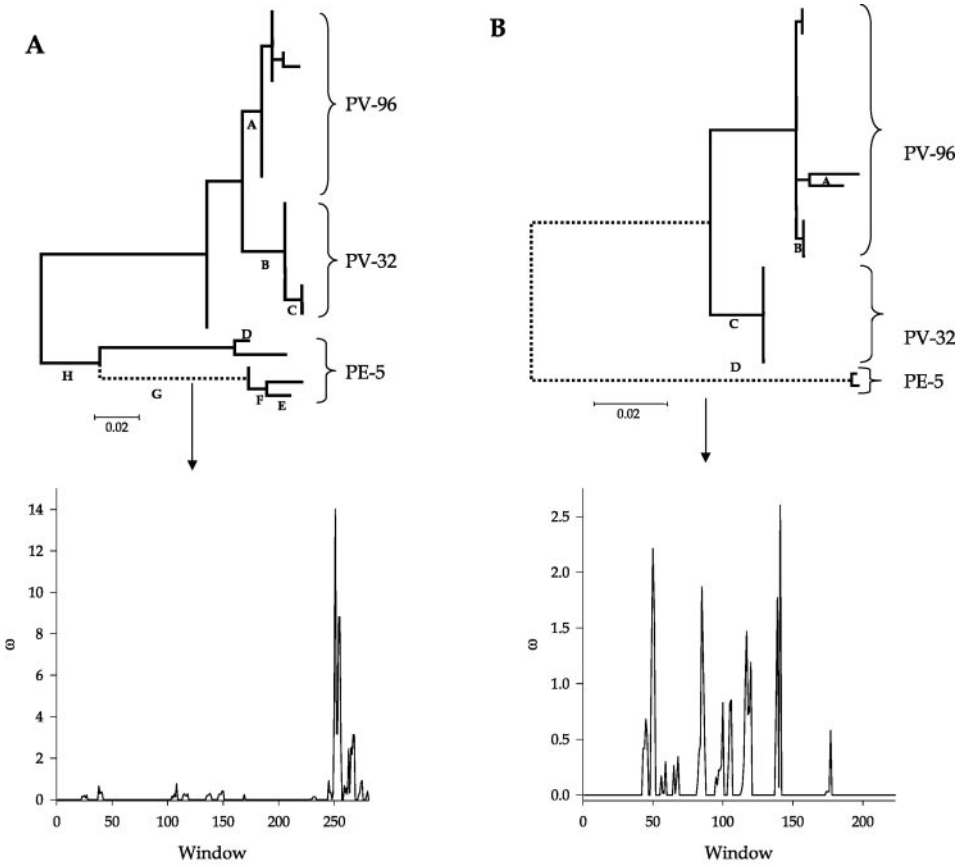

FIG. 2. SWAPSC analysis of selective constraints in MP (panel A) and CP (panel B). In this example, the branch detected under adaptive evolution is drawn as a dotted line. The plots show the distribution of the nonsynonymous to synonymous substitution rate ratio (ω) along the sequence. Peak heights are proportional to the intensities of positive selection in the corresponding regions.

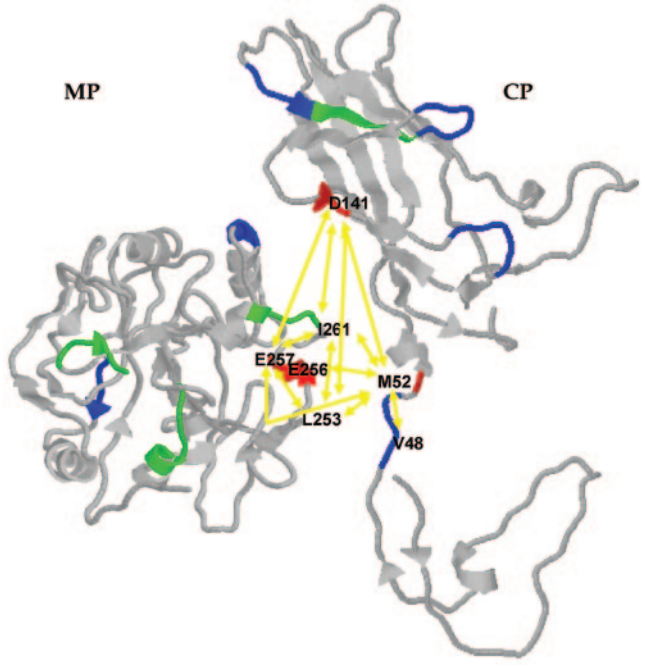

FIG. 3. Selected sites and covariation analysis of MP and CP. Color code: red, positively selected sites; blue, negatively selected sites; green, sites showing hypermutagenesis. Covarying amino acids are indicated and linked with yellow arrows.

after, we will call this covariation group MP CG. For CP, two different significant covariation groups (hereafter called CP CG1 and CG2) have been identified (Table 3). CP M52 belongs to both groups, and the covariation between V48 and D141 was statistically significant at the usual 5% significance level (MIC value = 0.161, $P = 0.040$), suggesting that both covariation groups can be reduced to a single one at the less stringent 5% confidence level. Nonetheless, in the following discussion we will maintain the distinction between two nondisjoint groups.

In a second step, we have explored the covariation across proteins (Table 3 and Fig. 3). Very interestingly, MP CG sig-

TABLE 3. Covariation groups within and between proteins

Comparison and covarying sites	MIC value	*P* value
Within MP CG: L253, E257, I261	≥0.243	≤0.008
Within CP		
CG1: V48, M52	0.229	0.007
CG2: M52, D141	0.261	0.003
Between proteins		
MP CG, CP CG2	≥0.243	≤0.010
MP E256, CP M52	0.251	0.007

nificantly covaries with CP CG2. In addition, MP E256 also covaries significantly with CP M52.

DISCUSSION

It has been previously reported that the variability of MP concentrates around amino acids 232 to 285 (1, 2, 19, 52). Our analyses confirm this observation and also show that positively selected and hypermutated sites also concentrate in this variable region of the molecule. Our results showing evidence of adaptive evolution within a region spanning amino acids 245 to 270 support the previously suggested antigenicity of this region (52) (Table 2, E252 to E257). Some of the changes are group specific, as replacements in amino acid positions V253, D256, D257, and P260 appear always in association with both PV groups but have a different amino acid composition in the isolates belonging to the PE-5 group (52). PE-5 isolates also share other group-specific mutations in this region, and interestingly, the operation of positive selection has been detected on the branch leading to this serotype, revealing that the differences between the PE-5 and PV groups may be due to changes in selective constraints. Furthermore, our finding that MP CG and site E256, all within this putative antigenic region, significantly covary with CP CG2 suggests that this region is involved in the establishment of interactions with CP that may be important for mediating cell-to-cell movement of PNRSV.

There are two domains of MP that are highly conserved, probably because of structural and/or functional constraints. These domains range, respectively, from L61 to S65 and from R212 to L216. In addition, amino acid sites 8 to 70 have been predicted to form a transmembrane domain and may be necessary for the protein to interact with the cell wall (42). Supporting the existence of such functional and/or structural constraints in this region, we have identified sites that are under the action of negative selection (Table 2, L61 to S65). In addition, it has also been recently shown that MP interacts with RNA molecules throughout amino acids 56 to 88 (21). Interestingly, this 32-amino-acid domain includes the above-mentioned amino acids detected as negatively selected (L61 to S65). These basic amino acids should be, in accordance with our threading structural predictions (data not shown), exposed to the solvent, both properties being typical of RNA- and DNA-binding motifs.

The variability in CP is concentrated in two regions, including one between N-terminal amino acids 48 and 56 (19) and another between C-terminal amino acids 139 and 145. These variable regions also contain most of the sites found to be under positive selection (Table 2, S49 to G53 and K139 to Q142). Amino acids 25 to 50 are essential for the protein to bind PNRSV RNA in vitro (3, 36). More generally, the N-terminal domain of members of the different genera within the family *Bromoviridae* (*Cucumovirus* [20], *Bromovirus* [44], and *Ilarvirus* [3, 4, 7]) is involved in capturing RNA during encapsidation and also plays an important role in the phenomenon of genomic activation in the case of ilarviruses (5, 6). Zhang et al. (56) predicted, on the basis of in silico experiments, that the main requirement for cowpea chlorotic mottle bromovirus RNA encapsidation should be an abundance of positively charged amino acids at the internal surface of the capsid. In good agreement with this prediction, changes S50N and D141N (both positively selected in branch D leading to group PE-5) and change G53R (positively selected in the branch leading to isolates ChrIt.lam1/Italy/Cherry and AprIt.caf1/Italy/Apricot of group PV-96) imply net gains of positive charges. Therefore, changes increasing the strength of the RNA-CP electrostatic interaction may have been selected during PNRSV diversification.

It has also been suggested that the amino acids around M110 should play an important role during the formation of CP dimers (9, 52). This importance is confirmed by our finding of a stretch of four amino acids (Table 2, L114 to L117) under the action of negative selection.

Our maximum-parsimony analysis of CP identified sites under positive selection at the origin of the PE-5 group (Table 2, S49 to G53). Precisely in this region of PE-5 CP, Vašková et al. (52) have predicted upon bioinformatic grounds the existence of an antigenic site. Amino acids K139 to Q142 (Table 2) have also been identified as under positive selection in the branch leading to isolates PchIt.mry1/Italy/Peach and ChrIt.bla1/Italy/Cherry of the PE-5 group. The biological significance of these sites is not obvious. However, since it is quite close to the region including A148 to K152, which is (i) under negative selection (and hence likely has an important function) and (ii) predicted to be in the surface of the folded protein (data not shown), we can speculate that it may be involved in the formation of the MP-CP complexes that allow PNRSV cell-to-cell movement. This suggestion is supported by our covariation analyses, which reveal that the C-terminal amino acids of MP are coevolving with the N-terminal region of CP (CP CG1). This suggestion gets further support from evidence for such contact in other members of the *Bromoviridae* family such as CMV (26), BMV (47), and AMV (41, 43).

Comparison of Tables 2 and 3 shows that all but one (MP I261) of the amino acids involved in covariation groups have been shown to be the target of natural selection. MP L253, E256, and E257 are under positive selection, as is the case for CP M52 and D141, whereas CP V48 is under purifying selection. This almost perfect match suggests that selection acts upon the maintenance of the right folded structures (i.e., within-molecule interactions) and interaction between proteins. From a practical standpoint, this match between the two types of analyses enhances the robustness of the results.

Within the inherent limitations of the threading technique used for predicting MP and CP tertiary structures, it is worth noting that the amino acids covarying between the proteins are not scattered along the molecules but it is possible to fit both molecules in space in such a way that the predicted intermolecular covariations start making structural sense (Fig. 3). The three amino acids belonging to MP CG plus MP E256 are all concentrated in the same surface region of the predicted globule, whereas the three amino acids belonging to CP CG1 and CG2 are located in a predicted hypothetical cavity (CP M52 occupying a central location) into which MP would fit to maximize the likelihood of physical interactions between the two predicted between-protein covariation groups.

The present in silico study opens new research avenues for researchers interested in experimentally exploring the in vivo interaction between MP and CP. In particular, it suggests which sites can be changed by site-directed mutagenesis to preclude the formation of MP-CP complexes. Also, site-di-

rected mutagenesis studies may shed light on the importance of the sites identified here as the target of selection.

ACKNOWLEDGMENTS

We thank F. Aparicio, V. Pallás, and J. A. Sánchez-Navarro for comments and discussion.

This work was supported by grants from the Spanish MEC-FEDER (BMC2003-00066 and BFU2005-23720-E/BMC), the Generalitat Valenciana (GV04B280 and GRUPOS03/064), and the EMBO Young Investigator Program to S.F.E. and from the Irish Science Foundation under the President of Ireland Young Researcher Award program to M.A.F. F.M.C. was the recipient of fellowships from the CSIC I3P Bioinformatics program and from the ESF Functional Genomics program.

REFERENCES

1. **Aparicio, F., A. Mirta, B. di Terlizzi, and V. Pallás.** 1999. Molecular variability among isolates of prunus necrotic ringspot virus (PNRSV) from different Prunus species. Phytopathology **89:**991–999.
2. **Aparicio, F., and V. Pallás.** 2002. The molecular variability analysis of the RNA 3 of fifteen isolates of prunus necrotic ringspot virus sheds light on the minimal requirements for the synthesis of its genomic RNA. Virus Genes **25:**75–84.
3. **Aparicio, F., M. Vilar, E. Perez-Payá, and V. Pallás.** 2003. The coat protein of prunus necrotic ringspot virus specifically binds to and regulates the conformation of its genomics RNA. Virology **313:**213–223.
4. **Baer, M. L., F. Houser, L. S. Loesch-Fries, and L. Gehrke.** 1994. Specific RNA binding by amino terminal peptides of alfalfa mosaic virus coat protein. EMBO J. **13:**727–735.
5. **Bol, J. F.** 1999. Alfalfa mosaic virus and ilarviruses: involvement of coat protein in multiple steps of the replication cycle. J. Gen. Virol. **80:**1089–1102.
6. **Bol, J. F.** 2005. Replication of alfamo- and ilarviruses: role of the coat protein. Annu. Rev. Phytopathol. **43:**39–62.
7. **Bol, J. F., B. Kraal, and F. T. H. Brederode.** 1974. Limited proteolysis of alfalfa mosaic virus: influence on the structural and biological function of the coat protein. Virology **58:**101–110.
8. **Bol, J. F., L. van Vloten-Doting, and E. M. J. Jaspars.** 1971. A functional equivalence of top component RNA and coat protein in the initiation of infection by alfalfa mosaic virus. Virology **46:**73–85.
9. **Choi, J., and S. Loesch-Fries.** 1999. Effect of C-terminal mutations of alfalfa mosaic virus coat protein on dimer formation and assembly *in vitro*. Virology **260:**182–189.
10. **Codoñer, F. M., J. M. Cuevas, J. A. Sánchez-Navarro, V. Pallás, and S. F. Elena.** 2005. Molecular evolution of the plant virus *Bromoviridae* based on RNA 3-encoded proteins. J. Mol. Evol. **61:**697–705.
11. **Codoñer, F. M., and S. F. Elena.** 2006. Evolutionary relationships among members of the *Bromoviridae* deduced from whole proteome analysis. Arch. Virol. **151:**299–307.
12. **Fares, M.** 2004. SWAPSC: sliding window analysis procedure to detect selective constraints. Bioinformatics **20:**2867–2868.
13. **Fares, M. A., S. F. Elena, J. Ortiz, A. Moya, and E. Barrio.** 2002. A sliding window-based method to detect selective constraints in protein-coding genes and its application to RNA viruses. J. Mol. Biol. **55:**509–521.
14. **Fauquet, C. M., M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball (ed.).** 2005. Virus taxonomy, VIIIth report of the ICTV. Elsevier/Academic Press, London, United Kingdom.
15. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
16. **Flasinski, S., A. Dzianott, S. Pratt, and J. J. Bujarski.** 1995. Mutational analysis of the coat protein gene of brome mosaic virus: effects on replication and movement in barley and in *Chenopodium hybridum*. Mol. Plant-Microbe Interact. **8:**23–31.
17. **George, J., and T. R. Davidson.** 1963. Pollen transmission of necrotic ringspot virus and sour cherry yellows viruses from tree to tree. Can. J. Plant Sci. **43:**276–278.
18. **Hammond, R. W.** 2003. Phylogeny isolates of prunus necrotic ringspot virus from the ilarvirus ringtest and identification of group-specific features. Arch. Virol. **148:**1195–1210.
19. **Hammond, R. W., and J. M. Crosslin.** 1998. Virulence and molecular polymorphism of prunus necrotic ringspot virus isolates. J. Gen. Virol. **79:**1815–1823.
20. **Harrison, S. C.** 1984. Multiple nodes of subunit association in the structures of simple spherical viruses. Trends Biochem. Sci. **9:**345–351.
21. **Herranz, M. C., and V. Pallás.** 2004. RNA-binding properties and mapping of the RNA-binding domain from the movement protein of prunus necrotic ringspot virus. J. Gen. Virol. **85:**761–768.
22. **Herranz, M. C., J. A. Sánchez-Navarro, A. Saurí, I. Mingarro, and V. Pallás.** 2005. Mutational analysis of the RNA-binding domain of prunus necrotic ringspot virus (PNRSV) movement protein reveals its requirement for cell-to-cell movement. Virology **339:**31–41.
23. **Hwang, M. S., S. H. Kim, J. H. Lee, J. M. Bae, K. H. Paek, and Y. I. Park.** 2005. Evidence for interaction between the 2a polymerase protein and the 3a movement protein of cucumber mosaic virus. J. Gen. Virol. **86:**3171–3177.
24. **Jaspars, E. M. J.** 1999. Genome activation in alfamo- and ilarviruses. Arch. Virol. **144:**843–863.
25. **Kaplan, I. B., L. Zhang, and P. Palukaitis.** 1998. Characterization of cucumber mosaic virus. V. Cell-to-cell movement requires capsid but not virions. Virology **246:**221–231.
26. **Kim, S. H., N. O. Kalinina, I. Andreev, E. V. Ryabov, A. G. Fitzgerald, M. E. Taliansky, and P. Palukaitis.** 2004. The C-terminal 33 amino acids of the cucumber mosaic virus 3a protein affect virus movement, RNA binding and inhibition of infection and translation. J. Gen. Virol. **85:**221–230.
27. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
28. **Korber, B. T. M., R. M. Farber, D. H. Wolpert, and A. S. Lapedes.** 1993. Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. Proc. Natl. Acad. Sci. USA **90:**7176–7180.
29. **Kullback, S.** 1959. Information theory and statistics. Wiley, New York, N.Y.
30. **Kumar, A., V. S. Reddy, V. Yusibov, P. R. Chipman, Y. Hata, I. Fita, K. Fukuyama, M. G. Rossmann, L. S. Loesch-Fries, T. S. Baker, and J. E. Jonhson.** 1997. The structure of alfalfa mosaic virus capsid protein assembled as a T = 1 icosahedral particle at 4.0-Å resolution. J. Virol. **71:**7911–7916.
31. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5:**150–163.
32. **Li, W. H.** 1993. Unbiased estimation of the rates of synonymous and non-synonymous substitutions. J. Mol. Evol. **36:**96–99.
33. **Meller, J., and R. Elber.** 2001. Linear programming optimization and a double statistical filter for protein threading protocols. PROTEINS: Struct. Funct. Genet. **45:**241–261.
34. **Mink, G. I., W. E. Howell, A. Cole, and S. Regev.** 1987. Three serotypes of prunus necrotic ringspot virus isolated from rugose mosaic-diseased sweet cherry in Washington. Plant Dis. **71:**91–93.
35. **Murphy, F. A., C. M. Fauquet, D. M. C. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. Mayo, and M. D. Summers.** 1995. Virus taxonomy: classification and nomenclature of viruses. Arch. Virol. Suppl. **10:**450–457.
36. **Pallás, V., J. A. Sánchez-Navarro, and J. Díez.** 1999. *In vitro* evidence for RNA binding properties of the coat protein of prunus necrotic ringspot ilarvirus and their comparison to related and unrelated viruses. Arch. Virol. **144:**797–803.
37. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.
38. **Rao, A. L. N., and G. L. Grantham.** 1995. Biological significance of the seven amino-terminal basic residues of brome mosaic virus coat protein. Virology **211:**42–52.
39. **Rao, A. L. N., and G. L. Grantham.** 1996. Molecular studies on bromovirus capsid protein. II. Functional analysis off the amino-terminal arginine-rich motif and its role in encapsidation, movement and pathology. Virology **226:**294–303.
40. **Sacher, R., and P. Ahlquist.** 1989. Effects of deletions in the N-terminal basic arm of brome mosaic virus coat protein on RNA packaging and systemic infection. J. Virol. **63:**4545–4552.
41. **Sánchez-Navarro, J. A., and J. F. Bol.** 2001. Role of alfalfa mosaic virus movement protein and coat protein in virus transport. Mol. Plant-Microbe Interact. **14:**1051–1062.
42. **Sánchez-Navarro, J. A., and V. Pallás.** 1997. Evolutionary relationships in the ilarviruses: nucleotide sequence of prunus necrotic ringspot virus RNA3. Arch. Virol. **142:**749–763.
43. **Sánchez-Navarro, J. A., M. C. Herranz, and V. Pallás.** 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.
44. **Sgro, J. Y., B. Jacrot, and J. Chroboczek.** 1986. Identification of regions of brome mosaic virus coat protein chemically cross-linked *in situ* to viral RNA. Eur. J. Biochem. **154:**69–76.
45. **Stimmer, K., and A. von Haeseler.** 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. USA **94:**6815–6819.
46. **Suzuki, Y., and M. Nei.** 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **18:**2179–2185.
47. **Takeda, A., M. Kaido, T. Okuno, and K. Mise.** 2004. The C-terminus of the movement protein of brome mosaic virus controls the requirement for coat protein in cell-to-cell movement and plays a role in long distance movement. J. Gen. Virol. **85:**1751–1761.
48. **Tamura, K., and M. Nei.** 1993. Estimation of the number of nucleotide

substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10:**512–526.
49. **Tenllado, F., and J. F. Bol.** 2000. Genetic dissection of the multiple functions of alfalfa mosaic virus coat protein in viral replication, encapsidation, and movement. Virology **268:**29–40.
50. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
51. **Uemoto, J. K., and S. W. Scott.** 1992. Important disease of prunus caused by viruses and other graft-transmissible pathogens in California and South Carolina. Plant Dis. **76:**5–11.
52. **Vašková, D., K. Petrzik, and R. Karesová.** 2000. Variability and molecular typing of the woody-tree infecting prunus necrotic ringspot ilarvirus. Arch. Virol. **145:**699–709.
53. **Xia, X., and Z. Xie.** 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. **92:**371–373.
54. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **337:**555–556.
55. **Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen.** 2000. Codon substitution models for heterogeneous selection pressures at amino acid sites. Genetics **155:**431–449.
56. **Zhang, D., R. Konecny, N. A. Baker, and J. A. McCammon.** 2004. Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. Biopolymers **75:**325–337.