

# The Evolution of the Heat-Shock Protein GroEL from *Buchnera*, the Primary Endosymbiont of Aphids, Is Governed by Positive Selection

Mario Ali Fares, Eladio Barrio, Beatriz Sabater-Muñoz, and Andrés Moya

Institut "Cavanilles" de Biodiversitat i Biologia Evolutiva and Department de Genètica, Universitat de València, Spain

The heat-shock protein GroEL is a double-ring-structured chaperonin that assists the folding of many newly synthesized proteins in *Escherichia coli* and the refolding in vitro, with the cochaperonin GroES, of conformationally damaged proteins. This protein is constitutively overexpressed in the primary symbiotic bacteria of many insects, constituting approximately 10% of the total protein in *Buchnera*, the primary endosymbiont of aphids. In the present study, we perform a maximum likelihood (ML) analysis to unveil the selective constraints in GroEL. In addition, we apply a new statistical approach to determine the patterns of evolution in this highly interesting protein. The main conclusion derived from our analysis is that GroEL has suffered an accelerated rate of amino acid substitution upon the symbiotic integration of *Buchnera* into the aphids. It is most interesting that the ML analysis of codon substitutions in the different branches of the phylogenetic tree strongly supports the action of positive selection in the different lineages of *Buchnera*. Additionally, the new sliding window analysis of the complete *groEL* sequence reveals different regions of the molecule under the action of positive selection, mainly located in the apical domain, that are important for both peptide and GroES binding.

## Introduction

The bacterial chaperonin GroEL and its cofactor GroES are two proteins essential for cell viability under normal growth conditions (Hemmingsen et al. 1988; Zeilstra-Ryalls, Fayet, and Georgopoulos 1991) and at all temperatures for the growth of *Escherichia coli* (Fayet, Ziegelhoffer, and Georgopoulos 1989). These two proteins are encoded in the same operon *groE* and are involved in the binding of nonnative polypeptides by hydrophobic contacts (Cheng et al. 1989; Fayet, Ziegelhoffer, and Georgopoulos 1989; Braig et al. 1994), assisting in the folding of newly synthesized proteins and the refolding of conformationally damaged proteins (Gething and Sambrook 1992; Hartl 1996; Fenton and Horwich 1997; Ranson, White, and Saibil 1998; Sigler et al. 1998; Xu and Sigler 1998). Thus, the importance of GroEL resides in its ability to prevent inappropriate intra- and intermolecular interactions, that would otherwise lead to irreversible, nonspecific aggregations (Gething and Sambrook 1992; Georgopoulos and Welch 1993; Ellis 1994), and to switch from a folding to a storing function (Llorca et al. 1998).

The GroEL protein is a homotetradecamer structured in two rings, each consisting of three structural domains (Braig et al. 1994; Braig, Adams, and Brünger 1995). The large equatorial (residues 5–132 and 408–522) and apical (residues 190–375) domains are located at the center and the ends of the tetradecameric complex, respectively, linked by the third smaller intermediate domain (residues 133–189 and 376–407). According to different detailed studies (Hartl 1996; Weissman et al.

1996; Fenton and Horwich 1997; Rye et al. 1997), the binding of the peptide-folding intermediates takes place in a hydrophobic groove in the apical domain, where the cofactor GroES also binds (Landry et al. 1993). The equatorial domain, which is involved in all the inter-ring and most of the intraring interactions, contains the ATP-binding sites that are essential for the GroEL-GroES-polypeptide-binding cycle (Braig et al. 1994). These three domains do not function independently but have been demonstrated, by means of mutation experiments, to be functionally connected (Kawata et al. 1999).

The GroEL-homologue protein (Hara et al. 1990) is abundantly synthesized in several endosymbiotic bacteria studied so far (Ishikawa 1982, 1984; Ahn et al. 1994; Aksoy 1995; Baumann P, Baumann L, and Clark 1996; Charles et al. 1997). GroEL has been reported to be highly expressed in bacteriocyte-harbored endosymbiotic bacteria in aphids (Ishikawa 1982, 1984; Hara et al. 1990), that are vertically transmitted from one generation to the next at an early stage of the host embryogenesis (Buchner 1965, pp. 210–332). In fact, GroEL corresponds to ≈10% of the total protein produced in *Buchnera* (Hara et al. 1990; Baumann, Baumann, and Clark 1996). Several studies pointed out that GroEL is secreted from *Buchnera* to the outer space of the bacteriocyte and is able to join viral particles in the hemolymph of the aphid and transmit them between plants (Young and Filichkin 1999; Hogenhout et al. 2000; Li et al. 2001). Two regions of distinct variability distinguish the GroEL of *Buchnera* from that belonging to their closest free-living relative, the *E. coli*: the hypervariable serine-rich tract (residues 526–538), located at the C-terminus of GroEL, and residues 339–347, belonging to the apical domain of this protein, that are involved in substrate polypeptide binding (Fenton et al. 1994). Precise complementation experiments with the GroE mutants of *E. coli* showed that both the *groES* and *groEL* genes from the endosymbiont codify for the functional molecular chaperones in *E. coli* (Ohtaka, Nakamura, and Ishikawa 1992).

The strong bottleneck suffered by endosymbiotic bacteria of aphids during their clonal transmission to the

Abbreviations: PCR, polymerase chain reaction; LRT, likelihood ratio test.

Key words: aphid endosymbionts, *Buchnera aphidicola*, *groEL*, positive selection, rates of evolution.

Address for correspondence and reprints: Andrés Moya, Institut "Cavanilles" de Biodiversitat i Biologia Evolutiva, Universitat de València, Edifici d'Instituts del Campus de Paterna, P.O. Box 2085, E-46071 València, Spain. E-mail: andres.moya@uv.es.

*Mol. Biol. Evol.* 19(7):1162–1170. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

ovaries or during the infection of developing embryos (Buchner 1965; Hinde 1971; Moran, von Dohlen, and Baumann 1995; Baumann et al. 1995) leads to a strong reduction in their effective population sizes (Funk, Wernegreen, and Moran 2001). The expected effect of this bottleneck is the accumulation of slightly deleterious mutations by genetic drift (Ohta 1973), which might reflect on the acceleration of the fixation rate of nucleotide substitutions at nonsynonymous sites in symbiotic bacteria compared with their free-living relatives (Moran 1996; Wernegreen and Moran 1999). The difference in the rate of evolution between free-living bacteria and *Buchnera* was underlined in several studies (Moran, Von Dohlen, and Baumann 1995; Moran 1996; Brynne et al. 1998). Although all loci examined in *Buchnera* (Wernegreen and Moran 1999) have values of nonsynonymous to synonymous rate ratio ( $K_a/K_s$ ) significantly higher than that in *E. coli*, *groEL* showed the lowest nonsynonymous to synonymous rate ratio ( $K_a/K_s$  or  $d_N/d_S$ ) that was apparently caused by a more effective action of purifying selection on this gene.

In this study our goal is to demonstrate the role played by positive selection in the evolution of *groEL* in *Buchnera* and to highlight the specific amino acid regions of this gene that were fixed by positive selection. The assertion of this work is that *groEL* is an extremely important gene for ensuring the normal function of the endosymbiotic protein system and that only a few beneficial amino acid replacements can be fixed.

## Materials and Methods

### Sequence Data

Four free-living  $\gamma$ -proteobacteria were used as the closest relatives to *Buchnera*. These bacteria and their accession numbers are as follows: *Escherichia coli* (X07850), *Erwinia carotovora* (AB008152), *Salmonella typhimurium* (U01039), and *Enterobacter aerogenes* (AB008141). Ten *Buchnera* sequences from different aphid species were used; five of them were obtained from the EMBL database: *Rhopalosiphum padi* (U77380), *Schizaphis graminum* (AF008210), *Acyrtosiphon pisum* (X61150), *Sitobion avenae* (U77379), and *Myzus persicae* (AF003957), all belonging to the Aphidinae subfamily. The other five were sequenced for this study and belong to *Buchnera* of aphids whose species (and corresponding subfamilies) and accession numbers are as follows: *Chaitophorus leucomelas* (Chaitophorinae), AJ439087; *Pterocomma populeum* (Pterocommatinae), AJ439083; *Tetraneura caerulea* (Pemphiginae), AJ439084; *Thelaxes suberi* (Thelaxinae), AJ439085; and *Tuberolachnus salignus* (Lachninae), AJ439086.

### Aphid Collection and DNA Preparation

Aphids were field collected and stored at 4°C until further use. Several individuals from the same aphid colony were stored in 70% ethanol for species identification, and the remaining were used for DNA isolation.

Isolation of total aphid DNA from single individuals was carried out following the method described in Latorre, Moya, and Ayala (1986).

### PCR and Sequencing of the *groE* Operon

Two degenerated primers were designed based on the alignment of the *groE* operon sequence from *E. coli*, *Sitophilus oryzae* endosymbiont, *Lactobacillus lactis*, *Buchnera* (*S. graminum*), *Buchnera* (*A. pisum*), *Buchnera* (*M. persicae*), and *Buchnera* (*R. padi*): *groExF15'*-(ggaattc)ATGAAWATTCGTCCRTTRCAYGAYCG-3' and *groExR1* 5'-(ggaattc)TTACATCATKCCRCCATRCCACCCA-3' (van Ham et al. 2000). The oligo 5' sequences between brackets are *EcoRI* extensions.

PCR reactions were performed in the GeneAmp 2400 or 9700 System (Perkin-Elmer). Cycling conditions were as follows: 92°C for 2 min; 30 cycles of 92°C for 15 s, 52°C for 30 s, 72°C for 2 min; and a final extension of 4 min at 72°C. PCR conditions were as follows: 200 mM dNTPs (AP Biotech), 20 pmol of each primer, 1 × *Taq* buffer, 1.5 U of *Taq* pol (AP Biotech), and 10–40 ng of template. PCR products (~1.9 kb in length) were excised from a 0.8% agarose gel in 0.5 × TBE and extracted with QIAEX (QIAGEN). PCR products were cloned into pGEM-T Easy (Promega). Plasmid extractions were carried out using CONCERT<sup>®</sup> Rapid miniprep system (GIBCO-BRL). Sequencing reactions were performed in Applied Biosystems automatic sequencers/ABI 373, 377, or 3700 using d-Rhodamine ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer). In order to avoid *Taq*-mistakes, several clones were sequenced. T7 and SP6 vector primers and internal primers (sequence available upon request) were used to cover the complete length of the PCR product in both the chains.

### Phylogenetic Reconstruction and Nucleotide Substitution Models

Alignment of sequences was performed with CLUSTAL X, a Windows version of the CLUSTAL W program (Thompson, Higgins, and Gibson 1994) and corrected by eye. Phylogenetic analysis of the aligned sequences was performed using different methods: Neighbor-Joining (NJ) (Saitou and Nei 1987), maximum likelihood (ML) (Felsenstein 1981), and maximum parsimony (MP) (Fitch 1971). The MEGA program, version 2.01 (Kumar, Tamura, and Nei 1993) was used to obtain the NJ trees and the bootstrap support values for 1,000 pseudoreplicates. MP and ML trees were reconstructed using DNPARS and DNAML, respectively, from the PHYLIP package, version 3.5 for Windows (Felsenstein 1993).

For distance estimation and ML analysis, we determined the appropriate model to explain the evolution of *groES/L* sequences by using the likelihood ratio test (LRT) (Huelsenbeck and Crandall 1997). When two models are nested, they can be compared by the LRT; twice the log-likelihood difference follows a chi-square distribution, with the degrees of freedom being the difference between the numbers of free parameters between

the models compared. These tests were implemented using the program BASEML from the PAML package, version 3.0 (Yang 2000). The models tested for a given phylogenetic tree were those described by Jukes and Cantor (1969), Kimura (1980), Felsenstein (1981), Hasegawa, Kishino, and Yano (1985), Tamura and Nei (1993), and Yang (1994). In addition, to determine if the substitution rates are equal (Poisson distribution) or unequal ( $\Gamma$ -distributed rates model) among sites, the LRT was also applied.

#### Assessing Substitution Rates in Branches Leading to the Different Bacterial Groups

To assess the variation in the substitution rates among different lineages of the phylogenetic tree, we first measured nucleotide and amino acid distances in the different branches of the tree and tested the constancy of the substitution rates, using as an outgroup the *groES/L* sequence of the gamma proteobacteria *Pseudomonas aeruginosa*. Nucleotide distances were estimated using Tamura and Nei's model (1993) under a gamma distribution of substitution rates (Rzhetsky and Nei 1994) with a shape parameter ( $\alpha$ ) of 0.2997. This  $\alpha$  value was estimated with the PAMP and BASEML programs from the PAML package (Yang 2000). Accurate amino acid distances were obtained with the Poisson correction. Thereafter, rate distances were assessed using the two-cluster test (Takezaki, Rzhetsky, and Nei 1995; program LINTREE), which examines the equality of the average substitution rates for two clusters (A and B) linked by a node on the tree, using one or several out-group sequences.

#### Detection of Branches of the Phylogenetic Tree Under Positive Selection

To test the selective pressures in the different lineages of the phylogenetic tree, we estimated the non-synonymous to synonymous rate ratio ( $\omega = d_N/d_S$ ). Values of  $\omega = 1$ ,  $\omega > 1$ , and  $\omega < 1$  indicate neutrality, positive selection, and purifying selection, respectively. This is the most accepted and stringent way to detect positively selected changes in a sequence alignment (Sharp 1997; Akashi 1999; Crandall et al. 1999). To perform these analyses, we applied different models of codon evolution implemented in the CODEML program of the PAML package (Yang 2000). The Goldman and Yang (1994) model assumes a single  $\omega$  value for all the lineages of the phylogenetic tree and for all codon sites of the sequence alignment. In contrast, the free-ratio model allows the  $\omega$  value to vary among different lineages of the tree. Both the models are nested and hence comparable by the LRT because it can be approached to a chi-square distribution with the degrees of freedom being the total number of estimated rate ratios ( $\omega$ ) - 1.

#### Highlighting Specific Regions Under Selective Constraints

We have developed a new statistical method, based on a sliding-window approach, to detect selective con-

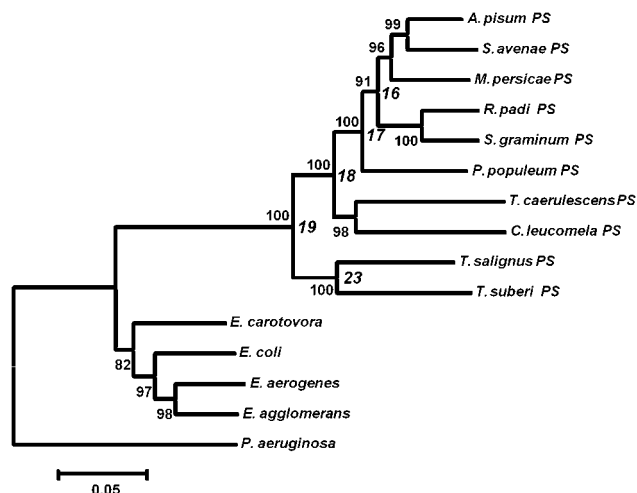


FIG. 1.—Neighbor-Joining tree based on nucleotide distances between bacterial *groEL* sequences corrected by the Tamura and Nei's model (1993) assuming a gamma distribution of nucleotide substitution rates with a shape parameter of  $\alpha = 0.2972$ . Numbers on the nodes correspond to bootstrap values based on 1,000 pseudoreplicates. Italic numbers are used to identify nodes of interest in the analysis of positive selection (see table 3).

straints in protein-coding genes (M. A. Fares et al., unpublished data). In brief, the method tests the significance of the deviation of nonsynonymous and synonymous substitutions from that expected under neutrality and allows the testing of different hypotheses, such as saturation of synonymous or nonsynonymous sites, high substitution rates, or the action of selection in a specific region of the protein, that could be important from a structural or functional viewpoint. The fundamentals of the method are similar to those of Suzuki and Gojobori (1999), but differ from it in that our method uses a Poisson approach instead of a binomial and enables us to use a statistically appropriate window size to analyze the data. The method is described in the Supplementary Material posted on the MBE web page (<http://www.molbiolvol.org>).

## Results

### Accelerated Substitution Rates in the Lineages Leading to the Primary Symbiotic Bacteria of Aphids

All the three phylogenetic tree reconstruction methods (NJ, MP, and ML) applied to *groES/L* data gave the same tree topology (fig. 1). The LRT revealed that irrespective of the models compared, a gamma distribution of the nucleotide substitution rates significantly improved the log-likelihood values, indicating that, among the nucleotide substitution models, those considering heterogeneous rates of nucleotide substitutions must be used. A comparison of the models, that took into account a gamma distribution of nucleotide substitutions using the LRT showed that the best nucleotide substitution model to explain the evolution of the *groE* alignment, given the phylogenetic tree of figure 1, is the Tamura and Nei model (1993). In fact, the comparison of the models of Tamura and Nei and Hasegawa-Kishino and Yano by the LRT gives a value of  $2\Delta l = 114.626$



**Table 1**  
**Two-Cluster Test of Nucleotide or Amino Acid Substitution Rate Constancy**

Data	Cluster A	Cluster B	$b_A$	$b_B$	$\delta$	Z	P
Nucleotides . . .	Free-bacteria	Symbionts	0.025	0.342	0.317	12.628	<0.001
	Aphidinae symbionts	<i>Cle-Tca</i>	0.076	0.124	0.048	5.919	<0.001
Amino acids . . . . .	Free-bacteria	Symbionts	0.057	0.101	0.044	4.205	<0.001
	Rest of symbionts	<i>Tsa-Tsu</i>	0.041	0.066	0.025	4.507	<0.001
	Aphidinae symbionts	<i>Cle-Tca</i>	0.041	0.063	0.022	3.785	<0.001

NOTE.— $b_A$  and  $b_B$  is the distance between the root and cluster A and B, respectively.  $\delta$  is the absolute difference between  $b_A$  and  $b_B$ . Specific endosymbionts are indicated by their host abbreviation as follows: *Cle*, *C. leucomelas*; *Tca*, *T. caerulescens*; *Tsa*, *T. salignus*; and *Tsu*, *T. suberi*.

( $P < 0.001$ ) with an ML-estimated gamma shape parameter  $\alpha = 0.2972$ .

The application of the two-cluster test of evolutionary rates, using *P. aeruginosa* as the most adequate out-group, showed an acceleration of the nucleotide substitution rates in the lineage leading to the endosymbiotic bacteria (table 1). In this table only significant rate comparisons are shown. Thus, only two clusters seem to be accelerated; the first includes the complete endosymbiotic bacterial group, where the rate for this group is significantly higher than that for the free-living bacteria ( $P < 0.001$ ), and the second corresponds to the group including *T. salignus* and *T. suberi* endosymbionts ( $P < 0.001$ ) compared with the rest of the endosymbionts. The application of the same test to the Poisson-corrected distances of amino acid sequences also showed an acceleration of the amino acid substitutions in the branches leading to the same groups, all with significant values ( $P < 0.001$ ).

#### Positive Selection as the Main Force Responsible for the Fixation of Amino Acid Replacements in *Buchnera* GroEL

A comparison of the log-likelihood values of the one-ratio and free-ratio models, to test which model better explains the evolution of *Buchnera groEL*, significantly indicated that the free-ratio model is the most appropriate ( $\chi^2 \approx 2\Delta\ell = 185.965$ ;  $P < 0.001$ ) (table 2). Under the free-ratio model,  $\omega$  values greater than 1 were detected in the branch connecting the cluster constituted by the symbionts of *T. salignus* and *T. suberi* ( $\omega = 10.5602$ ).

**Table 2**  
**Likelihood Ratio Test to Compare Models of Codon Evolution Considering Variable  $d_N/d_S$  Ratios Along Phylogenetic Lineages or a Single Ratio**

GroEL region	$\ell$	$\ell$	$2\Delta\ell$	P
	(free-ratio model)	(one-ratio model)		
Complete . . . . .	-9544.159	-9637.141	185.964	<0.01
Apical . . . . .	-3340.015	-3743.4808	802.931	<0.01
Intermediate . . .	-1573.601	-1597.456	47.710	<0.01
Equatorial . . . .	-4709.947	-5126.3182	832.000	<0.01

NOTE.—Models compared were the free-ratio and the one-ratio models implemented in PAML (Yang 2000).  $\ell$  corresponds to the log-likelihood value, twice the difference between  $\ell$  values can be compared with chi-square distribution. P is the probability of each  $2\Delta\ell$  value.

To determine if functionally and structurally specific regions of *groEL* show different  $\omega$  values, and to unveil which region (apical, equatorial, or intermediate) is responsible for these high  $\omega$  values, we performed the same analysis for each region. For all three regions, the free-ratio model always gave significantly higher log-likelihood values than did the one-ratio model (table 2). However, only the apical domain gave  $\omega$  values significantly higher than 1. In fact, when the free-ratio model was used to analyze the apical domain, three branches were detected to be under strong positive selection. These branches were those connecting *T. salignus* and *T. suberi* ( $\omega = 1.7$ ), the branch connecting the symbionts from species of the Aphididae family ( $\omega = 23.8$ ), and one of the branches within the *Buchnera* from the Aphididae family ( $\omega = 41.12$ ) (fig. 1). Amino acid replacements in the apical domain for these three branches detected to be under positive selection, with their posterior Bayesian probabilities, are shown in table 3. The 11 amino acid replacements were located close to important peptide-binding sites, and 5 of them correspond to dramatic changes in the amino acid nature (position 336 G-D, 340 E-N, 344 H-N, 346 Q-K, and 355 E-Q).

To ensure that there was no saturation of synonymous sites caused by multiple hits that could give a  $d_N/d_S$  ratio greater than 1, we plotted the synonymous ver-

**Table 3**  
**Amino Acid Replacements Subjected to Positive Selection in the Apical Domain Occurred in the Different Branches of the Phylogenetic Tree Depicted in Figure 1**

Branch Between Nodes	Codon Position	Replacement	P
16-17 . . . . .	231	V-I	0.98
	302	S-T	0.99
	336	G-D	0.99
17-18 . . . . .	344	H-N	0.94
	350	S-T	0.55
	355	E-Q	0.96
23-19 . . . . .	310	M-L	0.98
	315	A-T	0.97
	338	V-I	0.95
	340	E-N	0.89
	346	Q-K	0.95

NOTE.—Ancestral nodes are the internal nodes connecting two clusters and are drawn in the tree of figure 1. Codon positions of the amino acid replacements and their posterior Bayesian probability are also shown.

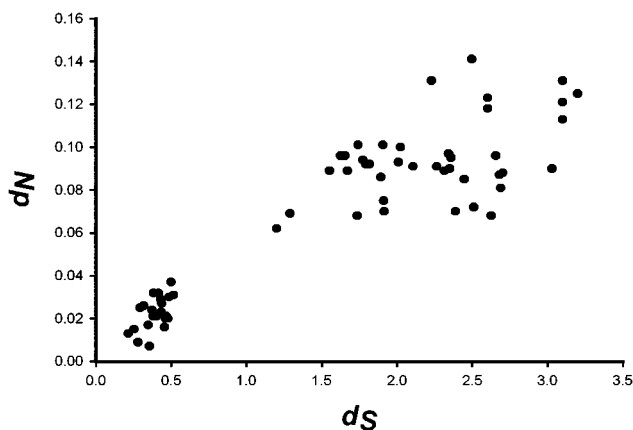


FIG. 2.—Plot of synonymous ( $d_S$ ) versus nonsynonymous ( $d_N$ ) pairwise distances estimated by Nei and Gojobori's modified method.

synonymous distances, both estimated by Nei and Gojobori's modified method (Zhang, Rosenberg, and Nei 1998), and fitted linear and logarithmic regression models to the data (fig. 2). Saturation of synonymous sites might lead to a better fit to the logarithmic model compared with the linear one, whereas no significant differences are expected if there is no saturation of synonymous nucleotide sites. When the linear model was applied, the correlation coefficient was  $r = 0.9066$  ( $P < 0.001$ ), and the logarithmic model gave a correlation coefficient of  $r = 0.9267$  ( $P < 0.001$ ). However, the logarithmic model did not improve the linear regression model ( $t = 0.266$ ;  $P = 0.4173$ ), which means that the linear model is a good explanation of the correlation of synonymous distances with respect to nonsynonymous ones.

When the same approach was applied for each bacterial group (free-living bacteria and *Buchnera* endosymbionts), we obtained the same results. Thus, for free-living bacteria, the correlation coefficient under the linear model was  $r = 0.9629$  ( $P < 0.001$ ), being the difference between the sum of squares not significant between the linear and logarithmic models ( $t = 1$ ,  $P = 0.2500$ ). Finally, the coefficient of correlation under the linear model for *Buchnera* was  $r = 0.7086$  ( $P < 0.001$ ), and the comparison of the two models again was not significant ( $t = 0.0530$ ;  $P = 0.4836$ ).

#### Detecting Regions Under Positive Selection by the Sliding Window Method

First of all we calculated the probability of nonsynonymous and synonymous nucleotide substitutions following equation (3) (see Supplementary Material) and using a random set of simulated sequence alignments. The estimated probabilities of nonsynonymous and synonymous substitutions,  $P(\theta_N) = 0.1$  and  $P(\theta_S) = 0.9$ , were used to determine the appropriate window size to analyze the sequence data, which resulted in 4 codons because its lower 5%  $P(d_N)$  estimated value was higher than 0.05 (fig. 3). Therefore, we might not expect significant results by chance if we slid a window of 4 codons along the real sequence alignment and if no selec-

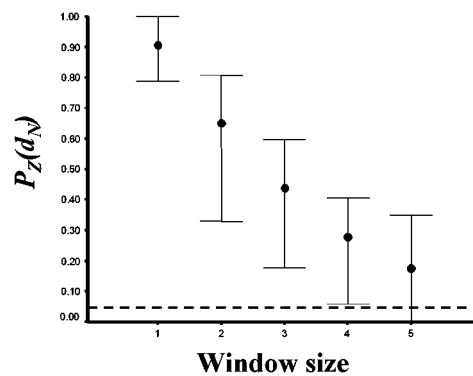


FIG. 3.—Representation of the Poisson probability of nonsynonymous nucleotide substitutions in each window size. Bold dots are the mean probability in each window, upper and lower limits are the 5% higher and lower probability estimates, respectively, and the horizontal dashed line indicates the threshold probability of 0.05.

tive constraints are affecting the evolution of any region of the sequence.

When a window of four codons was slid along the real alignment, we detected several regions under positive selection; these regions, their probabilities of nonsynonymous substitutions, and their  $\omega$  values are represented in table 4. In this table, we represent the regions wherein positively selected codons were detected in endosymbiotic bacteria. By this method, nine regions were determined to be under positive selection, five of which corresponded to the apical domain of the protein.

The positively selected codon regions located in the equatorial domain show average  $\omega$  values ranging between 1.94 and 33.94. The intermediate region contains a single region under positive selection, located between codons 160 and 163, having an  $\omega$  value of 4.28. Finally, the apical domain contains five codon regions under positive selection with mean values of  $\omega$  ranging between 3.13 and 7.21.

The significance of these  $\omega$  values was tested against chance by calculating the analytical variance of  $\omega$  following the method of Weir (1996). The variance of  $\omega$  was estimated to be  $V(\omega) = 0.3304$ , and the probabilities of  $\omega$  values in each region of the *groEL* alignment sequence applying a normal test are summarized in table 4.

**Table 4**  
Codon Intervals Subjected to Positive Selection in Each Structural Region of GroEL

GroEL Region	Codon Interval	$\omega$	One-tailed $P$
Apical . . . . .	246–249	3.53	<0.0001
	339–342	7.2	<0.0001
	344–347	3.52	<0.0001
	348–351	3.54	<0.0001
	350–353	3.578	<0.0001
Intermediate . . . . .	160–163	4.278	<0.0001
Equatorial . . . . .	434–437	3.877	<0.0001
	534–537	6.249	<0.0001
	543–546	33.94	<0.0001

NOTE.— $\omega$  values are given for each codon interval with its probability estimated under a normal distribution.

On the other hand, we estimated the proportion of sites that are under different selective constraints. To do so, we distinguished between three sets of codons: those that are under neutrality (where  $\omega$  is not significantly different from 1 or from the average estimate of  $\omega$ ), those that are under strong purifying selection (with  $\omega$  values significantly smaller than 1 and very close to 0), and those that evolved under positive selection (with  $\omega$  values significantly higher than the average  $\omega$  values of the alignment and higher than 1). Only 0.67% of codon sites showed  $\omega$  values not significantly different from 1, whereas the vast majority of the codon sites (97.73%) have  $\omega$  values significantly smaller than 1 (ranging between 0.004 and 0.25), which means that GroEL evolved mainly under strong purifying selection. Among the codons analyzed, a small fraction of them (1.6%) showed  $\omega$  values significantly higher than 1 (ranging from 3.53 to 33.94) and than the average  $\omega$  value estimated from the alignment by the sliding window-based method.

## Discussion

In the light of the nearly neutral theory, effective population size ( $N_e$ ) and selective constraints are the two main factors that control the rate of fixation of substitutions in the population. The effective population size, the mode of transmission, and the probability of recombination among bacteria confirm the main parameters that determine the strength of selection acting on the genome of *Buchnera*. In fact, the slower the recombination rate and smaller the effective population size, the higher is the proportion of mutations that drift to fixation and, hence, the weaker is the effect of selection acting against these mutations. The consequence of this higher genetic drift is an increased fixation of slightly deleterious mutations (Ohta 1973, 1992).

In contrast to free-living bacteria, which have large effective population sizes (Selander, Caugant, and Whitam 1987) and whose evolutionary history is highly controlled by recombination between individuals (Maynard Smith, Dowson, and Spratt 1991; Maynard Smith et al. 1993; Dykhuizen and Green 1993), strictly intracellular symbiotic bacteria have very small effective population sizes, apparently lack recombination, and are maternally transmitted in a clonal manner (Funk, Wernegreen, and Moran 2001). All these characteristics make endosymbiotic bacteria an ideal model for studying the effect of genetic drift in the accumulation of mutations on their genomes and to determine how this accumulation is biased to the fixation of deleterious mutations because of their asexuality (Muller 1964; Lynch 1996, 1997).

Several previous studies showed an accelerated fixation rate of amino acid replacements in *Buchnera aphidicola* genome, the primary symbiont of the aphids (Moran 1996; Brynel et al. 1998; Lambert and Moran 1998; Clark, Moran, and Baumann 1999; Wernegreen and Moran 1999). In agreement with the data obtained by Wernegreen and Moran (1999), the relative rate test evidences 1.3 to 6.9 times higher rates of evolution in *Buchnera* genes compared with those from *E. coli*.

When the authors examined the rates of nonsynonymous to synonymous nucleotide substitutions ( $K_a/K_s$  or the corrected  $\omega = d_N/d_S$ ) in different genes, they realized that, in a vast majority of cases, the  $K_a/K_s$  of *Buchnera* genes significantly exceeded that obtained when the out-group sequence *S. typhimurium* was compared with one of the closest free-living relatives of *Buchnera*, *E. coli*. Because the distribution of amino acid replacements was random along the entire protein-coding gene, they asserted that one cannot expect that these changes have been fixed by positive selection, and hence, the best explanation is that amino acid changes in the lineage of *Buchnera* were slightly deleterious mutations fixed by genetic drift. This accumulation of slightly deleterious mutations has previously been attributed to mutational bias, not buffered because of the absence of recombination, and to small population sizes (Moran 1996; Wernegreen and Moran 1999; Funk, Wernegreen, and Moran 2001). However, an interesting result was that the difference between the two rate ratios was slightly smaller when the heat-shock protein GroEL was examined; thus, it can be concluded that compared with other genes in *Buchnera*, purifying selection is apparently more effective against nonsynonymous substitutions in this gene. Our results demonstrate that *groEL* is subjected to a strong purifying selection (97.73% of the codon sites have  $\omega < 1$  and close to 0), and only very few codons (0.67%) have  $\omega$  values not significantly different than 1, which means that *groEL* is not allowed to accumulate nonsynonymous substitutions even under the strong bottleneck to which *Buchnera* is subjected. This result suggests that, in contrast to the rest of the protein-coding genes that accumulated nonsynonymous changes because of the effect of genetic drift, *groEL* cannot accumulate slightly deleterious amino acid replacements by genetic drift because this protein may be functionally important to buffer the conformation loss of the rest of the proteins and any amino acid replacement could affect its function, and hence, disable *groEL* from folding the damaged proteins.

Our results suggest that an acceleration of nonsynonymous (amino acid) substitution rates occurred during the evolution of the symbiotic lineage but also showed acceleration among lineages within the cluster of *Buchnera*. This acceleration of rates affected nonsynonymous sites in *groEL*, which is consistent with the results obtained by Moran (1996).

However, ML codon-based models applied in this study revealed that, although purifying selection explains the evolution of *groEL* in free-living bacteria and within each aphid family, positive selection is the alternative explanation for the fixation of amino acid replacements in some branches of the tree connecting the aphid endosymbionts. Thus, ML analyses revealed that positive selection is the most likely explanation for the fixation of the amino acid changes in the lineage leading to *T. suberi* and *T. caerulescens* aphid symbionts and the branch connecting *Aphididae* symbionts. These results agree partially with the results obtained by the analysis of the constancy of rates using the two-cluster test. In this way, accelerated amino acid changes were

detected in the same branch leading to the cluster of *T. salignus*-*T. suberi* symbionts. This amino acid replacement acceleration is compatible with the fixation of amino acid changes by positive selection as indicated by the ML analyses. Furthermore, although no accelerated rates were evidenced in the branch leading to the symbionts of the *Aphididae* family, fixation of amino acid replacements by positive selection was also detected in this lineage. These results agree with the idea that, on average, primary symbiotic bacteria of aphids have accumulated amino acid replacements mainly by their fixation because of genetic drift. However, the fixation of amino acid replacements in GroEL in the lineages leading to endosymbionts cannot be explained under the genetic drift hypothesis but by positive selection.

In agreement with these results, the application of the same analyses to the different *groEL* regions, coding for different functional and structural regions of the GroEL protein, consistently detected positive selection in the same branches as in the case of the complete *groEL* alignment. In addition, the analysis of the equatorial and intermediate domains did not show any branch of the phylogenetic tree to be under positive selection.

The main result obtained when ML codon-based models were applied to detect lineages with positively selected amino acid changes is that all the amino acids are located within GroEL regions involved in peptide and GroES binding. Moreover, five out of the 11 replacements detected in these branches constitute important changes in the nature of the amino acids. Three of them (E<sup>340</sup>, H<sup>344</sup> and Q<sup>346</sup>) are located within the domain 339–347 of the apical region involved in substrate polypeptide binding (Fenton et al. 1994). Site-directed mutation experiments demonstrated that changes in these amino acid positions resulted in the loss of GroEL function. Also, position 191 (E–K) was detected as fixed in the *Buchnera* lineage by positive selection. This amino acid is conformationally important because it is located next to residue G<sup>192</sup> that controls the block massive movements of the GroEL apical domain (Braig et al. 1994; Xu, Horwich, and Sigler 1997).

From these results we can conclude that strong purifying selection prevents deleterious amino acid fixation in *groE* by genetic drift even during the strong bottlenecks caused by the maternal transmission of the symbiont to the next generation and the absence of recombination. Most of the amino acid replacements observed in the endosymbiotic *groEL* were fixed by positive selection as indicated by our results. A second conclusion derived from our analysis is that the majority of the amino acid sites are extremely conservative (97.73% of codons are under a strong purifying selection), which indicates that any nonsynonymous substitution in *groEL* has a strong antagonistic effect on its chaperonin function; hence, it can lead to a loss of in cell viability.

An objection to the use of ML methods based on a phylogenetic tree (Yang 1999) to detect positive selected branches is that these methods rely on the assumption that all amino acid sites are under the same selective constraints, and hence, have the same rate of amino acid fixation. Far from realistic, this assumption

does not collect the information to analyze selective constraints in a specific region of the sequence alignment.

Our new method allowed us to answer if the amino acid replacement acceleration observed is mainly caused by the action of positive selection, or, alternatively, is randomly distributed along the sequence and can be explained by chance (genetic drift). According to our results, the nonsynonymous substitutions causing amino acid replacements were not randomly distributed along the protein-coding gene *groEL* but occurred in key positions for the peptide- and GroES-binding function. All the regions under positive selection showed higher amino acid replacements than that expected by chance, and their  $\omega$  values were always significantly higher than the expected value under neutrality. However, in contrast to the results obtained using the ML method, we also detected with the new method several *groEL* regions coding for parts of the equatorial domain and the intermediate domain as subjected to positive selection. When these regions were examined, we realized that they are located in the hypervariable region consisting in a serine-rich tract.

The main conclusion from our study is that the heat-shock protein (chaperonin) GroEL appears to have been subjected to different selective constraints along the evolution of the symbiotic lineages. Our results indicate that, with the exception of a very small proportion of amino acid replacements in the lineage leading to primary symbiotic bacteria of aphids that were fixed by genetic drift, a vast majority of amino acid substitutions were fixed by positive selection in the lineages leading to the symbionts of the *Aphididae* family and to the *T. salignus* and *T. suberi* symbiont cluster.

These results, together with the fact that GroEL has been reported as having the highest expression levels in *Buchnera*, suggest a very important role of GroEL in the functional maintenance of the endosymbiont proteome. In accordance with the hypothesis postulated by Moran (1996), a plausible explanation for the different pattern of protein expression and evolution of GroEL is that this protein may act by buffering the effect of the accumulation of mildly deleterious amino acid replacements in the symbiont proteins caused by genetic drift, during the strong bottlenecks suffered by these symbionts, maintaining the appropriate functional protein folding.

## Acknowledgments

This work was supported by grant BFM 2000-1383 from the Spanish MEC. M.A.F. acknowledges a fellowship from Conselleria de Cultura, Educació i Ciència, Generalitat Valenciana.

## LITERATURE CITED

- AHN, T. I., S. T. LIM, H. K. LEEU, J. E. LEE, and K. W. JEON. 1994. A novel strong promoter of the *groEx* operon of symbiotic bacteria in *Amoeba proteus*. *Gene* **148**:43–49.



- AKASHI, H. 1999. Within- and between-species DNA sequence variation and the "footprint" of natural selection. *Gene* **238**:39–51.
- AKSOY, S. 1995. Molecular analysis of the endosymbiont of tsetse flies: 16S rDNA and overexpression of a chaperonin. *Insect Mol. Biol.* **4**:29–32.
- BAUMANN, P., L. BAUMANN, and M. CLARK. 1996. Levels of *Buchnera aphidicola* chaperonin groEL during growth of the aphid *Schizaphys graminum*. *Curr. Microbiol.* **32**:279–285.
- BAUMANN, P., L. BAUMANN, C. LAI, D. ROUHBAKHSH, N. MORAN, and M. CLARK. 1995. Genetic, physiology and evolutionary relationships of the genus *Buchnera*: intracellular symbiont of aphids. *Ann. Rev. Microbiol.* **49**:55–94.
- BRAIG, K., P. D. ADAMS, and A. T. BRÜNGER. 1995. Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nat. Struct. Biol.* **2**:1083–1094.
- BRAIG, K., Z. OTWINOWSKI, R. HEGDE, D. C. BOISVERT, A. JOACHIMIAK, A. L. HORWICH, and P. B. SIGLER. 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**:578–586.
- BRYNNEL, E. U., C. G. KURLAND, N. A. MORAN, and S. G. E. ANDERSSON. 1998. Evolutionary rates for *tuf* genes in endosymbionts of aphids. *Mol. Biol. Evol.* **15**:574–582.
- BUCHNER, P. 1965. Endosymbiosis of animals with plant microorganisms. Wiley and Sons, New York.
- CHARLES, H., A. HEDDI, J. GUILLAUD, C. NARDON, and P. NARDON. 1997. A molecular aspect of symbiotic interactions between the weevil *Sitophilus oryzae* and its endosymbiotic bacteria: over-expression of a chaperonin. *Biochem. Biophys. Res. Commun.* **239**:769–774.
- CHENG, M. Y., F. U. HARTL, J. MARTIN, R. A. POLLOCK, F. KALOUSEK, W. NEUPERT, E. M. HALLBERG, R. L. HALLBERG, and A. L. HORWICH. 1989. Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature* **337**:620–625.
- CLARK, M. A., N. A. MORAN, and P. BAUMANN. 1999. Sequence evolution in bacterial endosymbionts having extreme base composition. *Mol. Biol. Evol.* **16**:1586–1598.
- CRANDALL, K. A., C. R. KELSEY, H. IMANICHI, H. C. LANE, and N. P. SALZMAN. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol. Biol. Evol.* **16**:372–382.
- DYKHUIZEN, D. E., and L. GREEN. 1993. Recombination in *E. coli* and the definition of biological species. *J. Bacteriol.* **173**:7257–7268.
- ELLIS, R. J. 1994. Molecular chaperones: opening and closing the anfinen cage. *Curr. Biol.* **4**:633–635.
- FAYET, O., T. ZIEGELHOFFER, and C. GEORGOPOULOS. 1989. The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* **171**:1379–1385.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**:368–376.
- . 1993. PHYLIP (phylogenetic inference package). Version 3.5c. Department of Genetics, University of Washington, Seattle.
- FENTON, W. A., and A. L. HORWICH. 1997. GroEL-mediated protein folding. *Protein Sci.* **6**:743–760.
- FENTON, W. A., Y. KASHI, K. FRUTAK, and A. L. HORWICH. 1994. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371**:614–619.
- FITCH, W. M. 1971. Towards defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20**:406–416.
- FUNK, D. J., J. J. WERNEGREN, and N. MORAN. 2001. Intra-specific variation in symbiont genomes: bottlenecks and the aphid-buchnera association. *Genetics* **157**:477–489.
- GEORGOPOULOS, C., and W. J. WELCH. 1993. Role of the major heat shock proteins as molecular chaperones. *Annu. Rev. Cell. Biol.* **9**:601–634.
- GETHING, M.-J., and J. SAMBROOK. 1992. Protein folding in the cell. *Nature* **355**:33–45.
- GOLDMAN, N., and Z. YANG. 1994. A codon-based model of nucleotide substitutions for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**:725–736.
- HARA, E., T. FUKATSU, K. KAKEDA, M. KENGAKU, C. OHTAKA, and H. ISHIKAWA. 1990. The predominant protein in an aphid endosymbiont is homologous to an *E. coli* heat shock protein. *Symbiosis* **8**:271–283.
- HARTL, F. U. 1996. Molecular chaperone in cellular protein folding. *Nature* **381**:571–580.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- HEMMINGSSEN, S. M., C. WOOLFORD, S. M. VAN DER VIES, K. TILLY, D. T. DENNIS, C. P. GEORGOPOULOS, R. W. HENDRIX, and R. J. ELLIS. 1988. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* **333**:330–334.
- HINDE, R. 1971. The control of the mycetocyte symbiotes of the aphids *Brevicoryne brassicae*, *Myzus persicae* and *Macrosiphum rosae*. *J. Insect Physiol.* **17**:1791–1800.
- HOGENHOUT, S. A., F. VAN DER WILK, M. VERBEEK, R. W. GOLDBACH, and J. F. VAN DEN HEUVEL. 2000. Identifying the determinants in the equatorial domain of Buchnera GroEL implicated in binding Potato leafroll virus. *J. Virol.* **74**:4541–4548.
- HUELSENBECK, J. P., and K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* **28**:437–466.
- ISHIKAWA, H. 1982. Host-symbiont interactions in the protein synthesis in the pea aphid, *Acyrtosiphon pisum*. *Insect Biochem.* **12**:613–622.
- . 1984. Characterization of the protein species synthesized in vivo and in vitro by an aphid endosymbiont. *Insect Biochem.* **14**:417–425.
- JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–123 in H. N. MUNRO, ed. *Mammalian protein metabolism*. Academic Press, New York.
- KAWATA, Y., M. KAWAGOE, K. HONGO, T. MIYAZAKI, T. HIGURASHI, T. MIZOBATA, and J. NAGAI. 1999. Functional communications between the apical and equatorial domains of GroEL through the intermediate domain. *Biochem.* **38**:15731–15740.
- 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.
- KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.01. The Pennsylvania State University, University Park, Pennsylvania.
- LAMBERT, J. D., and N. A. MORAN. 1998. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. USA* **95**:4458–4462.
- LANDRY, S. J., J. ZEILSTRA-RYALLS, O. FAYET, C. GEORGOPOULOS, and L. M. GIERASCH. 1993. Characterization of a functionally important mobile domain GroES. *Nature* **15**:255–258.



- LATORRE, A., A. MOYA, and F. J. AYALA. 1986. Evolution of mitochondrial DNA in *Drosophila subobscura*. Proc. Natl. Acad. Sci. USA **83**:8649–8653.
- LI, C., D. COX-FOSTER, S. M. GRAY, and F. GILDOW. 2001. Vector specificity of barley yellow dwarf virus (BYDV) transmission: identification of potential cellular receptors binding BYDV-MAV in the aphid, *Sitobion avenae*. Virology **286**:125–133.
- LORCA, O., A. GALAN, J. L. CARRASCOSA, A. MUGA, and J. M. VALPUESTA. 1998. GroEL under heat-shock. Switching from a folding to a storing function. J. Biol. Chem. **273**:32587–32594.
- LYNCH, M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol. Biol. Evol. **13**:209–220.
- . 1997. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. Mol. Biol. Evol. **14**:914–925.
- MAYNARD SMITH, J., C. B. DOWSON, and B. G. SPRATT. 1991. Localized sex in bacteria. Nature **349**:29–31.
- MAYNARD SMITH, J., N. SMITH, M. O'ROURKE, and B. SPRATT. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90**:4384–4388.
- MORAN, N. A. 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93**:2873–2878.
- MORAN, N. A., C. D. VON DOHLEN, and P. BAUMAN. 1995. Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. J. Mol. Evol. **41**:727–731.
- MULLER, H. J. 1964. The relation of the recombination to mutational advance. Mutat. Res. **1**:2–9.
- OHTA, T. 1973. Slightly deleterious mutant substitutions in evolution. Nature **246**:96–98.
- . 1992. The nearly neutral theory of molecular evolution. Ann. Rev. Ecol. Syst. **23**:263–286.
- OHTAKA, C., H. NAKAMURA, and H. ISHIKAWA. 1992. Structure of chaperonins from an intracellular symbiont and their functional expression in *Escherichia coli* groEL mutants. J. Bacteriol. **174**:1869–1874.
- RANSON, N. A., H. E. WHITE, and H. R. SAIBIL. 1998. Chaperonins. Biochem. J. **333**:233–242.
- RYE, H. S., S. G. BURSTON, W. A. FENTON, J. M. BEECHEM, Z. XU, P. B. SIGLER, and A. L. HORWICH. 1997. Distinct actions of *cis* and *trans* ATP within the double ring of the chaperonin GroEL. Nature **388**:792–798.
- RZHETSKY, A., and M. NEI. 1994. Unbiased estimates of the number of nucleotide substitutions when substitution rate varies among different sites. J. Mol. Evol. **38**:295–299.
- SAITOU, N., and M. NEI. 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
- SELANDER, R. K., D. A. CAUGANT, and T. S. WHITTAM. 1987. Genetic structure and variation in natural populations of *Escherichia coli*. Pp. 1625–1648 in F. NEIDHARDT, ed. *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society of Microbiology, Washington, D.C.
- SHARP, P. M. 1997. In search of molecular Darwinism. Nature **385**:111–112.
- SIGLER, P. B., Z. XU, H. S. RYE, S. G. BURSTON, W. A. FENTON, and A. L. HORWICH. 1998. Structure and function in GroEL-mediated protein folding. Annu. Rev. Biochem. **67**:581–608.
- SUZUKI, Y., and T. GOJOBORI. 1999. A method for detecting positive selection at single amino acid site. Mol. Biol. Evol. **16**:1315–1328.
- TAKEZAKI, N., A. RZHETSKY, and M. NEI. 1995. Phylogenetic test of the molecular clock and linearized tree. Mol. Biol. Evol. **12**:823–833.
- 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.
- THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.
- VAN HAM, R. C. H. J., F. GONZÁLEZ-CANDELAS, F. J. SILVA, B. SABATER, A. MOYA, and A. LATORRE. 2000. Postsymbiotic plasmid acquisition and evolution of the repA1-replicon in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. USA **97**:10855–10860.
- WEIR, B. S. 1996. Genetic data analysis II: method for discrete population genetic data. Sinauer Associates, Sunderland, Mass.
- WEISSMAN, J. S., H. S. RYE, W. A. FENTON, J. M. BEECHEM, and A. L. HORWICH. 1996. Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. Cell **84**:481–490.
- WERNEGREEN, J. J., and N. A. MORAN. 1999. Evidence for genetic drift in endosymbionts (*Buchnera*): analysis of protein-coding genes. Mol. Biol. Evol. **16**:83–97.
- XU, Z., A. L. HORWICH, and P. B. SIGLER. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)<sub>7</sub> chaperonin complex. Nature **388**:741–750.
- XU, Z., and P. B. SIGLER. 1998. GroEL/GroES: structure and function of the two-stroke folding machine. J. Struct. Biol. **124**:129–141.
- YANG, Z. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.
- . 2000. PAML: phylogenetic analysis by maximum likelihood. Version 3. University College London, UK.
- YOUNG, M. J., and S. A. FILICHKIN. 1999. Luteovirus interactions with aphid vector cellular components. Trends Microbiol. **7**:346–347.
- ZEILSTRA-RYALLS, J., O. FAYET, and C. GEORGOPOULOS. 1991. The universally conserved GroE (Hsp60) chaperonins. Ann. Rev. Microbiol. **45**:301–325.
- ZHANG, J., H. F. ROSENBERG, and M. NEI. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

WILLIAM MARTIN, reviewing editor

Accepted March 11, 2002