

Molecular Evolution of the Plant Virus Family *Bromoviridae* Based on RNA 3-Encoded Proteins

Francisco M. Codoñer,¹ José M. Cuevas,² Jesús A. Sánchez-Navarro,¹ Vicente Pallás,¹ Santiago F. Elena¹

¹ Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, Avenida de los naranjos s/n, 46022 Valencia, Spain

² Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

Received: 26 January 2005 / Accepted: 26 May 2005 [Reviewing Editors: Dr. Martin Kreitman and Dr. James Bull]

Abstract. We have carried out an evolutionary study of the two proteins encoded by the RNA 3 from members of the plant virus family *Bromoviridae*. Using maximum likelihood methods, we have inferred the patterns of amino acid substitution that better explain the diversification of this viral family. The results indicate that the molecular evolution of this family was rather complex, with each protein evolving at different rates and according to different patterns of amino acid substitution. These differences include different amino acid equilibrium frequencies, heterogeneity in substitution rates among sites, and covariation among sites. Despite these differences, the model of protein evolution that better fits both proteins is one specifically proposed for the evolution of globular proteins. We also found evidence for coevolution between domains of these two proteins. Finally, our analyses suggest that the molecular clock hypothesis does not hold, since different lineages evolved at different rates. The implications of these results for the taxonomy of this important family of plant viruses are discussed.

Key words: *Bromoviridae* — Maximum likelihood — Plant viruses — Protein evolution — Viral evolution — Virus taxonomy

Introduction

The family *Bromoviridae* constitutes one of the most important families of plant viruses. A large number of its members are distributed worldwide, and several are responsible for major disease epidemics in crop plants. Their host range covers more than 1000 species, including herbaceous plants, shrubs, and trees. Most of the type members of the different genera, such as cucumber mosaic virus (CMV), brome mosaic virus (BMV), and alfalfa mosaic virus (AMV), constitute excellent model systems in plant virology. The genome of the *Bromoviridae* is approximately 8 kb long and organized into three linear, positive sense ssRNAs. The genomic RNAs 1 and 2 each encode a single large ORF (the replicase proteins P1 and P2), and in some genera, RNA 2 also encodes a second ORF that is apparently translated from a subgenomic RNA (sgRNA). RNA 3 is bicistronic and has two ORFs encoding the movement protein (MP) and the coat protein (CP), the latter being translated from a sgRNA. Besides all these common features, the genera *Ilarvirus* and *Alfavirus* share the phenomenon of genome activation (Bol et al. 1971). Genome activation is the binding of CP to specific sites near the 3' end of the RNA to initiate infection.

The present taxonomy of the *Bromoviridae* divides the family into five genera (Table 1), named after their most representative member. Genera are distinguished by common biological properties of their members. The *Ilarvirus* genus, which includes more members than the other genera, has been subdivided into seven groups according to their antigenic

Correspondence to: Santiago F. Elena; email: sfelena@ibmcp.upv.es

Table 1. Viruses previously assigned to genera of the *Bromoviridae* family: Only viruses whose RNA 3 sequence has been determined are included

Genus	ICTV subgroup	Virus	Accession No.
<i>Alfavirus</i>		<i>Alfalfa mosaic virus</i> (AMV)	M59241
<i>Bromovirus</i>		<i>Broad bean mottle virus</i> (BBMoV)	M60291
		<i>Brome mosaic virus</i> (BMV)	X58458
		<i>Cowpea chlorotic mottle virus</i> (CCMV)	M28818
<i>Cucumovirus</i>		<i>Cucumber mosaic virus</i> (CMV)	D00385
		<i>Peanut stunt virus</i> (PSV)	D00668
		<i>Tomato aspermy virus</i> (TAV)	L15335
<i>Ilarvirus</i>	Subgroup 1	<i>Hydrangea mosaic virus</i> (HdMV)	U35145
		<i>Tobacco streak virus</i> (TSV)	X00435
		<i>Asparagus virus 2</i> (AV-2)	X86352
	Subgroup 2	<i>Citrus leaf rugose virus</i> (CiLRV)	U17390
		<i>Citrus variegation virus</i> (CVV)	U17389
		<i>Elm mottle virus</i> (EMoV)	U85399
		<i>Spinach latent virus</i> (SpLV)	U93194
		<i>Tulare apple mosaic virus</i> (TAMV)	AF226162
	Subgroup 3	<i>Apple mosaic virus</i> (ApMV)	U15608
		<i>Prunus necrotic ringspot virus</i> (PNRSV)	Y07568
	Subgroup 4	<i>Prune dwarf virus</i> (PDV)	L28145
	Subgroup 5	<i>American plum line pattern virus</i> (APLPV)	AF235166
	Subgroup 7	<i>Lilac ring mottle virus</i> (LiRMoV)	U17391
		<i>Olive latent virus 2</i> (OLV-2)	X76993
	<i>Oleavirus</i>		<i>Pelargonium zonate spot virus</i> (PZSV)
<i>Incertae sedis</i>			

properties (van Regenmortel et al. 2000). One more virus, pelargonium zonate spot virus (PZSV), although tentatively assigned to a new proposed genus, *Ourmiavirus*, presents some features characteristic of the *Bromoviridae* family and its exact taxonomic status within the family has not been explored yet. However, this taxonomy has recently been jeopardized. Studies of evolutionary relationships, considering, first, RNA 3 (Sánchez-Navarro and Pallás 1997) and, later, RNAs 1 and 2 (Rampitsch and Eastwell 1997; Scott et al. 1998; Shiel and Berger 2000), led to the proposal that AMV should be considered a true *Ilarvirus* instead of forming a distinct genus. In this scenario, it would be very convenient to study the mechanisms of molecular evolution driving the diversification of this important family of plant viruses, evaluating its consequences for the inference of correct taxonomic relationships. To address these questions we selected the two proteins encoded by genes located at the same RNA 3 molecule, which permits exclusion of pseudorecombination as a source of variation.

We focus our evolutionary study on amino acid sequences rather than on nucleotide ones because of the enormous divergence between the latter. Amino acid sequences of conserved proteins are widely used to infer distant phylogenetic relationships (Oppendoes 2003). For distant relationships, the use of nucleic acid sequences can be problematic, as alignment may be difficult, base frequencies may vary among species, and saturation of substitutions may have diluted phylogenetic information. In such cases, the use of protein sequences may be advantageous.

Models of amino acid substitutions have recently been developed (Yang et al. 1998).

Therefore, the aim of the present study is twofold. First, we study the mechanisms and patterns of molecular evolution that gave rise to the different members of the *Bromoviridae* family. Second, as a consequence of these analyses, we reevaluate the taxonomic relationships of the family in an attempt to reconcile, whenever possible, the patterns of molecular evolution observed with the existence of taxonomic groups previously defined on the basis of antigenic, morphologic, and host-related properties.

Materials and Methods

Sequence Data

Sequences for each member of the *Bromoviridae* family were downloaded from GenBank. Table 1 lists the viruses included in this study as well as their current taxonomic position within the different genera. The accession numbers for the complete sequence of their RNA 3 are also reported in Table 1. As outgroup, we used the two homologous sequences from tobacco mosaic virus (TMV; accession number NC001367). Amino acid sequences were aligned with CLUSTAL-X version 1.83 with its default parameters (Thompson et al. 1997) and with MUSCLE version 3.52 (Edgar 2004). In both cases, manual adjustments were necessary in order to preserve previously described homologies (Sánchez-Navarro and Pallás 1997) and the domains defined in the Pfam HMM library (<http://pfam.wustl.edu>), including the 30K domain of the MP (Melcher 2000). After manual editing, the alignments were almost identical, so we used only the CLUSTAL-X hand-improved ones for further analyses. Alignments are available upon request.

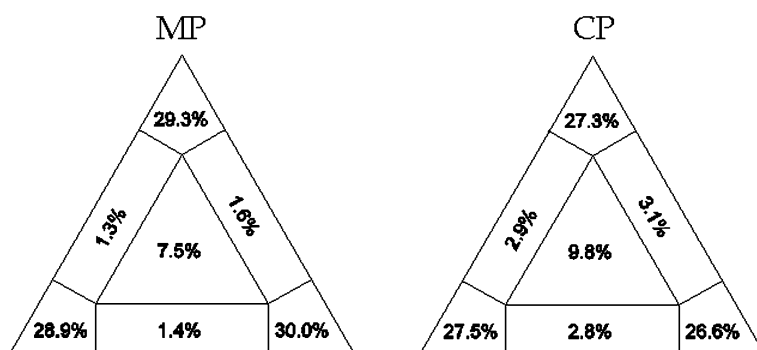


Fig. 1. Likelihood-mapping analysis for the MP and CP proteins.

Phylogenetic Reconstruction and Selection of the Model of Amino Acid Substitution

The information content for evolutionary reconstruction from the protein alignments was evaluated by likelihood mapping using TREEPUZZLE version 5.0 (Strimmer and von Haeseler 1997). This method is especially suitable for analyzing controversial phylogenies and testing the reliability of the internal edges of a tree.

Phylogenetic reconstruction and assessment of the goodness of fit for different models of amino acid substitution were done simultaneously using PROTTEST version 1.0.6 (Abascal et al. 2005). PROTTEST uses the neighbor-joining method for estimating the tree topology for each substitution model tested. Once a model of amino acid substitution was chosen, we used it to infer new phylogenetic trees using two alternative approaches: (1) the quartet-puzzling maximum likelihood method (Strimmer and von Haeseler 1996), as implemented in TREEPUZZLE, with 10,000 puzzling steps; and (2) Bayesian inference, as implemented in MRBAYES version 3.0b4 (Huelsenbeck and Ronquist 2001), with four simultaneous chains, each of a million generations, sampled every 100 generations. Trees sampled before the cold chain reached stationarity, as judged by plots of maximum likelihood scores, were discarded as “burn-in.” In addition, A. Rambaut and A. Drummond’s TRACER program (available from <http://evolve.zoo.ox.ac.uk>) was used to assess whether the length of the chains was appropriate for ensuring convergence. Convergence was already reached after 10,000 generations.

Statistical confidence for the deduced evolutionary tree was assessed by the quartet-puzzling probabilities (QPP), using TREEPUZZLE, and by Bayesian posterior probability (BPP), using MRBAYES.

We further improved the model of amino acid substitution using the CODEML program of the PAML package (Yang 1997) and the tree topology previously obtained. CODEML was run four times for each model to confirm convergence to a global peak in the likelihood surface.

Comparisons among models were done using the Bayesian information criterion (BIC) (Posada and Buckley 2004). The BIC statistic for a given model is defined as $BIC = -2l + p \log N$, where l is the estimated log-likelihood, p the number of free parameters in the model, and N the sample size, which is defined as the total number of characters in the alignment. The model with the lowest BIC was considered to be the most appropriate one. (Note that branch lengths have to be estimated for every model.)

Data for each protein were analyzed both independently and combined. The most general model for a combined analysis of data from multiple proteins uses independent parameters for each one. This model involves, for each protein, ϕ , an independent set of parameters, Θ (branch lengths, amino acid frequencies, and substitution rates). If D_ϕ stands for the sequence data for protein ϕ , then the log-likelihood function for the combined data set can be written $l = \sum_{\phi} \ln[P(D_\phi | \Theta_\phi)]$ (Yang

1996). At the other extreme, in the simplest model, all parameters can be assumed to be the same for the whole data set. Between these two extremes lies a wide variety of models, each one assuming some aspects of the evolutionary process to be the same and others to be different.

Test of the Molecular Clock Hypothesis

To test whether the rates of molecular evolution were homogeneous along all the branches of the phylogenetic trees, we used two approaches: first, the maximum-likelihood methods implemented in the PAML package and, second, the two-cluster test (Takezaki et al. 1995) on the topology obtained in our previous phylogenetic reconstruction with the branch lengths computed by the best model of amino acid substitution. The advantage of this method is that it says which branches had an accelerated rate of substitution. This procedure was implemented in the TPCV program of the LIN-TREE package (Takezaki et al. 1995).

Results

As the first step, we did a likelihood-mapping analysis to test the reliability of our data for phylogenetic inference. To do so, we used both JTT (Jones et al. 1992) and WAG (Whelan and Goldman 2001) transition matrixes to model amino acid substitution rates and incorporated a discrete gamma distribution (d Γ) with 16 categories to model rate heterogeneity among sites. Figure 1 shows the results of these analyses when the WAG matrix was used, but the conclusions were indistinguishable from those obtained using the JTT matrix. A strong phylogenetic signal was detected for both data sets. For MP the treelikeness (i.e., the sum of all well-resolved quartets that correspond to the percentages in the corners of the simplex) of the data was 88.2%, whereas for the CP protein it was 81.4%; in consequence, we concluded that our data sets contained a significant phylogenetic signal and we proceeded with the molecular evolution study.

Model Selection and Inference of Phylogenetic Trees

For both proteins, as well as for the combined data set, the model that produced the lowest BIC was the

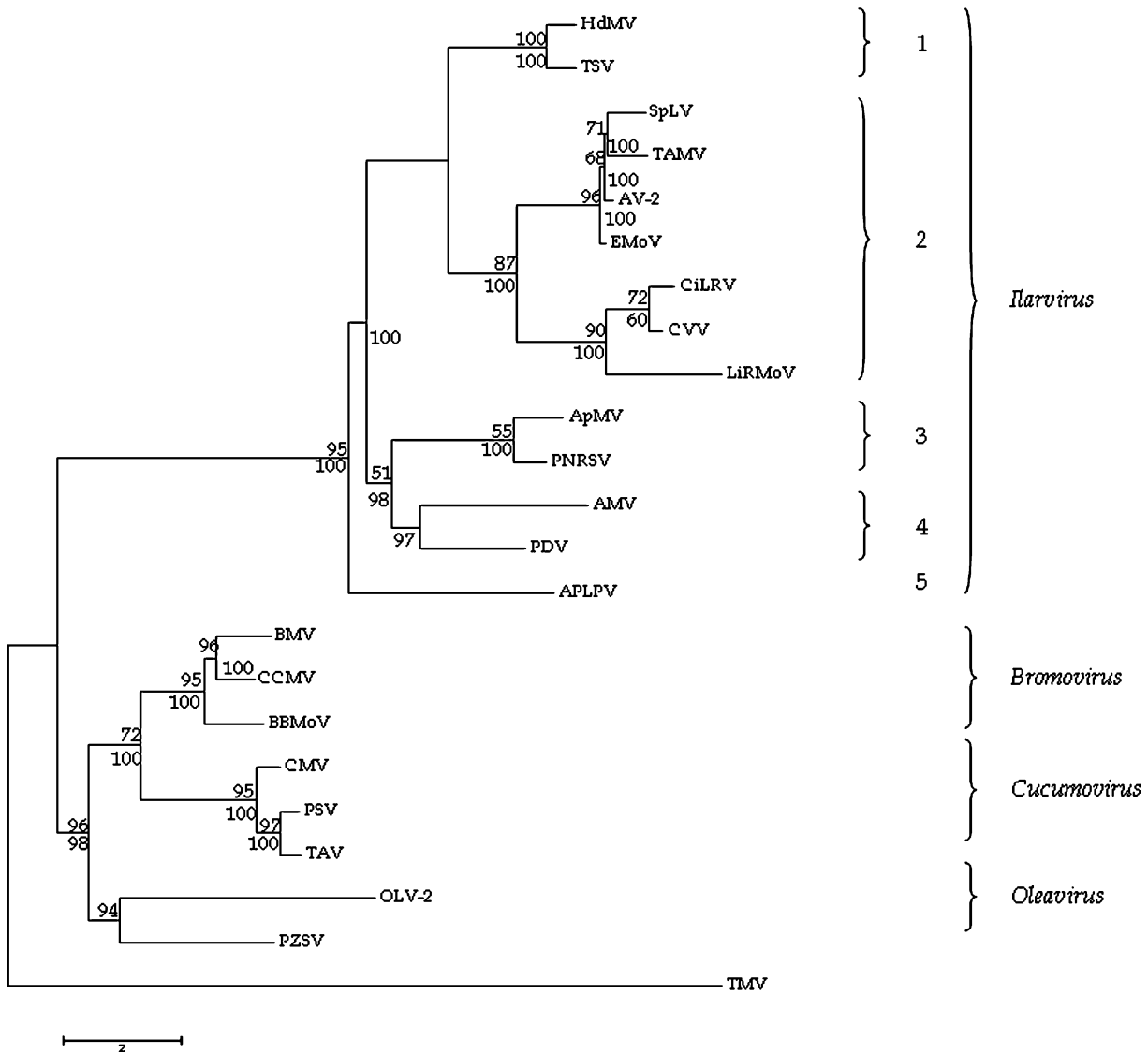


Fig. 2. Phylogenetic tree that best describes the evolution of the MP + CP genes. The branch lengths were computed according to a heterogeneous WAG + F + ad Γ + I model of protein evolution with different π_i , α , and ρ parameters for each protein. The numbers above each branch are the (QPP) probabilities, whereas those

below represent BPP. Only values with $P > 0.5$ are shown. Numbering of the different subgroups delineated for the *Ilarvirus* genus has been done following the grouping that best fits the one described by the ICTV (van Regenmortel et al. 2000).

so-called WAG + F + d Γ + I. In this model the WAG transition matrix is used to model substitution rates, equilibrium amino acid frequencies (π_i) are estimated as free parameters (F), heterogeneity among sites is incorporated by using d Γ , and a fraction of amino acid sites remains invariable along evolutionary time (I).

Once this model was chosen, we used it to infer phylogenetic trees using quartet-puzzling maximum likelihood and Bayesian inference for both proteins as well as for the concatenated data set. Regardless of the tree-making method employed, the tree topology obtained was the same for both proteins as well as for the concatenated data set. Figure 2 shows this tree topology, along with the statistical support obtained

by QPP (above branches) and the BPP (below branches) computed for the concatenated data set. Both methods significantly separated the *Ilarvirus* from the rest of the members of the family as a monophyletic group. According to QPP, there were two clusters within the *Ilarvirus* which are not well resolved ($P < 0.5$). The first corresponds with the splitting between groups 1 + 2 and groups 3 + 4 (Fig. 2). The second corresponds with the splitting between AMV and PDV. However, these two nodes were highly significant ($P = 1$ and $P = 0.97$, respectively) according to their BPPs. The clustering within the *Ilarvirus* challenges the ICTV subgroups classification (Table 1) in two ways. First, ICTV group 7 disappears, since LiRMoV clusters within subgroup 2 with very high

Table 2. Increasing the complexity of the homogeneous WAG + F + d Γ + I model: The number of free parameters (p), the most relevant parameters, the log-likelihood (l), and the BIC are reported for each model

Parameterization	P		l	BIC
= π , = α (protein homogeneity)	64	$\alpha = 3.392 \pm 0.344$	-24,064.553	48,545.380
= π , $\neq \alpha$	65	$\alpha_{MP} = 2.350 \pm 0.283$ $\alpha_{CP} = 7.254 \pm 1.530$	-24,050.768	48,524.314
$\neq \pi$, $\neq \alpha$ (independent proteins)	85	$\alpha_{MP} = 2.016 \pm 0.248$ $\alpha_{CP} = 6.131 \pm 1.236$	-23,720.117	47,993.099
$\neq \pi$, $\neq \alpha$, $\neq \rho$	87	$\alpha_{MP} = 0.915 \pm 0.218$ $\rho_{MP} = 0.966 \pm 0.013$ $\alpha_{CP} = 0.737 \pm 0.322$ $\rho_{CP} = 0.990 \pm 0.007$	-23,679.426	47,924.725

probability values. Second, according to BPP, AMV clearly belongs to the *Ilarvirus*, more concretely, to subgroup 4 along with PDV, and thus, its taxonomic status should be revised. The robustness of this result is explored in depth later.

We also found that PZSV, whose taxonomic status was not fully resolved before, belongs to the *Oleavirus*. Nonetheless, the corresponding node only has enough statistical support according to the QPP criterion ($P = 0.94$), and not according to the BPP one.

Refining the Model of Amino Acid Substitution

Next we explored whether modifications of the homogeneous WAG + F + d Γ + I substitution model provided a better fit to our data. Table 2 shows the relevant parameters for each of the increasing-complexity models tested. First, using the maximum likelihood approach encoded in CODEML we computed the parameters of the model under the assumptions that (i) the π_i values for both proteins are equal and (ii) the α parameter that characterizes the shape of the d Γ distribution is also the same for both proteins. The first obvious implementation to the basic homogeneous WAG + F + d Γ + I model consists in assigning different values of α to each protein but still assuming identical π_i . This model fitted significantly better to the data set, as shown by the lower BIC (Table 2). Next we tested whether a completely heterogeneous model, i.e., both α and π_i differing among proteins, significantly improved the goodness of fit. This heterogeneous model provided a better fit despite requiring 21 more free parameters, as shown by the lower BIC (Table 2). The final level of complexity explored was the replacement of the d Γ by an auto-d Γ (ad Γ) (Yang 1995). In addition to α , the ad Γ incorporates a second parameter, ρ , which is the correlation among rates at adjacent sites. Incorporating the ad Γ to the heterogeneous model has associated the lowest BIC among all models tested (Table 2). The estimates of the α parameters of the ad Γ for each gene suggest that the distribution of rates is highly skewed ($\alpha < 1$) and, therefore, sub-

stitution rates vary among sites for both proteins. Indeed, the degree of heterogeneity among sites was 24.1% higher for CP than for MP. The estimates of the ρ parameters for each gene suggest that the substitution rates of adjacent amino acid sites are not independent, but instead they change in a correlated manner (significant ρ values). Amino acid sites in the CP covary 2.5% more tightly than in the MP ($\rho_{MP} < \rho_{CP}$).

From the ad Γ distributions fitted, it is possible to estimate the expected amino acid substitution rates for each individual site in the proteins. Figure 3 shows the distribution of these rates along each protein. For the MP, the median substitution rate is 0.520 (range, 0.026–3.830). However, visual inspection of Fig. 3 suggests that the amino acid sites at the C-terminus of the molecule have, on average, higher substitution rates than those at the N-terminus. Indeed, after dividing the molecule into two fragments at residue 314, the median substitution rate for the N-terminus is 0.463 (range, 0.026–1.396), whereas for the C-terminus it is 2.591 (range, 1.250–3.830). In other words, the amino acids at the C-terminus part of the MP evolved on average, 5.60 times faster than those at the N-terminus, this difference being statistically significant (Mann–Whitney test: $U = 15$, $P < 0.001$). For the CP, the median substitution rate is 2.545 (range, 0.066–4.080), and apparently, substitution rates distribute evenly along the CP (Fig. 3). On average, CP sites evolve 5.50 times faster than sites at the MP N-terminus ($U = 1479.5$, $P < 0.001$) but at the same rate as amino acids at the MP C-terminus ($U = 7504$, $P = 0.286$).

Branch lengths shown in Fig. 2 were estimated by maximum likelihood fitting the heterogeneous WAG + F + ad Γ + I model to the tree topology previously obtained.

Test of the Molecular Clock Hypothesis

The first test of the molecular clock (i.e., constancy of evolutionary rates) can be done using the CODEML program. When the program was asked to fit the

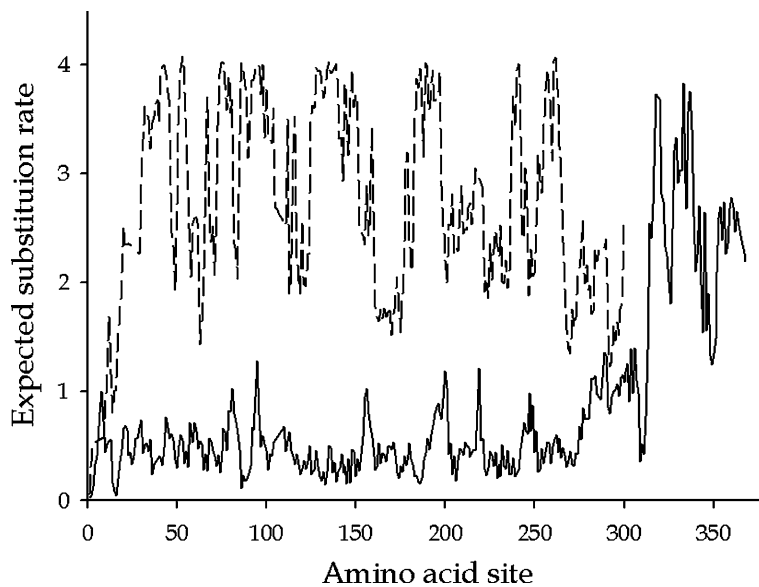


Fig. 3. Distribution of the estimated amino acid substitution rates for each site along the MP (solid line) and CP (dashed line) proteins.

heterogeneous WAG + F + ad Γ + I model but with contemporary tips to the topology shown in Fig. 2, the log-likelihood was $-25,301.888$. For this log-likelihood, we computed $BIC = 51,013.545$, which was higher than the value obtained without the molecular clock assumption. Therefore, we should reject the hypothesis of a global molecular clock during the radiation of the *Bromoviridae* family.

However, it is still possible that certain branches had evolved at a constant rate, whereas others had suffered episodes of rapid evolution. To test this possibility, we applied the two-cluster test. The tree topology and branch lengths used were those shown in Fig. 2. The results show acceleration in four branches. The first branch with a faster amino acid substitution rate was the one leading to the (LiRMoV, (CVV, CiLRV)) cluster, which evolved faster than the branch leading to its sister clade within *Ilarvirus* group 2 ($z = 2.291$, $P = 0.011$). The lineage leading to *Ilarvirus* group 3 (ApMV, PNRSV) evolved faster than the one leading to its sister group 4 (AMV, PDV) ($z = 2.150$, $P = 0.016$). The branch leading to APLPV also showed a faster evolution than the branch leading to the rest of the *Ilarvirus* ($z = 2.129$, $P = 0.017$). Finally, within the *Cucumovirus*, the branch leading to (TAV, PSV) had a faster rate of protein evolution than the branch leading to CMV ($z = 3.875$, $P < 0.001$).

Redefining Phylogenetic Units Within the *Bromoviridae*

Based on the phylogenetic tree shown in Fig. 2, the existence of four monophyletic clades within the family *Bromoviridae*—*Ilarvirus*, *Bromovirus*,

Cucumovirus, and *Oleavirus*—can be considered. This result would suggest a modification of the previous taxonomic status (van Regenmortel et al. 2000) in two major features: (i) AMV should be considered as a member of the *Ilarvirus*, and (ii) the division of the *Ilarvirus* genus into five subgroups reflecting a common phylogeny of their RNA 3 rather than into the seven phenetic subgroups classically defined on antigenic differences (van Regenmortel et al. 2000). Perhaps the most striking modification would be the exact status of AMV. Our phylogenetic analysis undoubtedly supports AMV as a member of the *Ilarvirus*. Notice that both the QPP and the BPP significantly supported the monophyly of the genus (Fig. 2). However, despite the robustness of these two methods, it is convenient to test whether the same conclusion can be obtained using a different statistical approach. To do so, we used the four-cluster likelihood-mapping technique. (Due to limitations in the TREEPUZZLE software, proteins were analyzed using a homogeneous WAG + F + d Γ model.) Figure 4 shows the results of this analysis. Ninety-nine percent of all the quartets between the four corresponding clusters supported the branching pattern that groups AMV and the *Ilarvirus* versus the remaining sequences. Therefore, based on likelihood mapping, we give extra support to the hypothesis that AMV is a true *Ilarvirus*.

Discussion

On the Patterns of Molecular Evolution

Our analyses suggest that the model of amino acid substitution that better describes the evolution of the

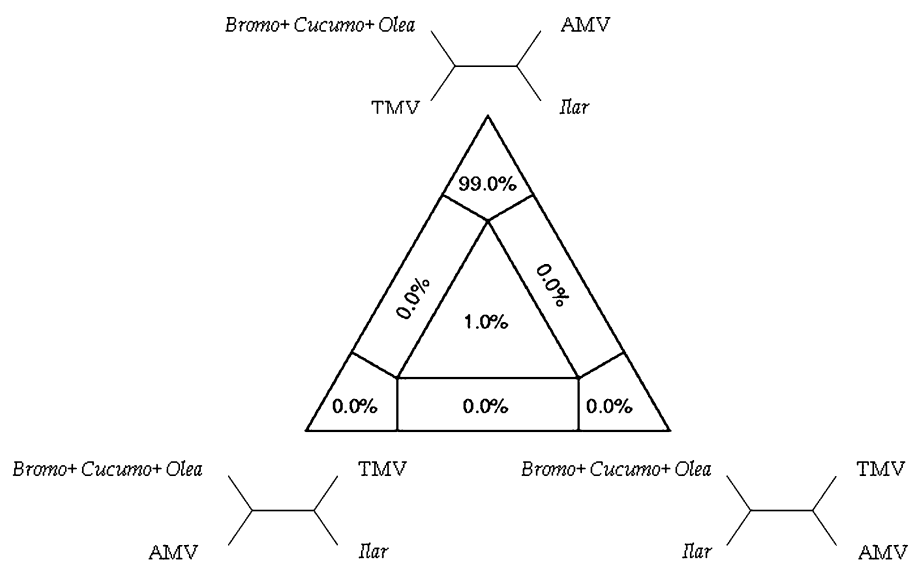


Fig. 4. Quartet likelihood-mapping analysis supporting the inclusion of AMV as a member of the *Ilarvirus* genus instead of as an independent genus (*Alfamovirus*).

MP and CP proteins is the one proposed by Wheland and Goldman (2001) to describe the evolution of globular proteins. Hence, it would be interesting to contrast whether this result is consistent with known structural data. Unfortunately, very little information is available on the crystallographic structure of these two proteins. Nonetheless, Kumar et al. (1997) obtained the structure of the CP from AMV at a resolution of 4 Å. At this resolution, it is possible to see that CP folds as a globule with both the C- and the N-terminus emerging from it. These two protrusions appear primarily to participate in the intersubunit interaction rather than contribute to the overall structure of the CP monomer (Choi and Loesch-Fries 1999; Tenllado and Bol 2001). Hence, these preliminary structural data agree with our model of globular protein evolution, giving support to its applicability.

We identified the presence of fewer amino acid sites with a strong tendency to change in the MP than in the CP ($\alpha_{MP} > \alpha_{CP}$). Furthermore, we found that whereas for the CP the substitution rates were evenly distributed along the molecule, this was not the case for the MP. Instead, amino acids near the MP C-terminus evolved faster than those at the N-terminus. An obvious explanation for this difference is that each part of the molecule is experiencing different functional constraints. Several lines of empirical evidence support this notion. The majority of AMV and BMV's N-terminal MP was required for the correct formation of microtubules, making any alteration in this region critical for the right interaction with host proteins (Sánchez-Navarro and Bol 2001). By contrast, deleting up to 44 amino acids from AMV's MP C-terminus apparently did not affect the ability to form microtubules but prevented the interaction with the CP, significantly reducing the specificity of virus

transport (Sánchez-Navarro and Bol 2001). This specific interaction between the C-termini of the MP and its cognate CP has also been demonstrated for BMV and CMV (Nagano et al. 1997; Takeda et al. 2004). Moreover, it has been reported that the variability of MP of different PNRSV isolates clusters mainly at its C-terminus (Hammond and Crosslin 1998; Aparicio and Pallás 2002). Taken together, these observations suggest a scenario for the MP in which the evolution of the N-terminus is governed by host requirements, whereas C-terminus evolution is driven by its interaction with the highly variable CP. Supporting the case of coevolution between both RNA 3-encoded proteins, we have found a similar evolutionary rate for the entire CP and the MP C-terminus (Fig. 3). It has been well documented that interacting proteins, or parts of proteins, evolve at similar rates to preserve the right interactions (Fraser et al. 2004a, b).

Our analyses were not consistent with the existence of a constant molecular evolutionary rate. Instead, we found that the rates of amino acid substitution have changed along lineages. Because RNA polymerase has no repair activity, it is unlikely that intrinsic rate of mutation will vary substantially between members of the *Bromoviridae* or that they have been slower or faster in different moments in evolutionary history. Instead, we envision two possible alternative explanations for this difference in amino acid substitution rates. The first possibility is that substantial rate acceleration was associated with jumps between host species. Different host species might impose different restrictions on viral replication, systemic movement, and transmission, and hence, amino acid substitutions would be quickly selected. The second possibility is that clock rates could vary between *Bromoviridae* if there are major

differences in rates of replication, that is, in virus generation times. Hence, if viruses differ dramatically in replication rate or have experienced periods when replication was either latent or extremely slow, this would reduce substitution rates in the long run in certain branches.

Phylogenetic Implications

Based on their serological relationships, *Ilarvirus* had been historically classified into 10 subgroups (Murphy et al. 1995). A recent reevaluation of the serological relationships led to a reduction to seven groups (van Regenmortel et al. 2000). Interestingly, this new clustering is very similar to the one proposed in our work with the exception of LiRMoV, which appears to be phylogenetically included in subgroup 2 in our analysis, whereas it was considered to belong to serological subgroup 7 together with *Fragaria chiloensis* latent virus (FCILV). The molecular characterization of FCILV would help to clarify this issue.

The topology presented in this study includes AMV in the *Ilarvirus* genus. This observation is supported by evidence from two sources. Functional analyses showed that the MP and/or CP coding sequences or the subgenomic promoter and 3' NTR of AMV can be exchanged by the corresponding ones of other unrelated *Ilarvirus* (Reusken et al. 1995; Sánchez-Navarro et al. 1997; Aparicio et al. 2001, 2003). Compatible exchange of MP full-length coding sequences between viruses of the *Bromoviridae* has been reported only between viruses of the same genera: CCMV and BMV (Mise and Ahlquist 1995; Sasaki et al. 2003). Such exchange between viruses of different genera was possible only when the C-terminus of the MP gene was deleted (Nagano et al. 1997). In a different context, quartet likelihood-mapping analysis, especially suitable for testing the internal edges of a tree, strongly supports the inclusion of AMV in the *Ilarvirus*. Taken together, the molecular evidence and the phylogenetic analysis reported so far support the idea that AMV should be considered as a true *Ilarvirus*. However, despite this evidence, there are several differences between AMV and the other *Ilarvirus*: (i) unlike the other *Ilarvirus*, AMV is transmitted by aphids; (ii) AMV virions are generally bacilliform, having a constant diameter of 18 nm, whereas those of the rest of *Ilarvirus* are quasi-isometric and about 30 nm in diameter; and (iii) there are differences in host range—whereas most *Ilarvirus* infect mainly woody plants, AMV mainly infects herbaceous hosts. Whether these biological and structural features must prevail over the phylogenetic relationships inferred from genomic sequences requires further discussion.

Acknowledgments. We thank John W. Randles and Mario A. Fares for critical reading of the manuscript. We also thank two anonymous reviewers for very helpful comments. F.M.C., J.M.C., and J.A.S.N. acknowledge fellowships from the Spanish CSIC (Bioinformatics I3P program), the Wellcome Trust, and the Ramón y Cajal program (MEC), respectively. This study was supported by Grant BIO02-04099-C02-02 (MCyT-FEDER) to V.P. and Grant BMC2003-00066 (MEC-FEDER), Grant GV04B280 (Generalitat Valenciana), and the EMBO Young Investigator Program to S.F.E.

References

- Abascal F, Zardoya R, Posada D (2004) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105
- Aparicio F, Pallás V (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 subgenomic RNA. *Virus Genes* 25:75–84
- Aparicio F, Sánchez-Navarro JA, Olsthoorn RC, Pallás V, Bol JF (2001) Recognition of *cis*-acting sequences in RNA 3 of *Prunus necrotic ringspot virus* by the replicase of *Alfalfa mosaic virus*. *J Gen Virol* 82:947–951
- Aparicio F, Vilar M, Pérez-Payá E, Pallás V (2003) The coat protein of *Prunus necrotic ringspot virus* specifically binds to and regulates the conformation of its genomic RNA. *Virology* 313:213–223
- Bol JF, van Vloten-Doting L, Jasper EMJ (1971) A functional equivalence of top component *a* RNA and coat protein in the initiation of infection by *Alfalfa mosaic virus*. *Virology* 46:73–85
- Choi J, Loesch-Fries LS (1999) Effect of C-terminal mutations of *Alfalfa mosaic virus* coat protein on dimer formation and assembly *in vitro*. *Virology* 260:182–189
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2004a) Evolutionary rate in the protein interaction network. *Science* 296:750–752
- Fraser HB, Hirsh AE, Wall DP, Eisen MB (2004b) Coevolution of gene expression among interacting proteins. *Proc Natl Acad Sci USA* 24:9033–9038
- Hammond RW, Crosslin JM (1998) Virulence and molecular polymorphism of *Prunus necrotic ringspot virus* isolates. *J Gen Virol* 79:1815–1823
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17:754–755
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *CABIOS* 8:275–282
- Kumar A, Reddy VS, Yusibov V, Chipman PR, Hata Y, Fita I, Fukuyama K, Rossmann MG, Loesch-Fries LS, Baker TS, Johnson JE (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
- Melcher U (2000) The '30K' superfamily of viral movement proteins. *J Gen Virol* 81:257–266
- Mise K, Ahlquist P (1995) Host-specificity restriction by *Bromovirus* cell-to-cell movement protein occurs after initial cell-to-cell spread of infection in nonhost plants. *Virology* 206:276–286
- Murphy FA, Fauquet CM, Bishop DHL, Ghabril SA, Jarvis AW, Martelli GP, Mayo MA, Summers MO (1995) Classification and nomenclature of viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag, Berlin
- Nagano H, Okuno T, Mise K, Furusawa I (1997) Deletion of the C-terminal 33 amino acids of *Cucumber mosaic virus* movement

- protein enables a chimeric *Brome mosaic virus* to move from cell to cell. *J Virol* 71:2270–2276
- Opperdoes FR (2003) Phylogenetic analysis using protein sequences. In: Salemi M, Vandamme AM (eds) *The phylogenetic handbook*. Cambridge University Press, Cambridge
- Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol* 53:793–808
- Rampitsch C, Eastwell KC (1997) The complete sequence of *Prune dwarf ilarvirus* RNA-1. *Arch Virol* 142:1911–1918
- Reusken CBEM, Neeleman L, Bol JF (1995) Ability of *Tobacco streak virus* coat protein to substitute for late functions of *Alfalfa mosaic virus* coat protein. *J Virol* 69:4552–4555
- Sánchez-Navarro JA, Bol JF (2001) Role of the *Alfalfa mosaic virus* movement protein and coat protein in virus transport. *Mol Plant Microbe Inter* 14:1051–1062
- Sánchez-Navarro JA, Pallás V (1997) Evolutionary relationships in the ilarviruses: nucleotide sequence of *Prunus necrotic ringspot virus* RNA 3. *Arch Virol* 142:749–763
- Sánchez-Navarro JA, Reusken CBEM, Bol JF, Pallás V (1997) Replication of *Alfalfa mosaic virus* RNA 3 with movement and coat protein genes replaced by corresponding genes of *Prunus necrotic ringspot ilarvirus*. *J Gen Virol* 78:3171–3176
- Sasaki N, Arimoto N, Nagano H, Mori M, Kaido M, Mise K, Okuno T (2003) The movement protein gene is involved in the virus-specific requirement of the coat protein in cell-to-cell movement of bromoviruses. *Arch Virol* 148:803–812
- Scott SW, Zimmerman MT, Ge X (1998) The sequence of RNA 1 and RNA 2 of *Tobacco streak virus*: additional evidence for the inclusion of *Alfalfa mosaic virus* in the genus *Iilarvirus*. *Arch Virol* 143:1187–1198
- Shiel PJ, Berger PH (2000) The complete nucleotide sequence of *Apple mosaic virus* (ApMV) RNA 1 and RNA 2: ApMV is more closely related to *Alfalfa mosaic virus* than to other ilarviruses. *J Gen Virol* 81:273–278
- Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol Biol Evol* 13:964–969
- Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci USA* 94:6815–6819
- Takeda A, Kaido M, Okuno T, Mise K (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
- Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol* 12:823–833
- Tenllado F, Bol JF (2000) Genetic dissection of the multiple functions of *Aalfalfa mosaic virus* coat protein in viral RNA replication, encapsidation, and movement. *Virology* 268:29–40
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, McGeoch DJ, Maniloff J, Mayo MA, Pringle CR, Winckner RB (2000) *Virus taxonomy. Classification and nomenclature of viruses*. Seventh ICTV Report. Academic Press, San Diego
- Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691–699
- Yang Z (1995) A space-time process model for the evolution of DNA sequences. *Genetics* 139:993–1005
- Yang Z (1996) Maximum-likelihood models for combined analyses of multiple sequence data. *J Mol Evol* 42:587–596
- Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13:555–556
- Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol Biol Evol* 15:1600–1611