

## Research Article

# New Insights on the Evolutionary History of Aphids and Their Primary Endosymbiont *Buchnera aphidicola*

Vicente Pérez-Brocal,<sup>1,2</sup> Rosario Gil,<sup>2,3</sup> Andrés Moya,<sup>1,2,3</sup> and Amparo Latorre<sup>1,2,3</sup>

<sup>1</sup>Área de Genómica y Salud, Centro Superior de Investigación en Salud Pública (CSISP), Avenida de Cataluña 21, 46020 Valencia, Spain

<sup>2</sup>CIBER Epidemiología y Salud Pública (CIBERESP), Spain

<sup>3</sup>Departament de Genètica, Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado Postal 22085, 46071 Valencia, Spain

Correspondence should be addressed to Vicente Pérez-Brocal, perez\_vicbro@gva.es

Received 13 October 2010; Accepted 24 December 2010

Academic Editor: Hiromi Nishida

Copyright © 2011 Vicente Pérez-Brocal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since the establishment of the symbiosis between the ancestor of modern aphids and their primary endosymbiont, *Buchnera aphidicola*, insects and bacteria have coevolved. Due to this parallel evolution, the analysis of bacterial genomic features constitutes a useful tool to understand their evolutionary history. Here we report, based on data from *B. aphidicola*, the molecular evolutionary analysis, the phylogenetic relationships among lineages and a comparison of sequence evolutionary rates of symbionts of four aphid species from three subfamilies. Our results support previous hypotheses of divergence of *B. aphidicola* and their host lineages during the early Cretaceous and indicate a closer relationship between subfamilies Eriosomatinae and Lachninae than with the Aphidinae. They also reveal a general evolutionary pattern among strains at the functional level. We also point out the effect of lifecycle and generation time as a possible explanation for the accelerated rate in *B. aphidicola* from the Lachninae.

## 1. Introduction

Aphids constitute a diversified group of insects widespread and of economical relevance as crop pests. The underlying reason of their ecological success is their novel capability to exploit ecological niches with little competitors, mainly due to their diet based on phloem, which is abundant and of easy access but represents an unbalanced source of nutrients, rich in sugars and poor in amino acids [1]. The clue to the use of new resources lies in the establishment of an obligate endosymbiotic relationship between the ancestor of aphids and a gamma-proteobacterium, the ancestor of *Buchnera aphidicola*. This single event of infection has been dated at least 150–200 million years ago (MYA) [2] according to the fossil record or to 80–150 MYA based on molecular data [3]. As a result of millions of years of cospeciation of host and endosymbiont, the current species of aphids carrying their specific strains of *B. aphidicola* emerged.

The vertical mode of transmission of *B. aphidicola*, from mother to eggs and embryos, together with the location in specific host cells (the bacteriocytes), determines a population scenario for this bacterium characterized by their low effective population size, with frequent bottlenecks and little chance of genetic recombination with other bacteria. As a result, the genome reductive process undergone by *B. aphidicola* encompasses a decrease in the genomic size due to the loss of unnecessary genes in the new intracellular context, the increase in A+T content compared to its free-living relatives, a significant acceleration in evolutionary rates, mainly due to the accumulation of nonsynonymous substitutions, the loss in codon bias, loss of many regulatory proteins and functions, as well as the retention of genes linked to their symbiotic role [4–9].

This particular history of genome reduction is pertinent to understand the coevolution between particular aphid hosts and *B. aphidicola*. Many of the genes that are involved

in recombination and/or genetic transference were lost at the beginning of the symbiotic association and, consequently, the *B. aphidicola* clones have evolved independently in each particular host with no or little chance of gene exchange among *B. aphidicola* from different aphid hosts [10].

The comparison of the topology of phylogenetic trees based on aphid genes and those from *B. aphidicola* reveals a perfect match [2, 11]. As a result of this parallel evolutionary pattern, *B. aphidicola* can be regarded as an excellent marker in order to elucidate the evolutionary relationship of aphids harboring particular *B. aphidicola* strains.

The analysis of *B. aphidicola* genes that follow an evolutionary pattern that agrees with the molecular clock hypothesis [12, 13] can be used to estimate the divergence time between pairs of aphids. This is possible because two aphid species, *Acyrtosiphon pisum* and *Schizaphis graminum* belonging to two tribes of the subfamily Aphidinae, have an estimated divergence time calibrated from their fossil record of 50 to 70 MY [14]. In addition, using molecular data from complete *B. aphidicola* genomes available, Pérez-Brocal and coworkers calculated the divergence time of aphids belonging to subfamilies Eriosomatinae (*Baizongia pistaciae*) and Lachninae (*Cinara cedri*) [15]. Based on morphological traits, the subfamilies Eriosomatinae and Lachninae have traditionally been considered very divergent. In fact, most phylogenetic hypotheses based both on morphological and molecular data consider the Lachninae as a sister group of the Aphidinae [11, 16]. However, the position of this subfamily remains controversial, as recent phylogenies based on molecular sequences located the subfamily in a basal position [17–19]. Here, we follow a genomic approach to deepen the evolutionary analyses and propose a phylogeny of the three subfamilies of aphids based on the genome sequence of their primary endosymbionts *B. aphidicola*. In addition, in order to detect if there is any selective effect related to the specific role of the genes, we also gave a closer look to the acceleration pattern of each functional category.

## 2. Materials and Methods

**2.1. Genome Sequences Used in This Study.** The genome sequences used in this study were retrieved from GenBank database. The four *B. aphidicola* strains are *B. aphidicola* *Acyrtosiphon pisum* str. APS (BAp, Accession no. BA000003 [20]), *B. aphidicola* *Schizaphis graminum* (BSg, Accession no. AE013218 [21]), *B. aphidicola* *Baizongia pistaciae* (BBp, Accession no. AE016826 [22]), and *B. aphidicola* *Cinara cedri* (BCc, Accession no. CP000263 [15]). *Escherichia coli* was used as out-group in all comparisons: *E. coli* str. K12 substr. MG1655 (Eco, Accession no. U00096).

**2.2. Sequence Alignments.** For protein-coding genes, nucleotide sequences were translated into amino acids using the ClustalW tool implemented in the MEGA4 package [23]. The generated amino acid sequences were used, in turn, as a template to align the corresponding nucleotides with MUSCLE v3.6 [24], to reduce ambiguities.

**2.3. Estimate of Strain-Specific Evolutionary Rates.** *B. aphidicola* BCc was used as a reference strain since it is the one

with the lowest gene complement of those analyzed. For each one of the genes present in *B. aphidicola* BCc having an orthologous in at least one of the other *B. aphidicola* strains, an analysis of relative substitution rates between pairs of *B. aphidicola* strains was carried out, using *E. coli* as out-group. Specifically, we applied a Tajima's relative rate test [25] with MEGA4, generating six comparisons for each of the aligned genes. Genes showing accelerated rates were grouped according to a nonredundant categories classification based on that used in the sequencing work on *Aquifex aeolicus* [26], with some modifications [27].

### 2.4. Estimate of Evolutionary Acceleration among Genomes.

The sequence from the 338 protein-coding genes shared by the four *B. aphidicola* strains plus *E. coli* was used to quantify the relative degree of evolutionary acceleration among strains. To do this, nucleotide sequences were concatenated with BioEdit and aligned using the ClustalW tool implemented in the MEGA4 [23]. Three different estimates of substitution rates per site between species *i* and *j* ( $K_{ij}$ ) were carried out with MEGA4, using (a) the total and (b) nonsynonymous nucleotide positions, under the Kimura 2-parameters and the modified Nei-Gojobori methods, respectively, and (c) amino acid sequences, using the JTT substitution matrix.  $K_{01}$  and  $K_{02}$  were calculated according to Moran [28], being taxon 0 the last common ancestor of the endosymbiont strains compared in each test (taxa 1 and 2). The calculation of total and nonsynonymous substitutions allowed us to account for the phenomenon of saturation. To check for saturation, the "transition and transversion versus divergence" plot was implemented by DAMBE v4.2.13 for the concatenation of shared genes using the first and second positions as well as the third one [29]. This method has been successfully used previously to estimate saturation due to divergence [30–33].

Additionally, for each protein-coding gene under study, the values of both synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) nucleotide substitutions were calculated, using a modified Nei-Gojobori model (Jukes Cantor) implemented by MEGA4 [23]. To calculate the synonymous ( $\lambda_S$ ) and nonsynonymous ( $\lambda_N$ ) nucleotide substitutions per million years, we used the expression  $\lambda = K/2T$ , where  $K$  is the number of nucleotide differences per site and  $T$  the estimated divergence time. The  $T$  values used in these analyses were 107 MY for (*B. aphidicola* BAp-BSg-)BBp, 111 MY for (*B. aphidicola* BAp-BSg-)BCc, and 112 MY for *B. aphidicola* BBp-BCc. These are the previously determined lowest values for each range of estimated divergence times among strains [15], based on the range of 50 to 70 MY since the strains used for calibration (*B. aphidicola* BAp and BSg) diverged as estimated from the fossil record [14]. The global average  $\lambda_S$  and  $\lambda_N$  values for each pair of *B. aphidicola* strains was calculated, as well as the partial average  $\lambda_S$  and  $\lambda_N$  values for each functional category [26, 27] between all the strain pairs.

**2.5. Phylogenetic Analyses.** Since saturation was achieved at the third position in all comparisons but BAp and BSg, in order to reduce the loss of phylogenetic signal we excluded this position when working with nucleotides to perform

our phylogenetic analyses. The concatenated sequence of the 338 protein-coding genes shared by the four *B. aphidicola* strains was used to reconstruct the phylogenetic relationships among them. Maximum Likelihood (ML) analyses were carried out with PAUP4.0b10 [34] for nucleotides, and PhyML.v2.4.5 [35] for amino acids, according to the best models of nucleotide (GTR+I+G) and amino acid (CpREV+I+G+F) substitutions for those genes derived from jModelTest [36] and ProtTest 1.4 [37], respectively. Nucleotides and amino acids were also used for Bayesian analysis, with MrBayes v3.1.2 [38], using four MCMC strands, 1,000,000 generations, with trees sampled every 100 generations. Consensus trees were produced after excluding an initial burn-in of 25% of the samples, as recommended.

In a previous study, the evolutionary analyses of the four *B. aphidicola* strains showed that only 21 genes fulfill the molecular clock hypothesis [15]. The topologies of the 21 phylogenetic trees based on these genes were obtained by ML using PAUP 4.0b10 [34], in order to determine the most plausible evolutionary relationships among strains and compare them with the phylogenetic reconstruction.

**2.6. Statistical Analyses.** All statistical analyses were performed using the software package R (<http://www.r-project.org>) [39]. A chi-square analysis was applied to the global distribution of the accelerated genes among *B. aphidicola* strains compared to the distribution within functional families, to test whether any particular functional category contains a significantly increased or reduced number of accelerated genes. Twelve comparisons with Yates' correction were carried out, at a significance level  $\alpha = 0.05$ .

The average rates of synonymous ( $\lambda_S$ ) and nonsynonymous ( $\lambda_N$ ) substitutions per site per million years of the six possible comparisons among *B. aphidicola* strains were compared using a one-way ANOVA analysis followed by Tukey's range tests to find which means are significantly different from one another.

### 3. Results

**3.1. Comparison of the Evolutionary Rates in *B. aphidicola* Strains at a Genome Level.** The relative rate test on the 338 concatenated protein-coding genes (Table 1) reveals that, since the last common ancestor of each pair of strains, the accumulation of both nucleotide and amino acid substitutions, as well as the nonsynonymous substitution rates follows different rates in the different strains, but the values obtained using all three parameters are equivalents for any given strain pair. Thus, for the nucleotide sequences, a similar pattern of relative evolutionary rates was observed when total and nonsynonymous substitution rates are considered. *B. aphidicola* BSg and BAp show a similar rate (1.12:1), the one in *B. aphidicola* BBp being slightly higher (1.3-1.4-fold that of *B. aphidicola* BSg and BAp) and *B. aphidicola* BCc being the one with more accelerated rates (1.7-fold that of *B. aphidicola* BBp and more than 2-fold that of *B. aphidicola* BAp and BSg). As for the amino acid sequences, the relative acceleration shows a similar pattern as the one observed for

the nucleotides, but with values in *B. aphidicola* BCc of 2 to 3-fold those of *B. aphidicola* BBp and BAp-BSg, respectively.

The evolutionary acceleration among genomes was also determined through the analysis of the synonymous ( $\lambda_S$ ) and nonsynonymous ( $\lambda_N$ ) nucleotide substitutions per million years. The results show that both rates exhibit an opposite pattern (Figure 1). Differences in both  $\lambda_S$  and  $\lambda_N$  are statistically significant (ANOVA test, significance level 0.05), clustering into three separate groups for  $\lambda_S$  and two groups for  $\lambda_N$ , according to Tukey's range tests. When synonymous substitutions (Figure 1(a)) are considered, the more accelerated rate is found in the comparison between strains *B. aphidicola* BAp and BSg, a second group includes *B. aphidicola* BBp with the two aforementioned, and the least accelerated one includes all rates in which *B. aphidicola* BCc is involved. A different pattern is found for nonsynonymous substitutions (Figure 1(b)), where the more accelerated group includes all the comparisons involving *B. aphidicola* BCc, and the other one includes the remaining three comparisons.

**3.2. Analyses of the Evolutionary Rates at a Functional Level.** The general pattern identified at a genomic level is reproduced at every functional category (see Section 2), with the same three and two groups of *B. aphidicola* strain pairs found in  $\lambda_S$  and  $\lambda_N$ , respectively (Figure 2). On the other hand, no significant differences are found among functional categories in any strain for  $\lambda_S$  (Figure 2(a)). However, a significant increase in  $\lambda_N$  is found for the genes involved in cell envelope in all the strains ( $P < .05$ ) and to a lesser extent in the category of poorly characterized genes (Figure 2(b)). This could be due to a significant acceleration of the flagellar genes still remaining in *B. aphidicola*, especially in BCc, the strain which has undergone the most drastic reduction in the flagellar machinery.

In a previous study, we determined the global relative distribution of accelerated genes displayed by the strains, using Tajima's relative rate test [15]. According to this test, *B. aphidicola* BCc presents a higher number of accelerated genes (56%–83%), while *B. aphidicola* BBp presents intermediate values (0.6%–35%), and the fewest appear in *B. aphidicola* BSg and specially BAp. This trait is observed in each functional category with no significant differences (Figure 3). This homogeneous distribution of the accelerated genes across functional categories was tested by the application of  $\chi^2$  tests, based on the observed number of accelerated genes for each category and the expected number of genes based on the totality of them for each pair of strains. None of the tests was statistically significant at  $P < .05$  (Table 2).

**3.3. Phylogenetic Analyses Show an Evolutionary Radiation Pattern.** According to the molecular clock hypothesis, two taxa sharing a common ancestor should have accumulated the same number of substitutions since they diverge. In the *B. aphidicola* case, only 21 genes do not reject the molecular clock hypothesis [15]. These genes can be used to identify the phylogenetic relationships among the strains under study, which will also reflect the relationships among their insect hosts. However, three different tree topologies appear in

TABLE 1: Relative rate tests for the 338 concatenated protein-coding genes shared by the four *B. aphidicola* strains included in this study plus *E. coli*<sup>a</sup>: (a) nonsynonymous sites, (b) all nucleotides, and (c) amino acids.

(a)							
Taxon 1	Taxon 2	Taxon 3	$\bar{K}_{12}$	$\bar{K}_{13}$	$\bar{K}_{23}$	$\bar{K}_{13} - \bar{K}_{23}$	$\bar{K}_{01}/\bar{K}_{02}$
BAp	BSg	Eco	0.152	0.339	0.348	-0.009	0.89
BAp	BBp	Eco	0.319	0.339	0.395	-0.056	0.70
BAp	BCc	Eco	0.380	0.339	0.494	-0.155	0.42
BSg	BBp	Eco	0.319	0.348	0.395	-0.047	0.74
BSg	BCc	Eco	0.377	0.348	0.494	-0.146	0.44
BBp	BCc	Eco	0.392	0.395	0.494	-0.099	0.60
(b)							
Taxon 1	Taxon 2	Taxon 3	$\bar{K}_{12}$	$\bar{K}_{13}$	$\bar{K}_{23}$	$\bar{K}_{13} - \bar{K}_{23}$	$\bar{K}_{01}/\bar{K}_{02}$
BAp	BSg	Eco	0.242	0.617	0.630	-0.013	0.89
BAp	BBp	Eco	0.421	0.617	0.685	-0.068	0.72
BAp	BCc	Eco	0.452	0.617	0.791	-0.174	0.50
BSg	BBp	Eco	0.417	0.63	0.685	-0.055	0.77
BSg	BCc	Eco	0.445	0.63	0.791	-0.161	0.47
BBp	BCc	Eco	0.463	0.685	0.791	-0.106	0.62
(c)							
Taxon 1	Taxon 2	Taxon 3	$\bar{K}_{12}$	$\bar{K}_{13}$	$\bar{K}_{23}$	$\bar{K}_{13} - \bar{K}_{23}$	$\bar{K}_{01}/\bar{K}_{02}$
BAp	BSg	Eco	0.350	0.814	0.842	-0.028	0.85
BAp	BBp	Eco	0.845	0.814	1.001	-0.187	0.64
BAp	BCc	Eco	1.126	0.814	1.410	-0.596	0.30
BSg	BBp	Eco	0.850	0.842	1.001	-0.159	0.68
BSg	BCc	Eco	1.180	0.842	1.410	-0.568	0.33
BBp	BCc	Eco	1.186	1.001	1.410	-0.409	0.48

<sup>a</sup>In each test, taxa 1 and 2 represent *B. aphidicola* strains, taxon 3 represents *E. coli*, and taxon 0 represents the last common ancestor of taxa 1 and 2.  $\bar{K}_{ij}$  is the estimate of substitutions per site between taxon *i* and taxon *j*.

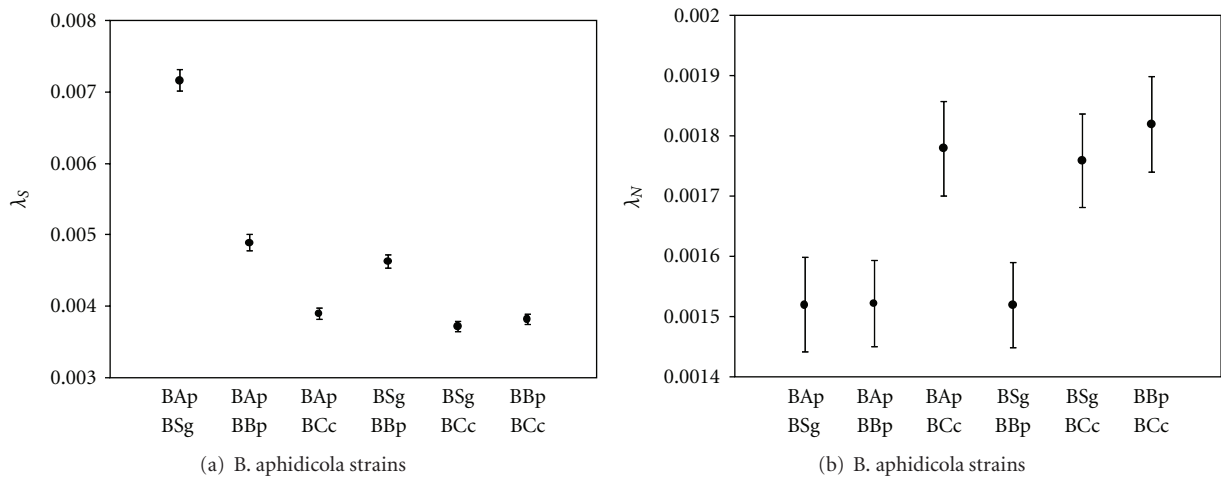


FIGURE 1: Global average values (and confidence interval of 95%) of (a) synonymous ( $\lambda_S$ ) and (b) nonsynonymous ( $\lambda_N$ ) nucleotide substitutions per site per million years. The divergence times among strains are 50 (BAp-BSg), 107 (BAp-BBp and BSg-BBp), 111 (BAp-BCc and BSg-BCc), and 112 (BBp-BCc) MY, respectively. The numbers of shared protein-coding genes are 348 (BAp-BSg), 347 (BAp-BBp), 354 (BAp-BCc), 343 (BSg-BBp), 350 (BSg-BCc), and 350 (BBp-BCc), respectively.

TABLE 2: Yates' chi-square tests for the accelerated genes classified by functional category in four *B. aphidicola* strains compared in pairs. Acceleration is based on Tajima's relative rate tests. The total number of comparisons for each particular category and pair of strains is shown in brackets ( ). A/B: number of accelerated genes in A compared to B and in B compared to A, respectively.

Observed Functional category	Pairs of strains					
	BAp/BSg	BAp/BBp	BAp/BCc	BSg/BBp	BSg/BCc	BBp/BCc
(1) Information storage and processing	5/12 (160)	0/54 (160)	0/137 (162)	0/47 (158)	0/122 (160)	1/86 (160)
(2) Protein processing, folding, and secretion	1/1 (25)	0/11 (24)	0/19 (25)	0/10 (24)	0/18 (25)	0/14 (24)
(3) Cellular processes	0/0 (10)	0/5 (10)	0/7 (10)	0/3 (10)	0/7 (10)	1/7 (10)
(4) Metabolism	2/9 (103)	3/34 (103)	0/86 (104)	2/32 (104)	0/88 (105)	0/61 (106)
(5) Cell envelope	0/0 (14)	1/7 (13)	0/12 (14)	1/7 (13)	0/12 (14)	0/10 (13)
(6) Poorly characterized	1/1 (33)	1/10 (34)	0/29 (35)	0/7 (32)	0/23 (33)	0/15 (34)
Total	9/23 (345)	5/121 (344)	0/290 (350)	3/106 (341)	0/270 (347)	2/193 (347)
Expected	Pairs of strains					
	BAp/BSg	BAp/BBp	BAp/BCc	BSg/BBp	BSg/BCc	BBp/BCc
(1) Information storage and processing	4.17/10.67	2.33/56.28	0.00/134.23	1.39/49.11	0.00/124.50	0.92/88.99
(2) Protein processing, folding and secretion	0.65/1.67	0.35/8.44	0.00/20.71	0.21/7.46	0.00/19.45	0.14/13.35
(3) Cellular processes	0.26/0.67	0.15/3.52	0.00/8.29	0.09/3.11	0.00/7.78	0.06/5.56
(4) Metabolism	2.69/6.87	1.50/36.23	0.00/86.17	0.91/32.33	0.00/81.70	0.61/58.96
(5) Cell envelope	0.36/0.93	0.19/4.57	0.00/11.60	0.11/4.04	0.00/10.89	0.07/7.23
(6) Poorly characterized	0.86/2.20	0.49/11.96	0.00/29.00	0.28/9.95	0.00/25.68	0.20/18.91
$\chi^2$ 's (with Yates' correction, 5 d.f.)	0.501/0.933	3.491/1.908	0.000/0.195	4.776/2.762	0.000/0.720	7.455/1.598
P value	.992/.968	.625/.862	1.000/.999	.444/.737	1.000/.982	.189/.902

a similar number of cases for these 21 genes (see Figure 4). Six genes generated the topology a (*B. aphidicola* BCc basal), seven the topology b (*B. aphidicola* BBp basal), and eight the topology c (*B. aphidicola* BCc and BBp clustered). Therefore, the analysis of these genes, individually considered, does not resolve the position of the *B. aphidicola* BCc and BBp strains. This result points at the possibility of a radiation within a relatively short period of time, giving rise to the subfamilies. To confirm this point, and in order to solve the deepest relationship among subfamilies, a more exhaustive phylogenetic reconstruction was carried out, based on all the concatenated protein-coding genes shared by the four *B. aphidicola* strains. The resulting phylogenetic tree (Figure 5) shows the same topology as tree c in Figure 4, that is, a well supported clade consisting of both members of the subfamily Aphidinae, as expected, and another clade that shows a clustering of *B. aphidicola* BBp and BCc, also with the maximum statistical support. The uneven branch length, being that of *B. aphidicola* BCc significantly longer, indicates the evolutionary acceleration experienced by this strain. The topology obtained using amino acid sequences is identical, but the relative length of *B. aphidicola* BCc's branch is even longer, reflecting a higher value of nonsynonymous substitutions.

## 4. Discussion

4.1. Reconstruction of the Evolutionary History of Aphids Belonging to Subfamilies Aphidinae, Eriosomatinae and Lachninae. Aphids emerged as a monophyletic group of

viviparous insects about 250 MYA as a divergent group from the oviparous Adelgidae and Phylloxeridae [11]. The basal radiation of the family Aphididae was dated by molecular data to the Cretaceous, 80 to 150 MYA [3]. Although the initial development of aphids took place on gymnosperms during the Mesozoic, most of their current diversity is linked to angiosperms, especially to grass [40]. The extraordinary diversity of aphids found today, affecting specially the subfamily Aphidinae, started during the Tertiary (Miocene), as a consequence of the proliferation of herbaceous angiosperms [41, 42].

The phylogenetic position of the subfamily Lachninae within the Aphididae is controversial. Traditionally, phylogenies based on both morphological characters [11, 16] and on mitochondrial rDNA [3] have placed them as a monophyletic group clustering with the Aphidinae. However, phylogenies based on sequences from both nuclear and mitochondrial aphid genes (long-wavelength opsin gene, the elongation factor 1 $\alpha$  gene, and mitochondrial genes encoding ATPase 6 subunit and the subunit II of the cytochrome oxidase), as well as those based on their primary endosymbiont *B. aphidicola* (16S rDNA and the  $\beta$  subunit of the F-ATPase complex) [17–19] place them as a basal group apart from the Aphidinae. This fact has implications about those aphids feeding on conifers (such as most members of the subfamily Lachninae, including *C. cedri*) being regarded as ancestral to groups feeding on angiosperms or, alternatively, as more recent secondarily derived conifer suckers.

Our phylogenetic analysis supports the presence of one clade clustering *B. aphidicola* BBp and BCc, and another

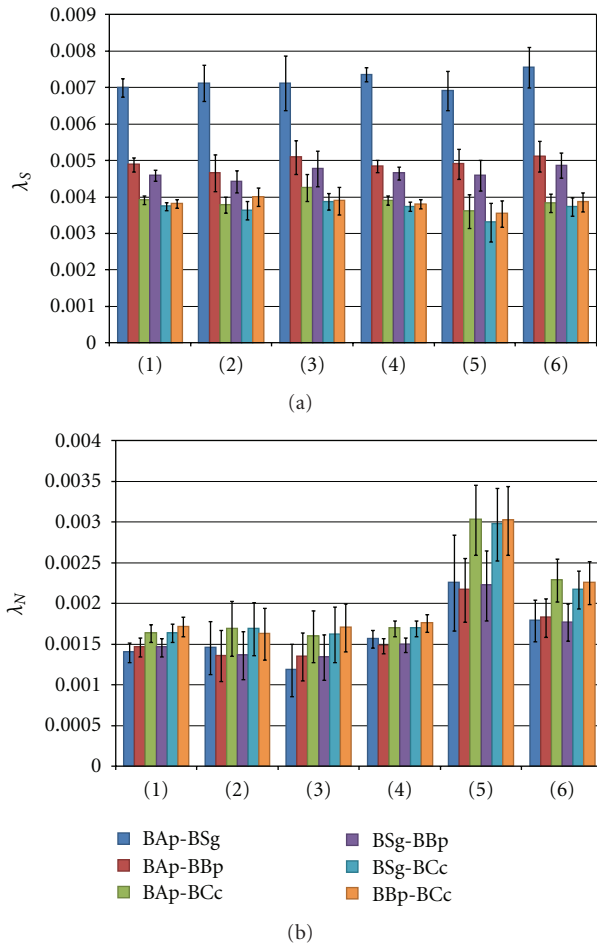


FIGURE 2: Average values (and confidence interval of 95%) of (a) synonymous ( $\lambda_S$ ) and (b) nonsynonymous ( $\lambda_N$ ) nucleotide substitutions per site per million years for each functional category. The numbers of shared protein-coding genes are 348 (BAp-BSg), 347 (BAp-BBp), 354 (BAp-BCc), 343 (BSg-BBp), 350 (BSg-BCc), and 350 (BBp-BCc), respectively. (1) Information storage and processing; (2) protein processing, folding, and secretion; (3) cellular processes; (4) metabolism; (5) cell envelope; (6) poorly characterized. Each given comparison is colored as illustrated above.

clade consisting of *B. aphidicola* BAp and BSg. This result is consistent with a panorama of a rapid evolutionary radiation of the main subfamilies of aphids, during the early Cretaceous (144-100 MYA), which seems concordant with previous proposals [3]. In addition, our evolutionary molecular data from *B. aphidicola* point out that aphids belonging to subfamilies Eriosomatinae and Lachninae share a common ancestor more closely related than compared to the members of subfamily Aphidinae. If true, our data refute the traditional phylogenetic reconstructions that placed Aphidinae and Lachninae as a monophyletic group [11]. However, we do not have evidence to conclude whether, within the subfamily Lachninae, tribes feeding on conifers are ancestral or more recent than those living on herbaceous angiosperms, since our analysis does not resolve which strain (and thus which host aphid) is basal compared to the others.

To solve this point, it would be necessary to sequence the genome of a greater number of *B. aphidicola* strains, including members of the different tribes from the subfamily Lachninae (work in progress). This would allow us to establish the date of divergence between those tribes and, thus, try to relate this fact to the change of vegetal host in either direction.

**4.2. Accelerated Evolutionary Rates in *B. aphidicola* within the Subfamily Lachninae.** From an evolutionary perspective, the protein-coding genes of *B. aphidicola* show higher ratios of nonsynonymous versus synonymous substitutions ( $d_N/d_S$ ) than those of free-living bacteria, due to an accelerated rate of nonsynonymous substitutions, a characteristic of bacterial endosymbionts [14, 28], where mutations with amino acid replacement are not efficiently eliminated by a relaxed purifying selection, leading to a greater accumulation of amino acid changes than in free-living bacteria. These nonsynonymous substitutions end up in fixation by genetic drift, due to the mode of transmission and the population dynamics of *B. aphidicola*. This acceleration of evolutionary rates is particularly evident in *B. aphidicola* BCc, presumably because factors promoting the accumulation of nonsynonymous substitutions are more intense in this strain. One of those factors is the extreme reduction of the repair machinery, barely able to counterbalance the accumulation of slightly deleterious mutations. In addition, there is a stronger effect of genetic drift that promotes the fixation of slightly deleterious mutations probably imposed by its coexistence within the aphid with a secondary symbiont, *Serratia symbiotica*, and its larger size compared to other *B. aphidicola* lineages [43]. A closer look at the particular genes that contribute to this acceleration observed in *B. aphidicola* BCc allows us to conclude that they are distributed among different functional categories, with none of them accumulating significant differences in the proportion of accelerated genes (as seen in Figure 3 and Table 2). This fact reveals that the process of gene degradation acts on any type of gene independently of their functional role. However, our results indicate that even if the accelerated genes are scattered homogeneously across all the functional categories in all *B. aphidicola* strains, genes of some functional categories, such as cellular envelope, are significantly more accelerated within all the lineages. That points to the ongoing action of selective constraints affecting nonsynonymous substitution rates.

Regarding synonymous substitutions, when pairs of strains of *B. aphidicola* were compared based on the average number of synonymous substitutions per site ( $d_S$ ), a greater accumulation was observed in the *B. aphidicola* BBp strain compared to bacteria from aphids of the subfamily Aphidinae (*B. aphidicola* BAp and BSg), while the smallest value is found between the *B. aphidicola* BAp and BSg strains [15]. However, if the temporary factor is considered, the rates of synonymous nucleotide substitutions per site and million years are greater in the endosymbionts from the Aphidinae (*B. aphidicola* BAp and BSg strains), registering the *B. aphidicola* BCc strain the smallest values. These results demonstrate that the synonymous substitution rate in *B. aphidicola* is a variable character, yet the explanation for

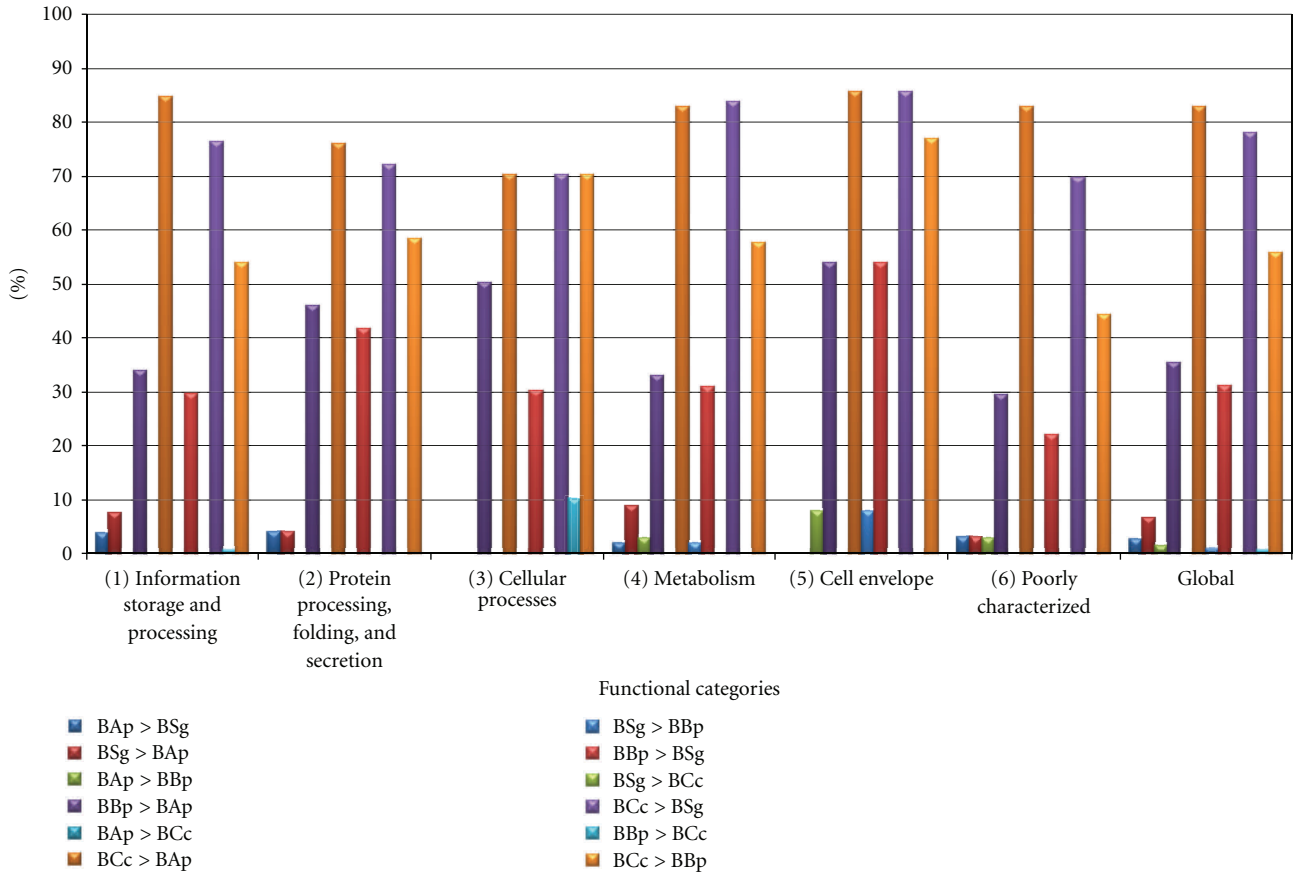


FIGURE 3: Relative distribution of the accelerated genes based on their functional category, between pairs of *B. aphidicola* strains. Accelerated genes were calculated by Tajima’s relative rate tests. A > B indicates a significantly higher accumulation of substitutions in strain A than in strain B ( $P < .05$ ).

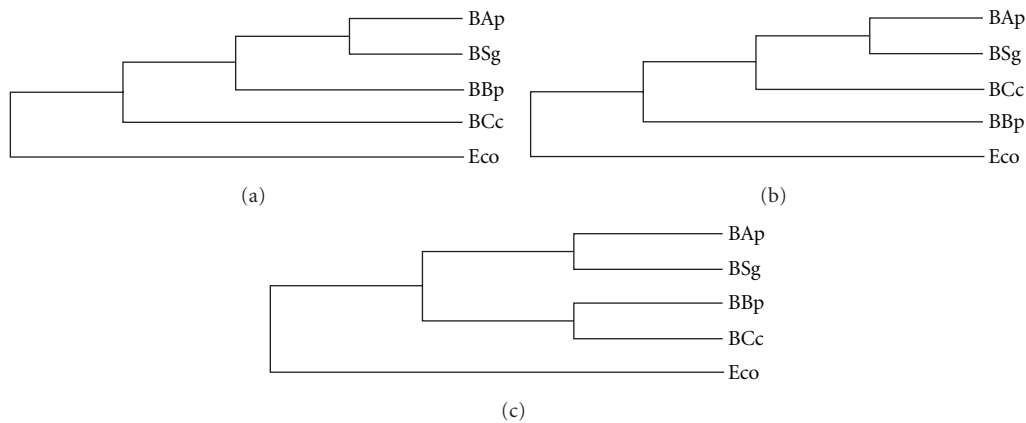


FIGURE 4: Topologies of the phylogenetic trees for the 21 genes that follow the hypothesis of molecular clock [15]. The trees were obtained by maximum likelihood, with the program PAUP 4.0b10.

these divergent patterns is not obvious. As stated elsewhere [7, 14, 44], these differences can be attributed to differences in the host’s life cycle, as well as ecological factors such as host-alternation and variations in the effective population size showed by the two members of the Aphidinae subfamily

compared to the other two aphid lineages. Additionally a differential mutation rate per generation cannot be ruled out. For example, endosymbionts from aphids with short generation times can accumulate more synonymous mutations per million years (case of the Aphidinae) than those

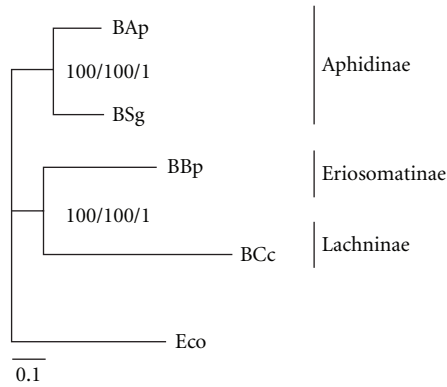


FIGURE 5: Phylogenetic tree obtained by maximum likelihood using PAUP4.0b10 on nucleotide sequences and the GTR+I+G evolutionary model. Topologies obtained from amino acid sequences, using Phym1\_v2.4.5 and MrBayes v3.1.2, are identical. Trees are based on the concatenated sequence of the 338 protein-coding genes shared by the four *B. aphidicola* strains and *E. coli*. Numbers beside the internal nodes are the maximum likelihood bootstrap values from 300 resamplings obtained with PAUP4.0b10, Phym1 and the Bayesian MCMC posterior probability, respectively. The scale bar represents the number of nucleotide substitutions per site.

with longer generation times, such as the Eriosomatinae and the Lachninae. Future studies are required to understand the evolutionary processes driving these patterns.

## Acknowledgments

Financial support was provided by Grant BFU2009-12895-C02-01/BMC (Ministerio de Educación y Ciencia, Spain) to A. Latorre and European Community's Seventh Framework Programme (FP7/2007–2013) under Grant Agreement number 212894 and Prometeo/2009/092 (Conselleria d'Educació, Generalitat Valenciana, Spain) to A. Moya.

## References

- [1] J. Sandstrom and J. Pettersson, "Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrtosiphon pisum*) performance," *Journal of Insect Physiology*, vol. 40, no. 11, pp. 947–955, 1994.
- [2] N. A. Moran, M. A. Munson, P. Baumann, and H. Ishikawa, "A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts," *Proceedings of the Royal Society B*, vol. 253, no. 1337, pp. 167–171, 1993.
- [3] C. D. von Dohlen and N. A. Moran, "Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation," *Biological Journal of the Linnean Society*, vol. 71, no. 4, pp. 689–717, 2000.
- [4] J. J. Wernegreen and N. A. Moran, "Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes," *Molecular Biology and Evolution*, vol. 16, no. 1, pp. 83–97, 1999.
- [5] L. Klasson and S. G. E. Andersson, "Evolution of minimal-gene-sets in host-dependent bacteria," *Trends in Microbiology*, vol. 12, no. 1, pp. 37–43, 2004.
- [6] J. J. Wernegreen, A. O. Richardson, and N. A. Moran, "Parallel acceleration of evolutionary rates in symbiont genes underlying host nutrition," *Molecular Phylogenetics and Evolution*, vol. 19, no. 3, pp. 479–485, 2001.
- [7] T. Itoh, W. Martin, and M. Nei, "Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 20, pp. 12944–12948, 2002.
- [8] J. J. Wernegreen, "Genome evolution in bacterial endosymbionts of insects," *Nature Reviews Genetics*, vol. 3, no. 11, pp. 850–861, 2002.
- [9] A. Mira and N. A. Moran, "Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria," *Microbial Ecology*, vol. 44, no. 2, pp. 137–143, 2002.
- [10] The International Aphid Genomics Consortium, "Genome aequence of the pea aphid *Acyrtosiphon pisum*," *PLoS Biology*, vol. 8, no. 2, Article ID e1000313, 2010.
- [11] O. E. Heie, "Palaeontology and phylogeny," in *Aphids: Their Biology, Natural Enemies and Control*, A. K. Minks and P. Harrewijn, Eds., vol. 2A, pp. 367–391, Elsevier, Amsterdam, The Netherlands, 1987.
- [12] E. Zuckerkandl and L. Pauling, "Molecular disease, evolution, and genetic heterogeneity," in *Horizons in Biochemistry*, M. Kasha and B. Pullman, Eds., pp. 189–225, Academic Press, New York, NY, USA, 1962.
- [13] E. Zuckerkandl and L. Pauling, "Evolutionary divergence and convergence in proteins," in *Evolving Genes and Proteins*, V. Bryson and H. J. Vogel, Eds., pp. 97–166, Academic Press, New York, NY, USA, 1965.
- [14] M. A. Clark, N. A. Moran, and P. Baumann, "Sequence evolution in bacterial endosymbionts having extreme base compositions," *Molecular Biology and Evolution*, vol. 16, no. 11, pp. 1586–1598, 1999.
- [15] V. Pérez-Brocá, R. Gil, S. Ramos et al., "A small microbial genome: the end of a long symbiotic relationship?" *Science*, vol. 314, no. 5797, pp. 312–313, 2006.
- [16] W. Wojciechowski, *Studies on the Systematic System of Aphids (Homoptera, Aphidinea)*, Uniwersytet Slaski, Katowice, Poland, 1992.
- [17] D. Martínez-Torres, C. Buades, A. Latorre, and A. Moya, "Molecular systematics of aphids and their primary endosymbionts," *Molecular Phylogenetics and Evolution*, vol. 20, no. 3, pp. 437–449, 2001.
- [18] B. Ortiz-Rivas, A. Moya, and D. Martínez-Torres, "Molecular systematics of aphids (Homoptera: Aphididae): new insights from the long-wavelength opsin gene," *Molecular Phylogenetics and Evolution*, vol. 30, no. 1, pp. 24–37, 2004.
- [19] B. Ortiz-Rivas and D. Martínez-Torres, "Combination of molecular data support the existence of three main lineages in the phylogeny of aphids (Hemiptera: Aphididae) and the basal position of the subfamily Lachninae," *Molecular Phylogenetics and Evolution*, vol. 55, no. 1, pp. 305–317, 2010.
- [20] S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa, "Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS," *Nature*, vol. 407, no. 6800, pp. 81–86, 2000.
- [21] I. Tamas, L. Klasson, B. Canbäck et al., "50 million years of genomic stasis in endosymbiotic bacteria," *Science*, vol. 296, no. 5577, pp. 2376–2379, 2002.
- [22] R. C. H. J. van Ham, J. Kamerbeek, C. Palacios et al., "Reductive genome evolution in *Buchnera aphidicola*," *Proceedings*



- of the National Academy of Sciences of the United States of America, vol. 100, no. 2, pp. 581–586, 2003.
- [23] K. Tamura, J. Dudley, M. Nei, and S. Kumar, “MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0,” *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1596–1599, 2007.
- [24] R. C. Edgar, “MUSCLE: multiple sequence alignment with high accuracy and high throughput,” *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.
- [25] F. Tajima, “Simple methods for testing the molecular evolutionary clock hypothesis,” *Genetics*, vol. 135, no. 2, pp. 599–607, 1993.
- [26] G. Deckert, P. V. Warren, T. Gaasterland et al., “The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*,” *Nature*, vol. 392, no. 6674, pp. 353–358, 1998.
- [27] R. Gil, F. J. Silva, E. Zientz et al., “The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 16, pp. 9388–9393, 2003.
- [28] N. A. Moran, “Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 7, pp. 2873–2878, 1996.
- [29] X. Xia and Z. Xie, “DAMBE: software package for data analysis in molecular biology and evolution,” *Journal of Heredity*, vol. 92, no. 4, pp. 371–373, 2001.
- [30] A. T. Marques, A. Antunes, P. A. Fernandes, and M. J. Ramos, “Comparative evolutionary genomics of the HADH2 gene encoding  $\text{A}\beta$ -binding alcohol dehydrogenase/17 $\beta$ -hydroxysteroid dehydrogenase type 10 (ABAD/HSD10),” *BMC Genomics*, vol. 7, article 202, 2006.
- [31] M. G. Fain and P. Houde, “Multilocus perspectives on the monophyly and phylogeny of the order Charadriiformes (Aves),” *BMC Evolutionary Biology*, vol. 7, article 35, 2007.
- [32] M. Farfán, D. Miñana-Galbis, M. C. Fusté, and J. G. Lorén, “Divergent evolution and purifying selection of the *flaA* gene sequences in *Aeromonas*,” *Biology Direct*, vol. 4, article 23, 2009.
- [33] M. Daly, L. C. Gusmão, A. J. Reft, and E. Rodríguez, “Phylogenetic signal in mitochondrial and nuclear markers in sea anemones (cnidaria, Actiniaria),” *Integrative and Comparative Biology*, vol. 50, no. 3, pp. 371–388, 2010.
- [34] D. L. Swofford, *PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). Version 4*, Sinauer Associates, Sunderland, Mass, USA, 2002.
- [35] S. Guindon and O. Gascuel, “A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood,” *Systematic Biology*, vol. 52, no. 5, pp. 696–704, 2003.
- [36] D. Posada, “jModelTest: phylogenetic model averaging,” *Molecular Biology and Evolution*, vol. 25, no. 7, pp. 1253–1256, 2008.
- [37] F. Abascal, R. Zardoya, and D. Posada, “ProtTest: selection of best-fit models of protein evolution,” *Bioinformatics*, vol. 21, no. 9, pp. 2104–2105, 2005.
- [38] J. P. Huelsenbeck and F. Ronquist, “MRBAYES: Bayesian inference of phylogenetic trees,” *Bioinformatics*, vol. 17, no. 8, pp. 754–755, 2001.
- [39] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010, <http://www.R-project.org>.
- [40] V. F. Eastop, “Biotypes of aphids,” in *Perspectives in Applied Biology*, A. D. Lowe, Ed., vol. 51 of *Bulletin of the Entomological Society of New Zealand*, pp. 40–51, 1973.
- [41] O. E. Heie, “Aphid ecology in the past and a new view on the evolution of Macrosiphini,” in *Individuals, Populations and Patterns in Ecology*, S. R. Leather, A. D. Watt, N. J. Mills, and K. F. A. Walters, Eds., pp. 409–418, Intercept, Andover, UK, 1994.
- [42] O. E. Heie, “The evolutionary history of aphids and a hypothesis on the coevolution of aphids and plants,” *Bollettino di Zoologia Agraria e di Bachicoltura*, vol. 28, pp. 149–155, 1996.
- [43] L. Gómez-Valero, A. Latorre, and F. J. Silva, “The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*,” *Molecular Biology and Evolution*, vol. 21, no. 11, pp. 2172–2181, 2004.
- [44] H. Ochman, S. Elwyn, and N. A. Moran, “Calibrating bacterial evolution,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 22, pp. 12638–12643, 1999.