A Conservative Test of Genetic Drift in the Endosymbiotic Bacterium Buchnera: Slightly Deleterious Mutations in the Chaperonin *groEL*

Joshua T. Herbeck,* Daniel J. Funk,† Patrick H. Degnan* and Jennifer J. Wernegreen*,[1]

**Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts 02543 and †Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235*

Manuscript received April 22, 2003
Accepted for publication October 8, 2003

ABSTRACT

The obligate endosymbiotic bacterium *Buchnera aphidicola* shows elevated rates of sequence evolution compared to free-living relatives, particularly at nonsynonymous sites. Because Buchnera experiences population bottlenecks during transmission to the offspring of its aphid host, it is hypothesized that genetic drift and the accumulation of slightly deleterious mutations can explain this rate increase. Recent studies of intraspecific variation in Buchnera reveal patterns consistent with this hypothesis. In this study, we examine inter- and intraspecific nucleotide variation in *groEL*, a highly conserved chaperonin gene that is constitutively overexpressed in Buchnera. Maximum-likelihood estimates of nonsynonymous substitution rates across Buchnera species are strikingly low at *groEL* compared to other loci. Despite this evidence for strong purifying selection on *groEL*, our intraspecific analysis of this gene documents reduced synonymous polymorphism, elevated nonsynonymous polymorphism, and an excess of rare alleles relative to the neutral expectation, as found in recent studies of other Buchnera loci. Comparisons with *Escherichia coli* generally show patterns predicted by their differences in N_e. The sum of these observations is not expected under relaxed or balancing selection, selective sweeps, or increased mutation rate. Rather, they further support the hypothesis that drift is an important force driving accelerated protein evolution in this obligate mutualist.

SEVERAL features characterize genome evolution in *Buchnera aphidicola*, the obligate bacterial endosymbiont of aphids. First, Buchnera shows extreme reduction of genome size compared to *Escherichia coli*, the most closely related free-living species in the γ-proteobacteria. Buchnera genomes range in size from 450 kb (Gil *et al.* 2002) to 641 kb (Shigenobu *et al.* 2000), while those of natural *E. coli* isolates vary from 4.5 to 5.5 Mb (Bergthorsson and Ochman 1995). The genomes of Buchnera are also extremely AT biased, at ~26% GC (Shigenobu *et al.* 2000). In addition, Buchnera experiences elevated rates of sequence evolution across the genome, especially at nonsynonymous sites (Moran 1996; Rouhbakhsh *et al.* 1997; Clark *et al.* 1999; Wernegreen *et al.* 2001). Similar patterns of genome reduction and increased evolutionary rates have been documented in other obligate endosymbionts of insects (*e.g.*, Aksoy 2000; Clark *et al.* 2001; Wernegreen *et al.* 2002), and accelerated 16S rDNA evolution also characterizes the maternally transmitted symbionts of mollusks (Peek *et al.* 1998). Various studies suggest that reduced effective population sizes (N_e) and increased genetic drift may underlie these observed changes in the mode and tempo of molecular evolution (Funk *et al.* 2001; Mira and Moran 2002).

Specific aspects of their endosymbiosis with aphids may contribute to reduced N_e in Buchnera. The exclusive occurrence of these bacteria within aphid cells and a lack of any free-living stage reflect their reciprocally obligate relationship, in which Buchnera provides essential amino acids to, and receives nutrients from, the host (Shigenobu *et al.* 2000). Maternal transmission of Buchnera ensures its inheritance by host offspring, but inflicts a population bottleneck since only a few bacterial cells infect each developing egg or embryo (Buchner 1965; Mira and Moran 2002). Congruence among Buchnera and host phylogenies indicates the high fidelity and evolutionary stability of this transmission mode throughout the 150–200 million years of this mutualism (Munson *et al.* 1991). Furthermore, the wind-borne colony founding and rapid clonal population growth of aphids (Hales *et al.* 1997) results in bottlenecks that reduce the N_e of host and endosymbiont alike and produce distinct polymorphism patterns at aphid mitochondrial genes (Funk *et al.* 2001; Abbot and Moran 2002). An apparent lack of horizontal transfer among Buchnera strains (Buchner 1965; Funk *et al.* 2000; Wernegreen and Moran 2001; Tamas *et al.* 2002) may accentuate the effects of genetic drift caused by bacterial and aphid population bottlenecks.

An elevated rate of fixation of slightly deleterious

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY372289–AY372318 and AY372485–AY372493.

[1]*Corresponding author:* Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL St., Woods Hole, MA 02543. E-mail: jwernegreen@mbl.edu

mutations under bottleneck-induced drift may generally explain the increased rates of nonsynonymous divergence observed in endosymbionts, including Buchnera. However, alternative processes must also be considered. For example, the mutualistic endosymbiotic lifestyle may relax selective constraints at specific genes that are redundant in the host cell or may relax selection across the genome as a result of decreased maximum replication rates or diminished severity of the intracellular environment compared to that experienced by related free-living bacteria. Effects of relaxed selection can resemble those of decreased N_e because both will reduce the parameter $N_e s$ and thus increase substitution rates, as predicted by the nearly neutral theory of molecular evolution (Ohta 1973, 1992). Alternatively, elevated mutation pressure due to the loss of DNA repair genes in small endosymbiont genomes may drive rate acceleration (Andersson and Andersson 1999; Shigenobu *et al.* 2000; Akman *et al.* 2002; Tamas *et al.* 2002). Positive selection may also elevate evolutionary rates, but such selection typically acts at specific loci and is not expected to produce the genome-wide rate acceleration seen in Buchnera (Wernegreen and Moran 1999).

Fully distinguishing the effects of drift, relaxed selection, and increased mutation pressure on sequence variation is difficult, since these forces often have similar effects and may act simultaneously. For example, recent studies of interspecific divergence (Wernegreen and Moran 1999) and intragenomic variation (Palacios and Wernegreen 2002) indicate that mutation bias and drift largely shape codon usage in Buchnera, in contrast to the adaptive codon bias seen in *E. coli*. Interspecific comparisons show elevated ratios of nonsynonymous to synonymous substitutions (d_N/d_S) across Buchnera genes of varied functional categories (Clark *et al.* 1999; Wernegreen and Moran 1999). Increased mutation pressure can be ruled out as a sole explanation for these observations because it should affect d_N and d_S similarly and thus not influence their ratio. However, these patterns are predicted by both relaxed selection and genetic drift, so their contributions cannot be distinguished by interspecific approaches.

Population genetic analyses can more fully distinguish the contributions of drift, selection, and mutational pressure because each of these forces has distinct predicted effects on variation within species. Reduced N_e is expected to reduce levels of neutral polymorphism due to a reduction in the time to fixation or loss under genetic drift, but should increase levels of slightly deleterious polymorphism (for which $|s| \leq 1/N_e$; Ohta 1992) because a greater number of mutations will fall into this category. These weakly deleterious mutations would otherwise be quickly removed by selection in large populations but are free to persist and fluctuate under drift in small populations. Likewise, under reduced N_e, ratios of nonsynonymous to synonymous changes are expected to be higher within than between species, because even those slightly deleterious nonsynonymous mutations that fluctuate for a time within species under drift are most often eliminated by selection prior to fixation (McDonald and Kreitman 1991). Unlike relaxed selection or increased mutation rates, the hypothesis of bottleneck-induced drift also predicts an excess of young (and therefore rare) alleles, since few mutations will predate the bottleneck or have had sufficient time to rise to high frequency within populations (Tajima 1989).

Applying the population genetic approach, intraspecific studies of Buchnera from two aphid species (*Uroleucon ambrosiae* and *Pemphigus obesinymphae*) demonstrated predicted effects of bottlenecks and genetic drift on patterns and levels of polymorphism (Funk *et al.* 2001; Abbot and Moran 2002). These studies found extremely low levels of synonymous polymorphism and a significant excess of young, rare alleles compared to that expected under a neutral equilibrium model. They also detected an excess of nonsynonymous polymorphisms at a minority of assayed Buchnera genes (one of four loci).

The current study extends these prior investigations through comparative and intraspecific analyses of nucleotide variation in the chaperonin gene *groEL* in Buchnera and *E. coli*. In *E. coli*, *groEL* assists in protein folding (Fayet *et al.* 1989) and prevents misfolding under conditions of environmental stress (Bochdareva *et al.* 1988). *groEL* is constitutively overexpressed in Buchnera (Baumann *et al.* 1996) and accounts for ~10% of all proteins produced (Ishikawa 1984; Hara *et al.* 1990). In Buchnera, *groEL* may buffer against the accumulation of slightly deleterious amino acid substitutions that would otherwise cause conformational problems across the proteome (Moran 1996). This compensatory process has been demonstrated experimentally in *E. coli*, using simulated vertical transmission events, mutation accumulation, and induced *groEL* overexpression (Fares *et al.* 2002b). *groEL* of Buchnera has also acquired phosphotransferase activity as a novel histidine kinase (Morioka *et al.* 1993, 1994; Matsumoto *et al.* 1999). As predicted from its critical functions, *groEL* in Buchnera experiences stronger purifying selection than other Buchnera genes do (Palacios and Wernegreen 2002). Despite this, *groEL* nonetheless experiences accelerated protein evolution and evolves 2.4 times faster in Buchnera than in *E. coli* (Moran 1996).

This study compares patterns of polymorphism and divergence at *groEL* with those reported for several additional Buchnera genes sampled in the previous complementary study (Funk *et al.* 2001). We also examine site-specific synonymous and nonsynonymous substitution rates in *groEL* across the phylogeny of Buchnera associated with different Uroleucon species to evaluate purifying selection at *groEL* compared to other Buchnera loci. This interspecific analysis also allows us to test for positive selection, which was recently invoked in a study of

groEL from divergent Buchnera lineages (Fares *et al.* 2002a). The known functional importance of Buchnera *groEL* makes this chaperonin a strong candidate for a conservative test of the drift hypothesis. That is, detecting the signature of genetic drift at *groEL* would provide especially strong evidence that reduced N_e and drift play a general and dominant role in endosymbiont protein evolution.

MATERIALS AND METHODS

Samples: Although criteria for defining bacterial species are controversial, any workable species concept must consider the ecological range of a particular bacterial lineage (Cohan 2002). Buchnera of all aphids are technically considered the same species (*B. aphidicola*), but symbionts of different aphid species do not transfer and may be considered distinct populations ecologically and genetically. That is, the fixation of a mutation in Buchnera may occur throughout a particular aphid host species, but not beyond this ecological boundary. Therefore, for the purpose of this study, an "intraspecific" sample of Buchnera refers to endosymbionts of the same aphid host species, and "interspecific" refers to endosymbionts of different aphid host species.

The intraspecific data set of Buchnera includes *groEL* sequences of the 21 geographically widespread North American isolates described in Funk *et al.* (2001). Each of these isolates is derived from a single individual of the aphid species *U. ambrosiae* (Table 1). Patterns of polymorphism at *groEL* were compared to those of *dnaN*, *leuBC*, and *trpEG* sequences analyzed previously (Funk *et al.* 2001). Buchnera from various Uroleucon host species were included in the interspecific analyses of newly collected *groEL* sequences and previously published *dnaN*, *leuBC*, and *trpEG* sequences (Wernegreen *et al.* 2001; Table 1). Genomic DNA of these diverse Uroleucon species was kindly provided by N. A. Moran. Collection information and original DNA extraction methods for the Buchnera strains can be found in Funk *et al.* (2001) and Moran *et al.* (1999).

The *E. coli* data set included nine isolates from the ECOR *E. coli* reference strain collection (Ochman and Selander 1984). The *E. coli* strains used in this study were deliberately selected to span distinct genetic groups within the ECOR collection, similar to other *E. coli* population genetic studies (Hall and Sharp 1992; Nelson and Selander 1992; Boyd *et al.* 1994; Guttman and Dykhuizen 1994). Strains were chosen to represent major divisions (A–E) as indicated by multilocus enzyme electrophoresis (MLEE) analysis of the ECOR collection (Herzer *et al.* 1990). Isolates were kindly provided by H. Ochman (University of Arizona). Our sample included ECOR isolates 4 and 17 (group A), 29 (group B1), 51 and 60 (group B2), 46 and 50 (group D), and 31 and 37 (group E). This nonrandom sample is not directly comparable to the sample of Buchnera-*U. ambrosiae* isolates, which were collected without any prior knowledge of their genetic differentiation. However, the *E. coli* sample does provide a useful reference point to compare overall levels of variation between species of free-living and endosymbiotic bacteria. Patterns of nucleotide variation in *groEL* were compared to those in other genes (*celC*, *gapA*, *gutB*, *mdh*, *pabB*, and *putP*) analyzed previously in *E. coli*, using similar sampling across the ECOR collection. Sequences and alignments for these genes were retrieved from http://lifesci.rutgers.edu/~heylab/ProgramsandData/sites_data_sets.htm; all sequences are also available from GenBank, using accession numbers supplied in the original publications.

TABLE 1

Bacterial strains for which *groEL* was sampled in this study and the corresponding GenBank accession numbers of *groEL* sequences

Taxa	GenBank accession no.
E. coli	
ECOR isolates[a]	AY372310–AY372318
4 and 17 (group A)	
29 (group B1)	
51 and 60 (group B2)	
46 and 50 (group D)	
31 and 37 (group E)	
Buchnera	
Aphid host	
U. ambrosiae isolates[b]	AY372289–AY372309
U. astronomus[c]	AY372485
U. caligatum	AY372486
U. helianthicola	AY372487
U. jaceae	AY372488
U. obscurum	AY372489
U. rudbeckiae	AY372490
U. rapunculoidis	AY372491
U. solidaginis	AY372492
U. sonchi	AY372493

[a] Groups A–E of the ECOR collection represent major genetic groups as indicated by MLEE analysis (Herzer *et al.* 1990).

[b] Funk *et al.* (2001) provide information on collection information for the 21 *U. ambrosiae* isolates and the GenBank accession numbers for other genes sampled (*dnaN, leuBC, trpEG*). One original isolate (from Georgia) was not available for *groEL* sequencing. This isolate was substituted with another that was also collected from Georgia and was identical to the original Georgia isolate at *trpEG*, the most variable of the loci sampled.

[c] GenBank accession numbers of *dnaN*, *leuBC*, and *trpEG* for the interspecific sample of Buchnera-Uroleucon are available in Table 1 in Wernegreen and Moran (2001).

Molecular techniques: Gene amplification and sequencing of Buchnera loci other than *groEL* were described previously (Funk *et al.* 2001; Wernegreen *et al.* 2001). In this study, *groEL* sequences of *E. coli* and Buchnera were obtained through polymerase chain reaction (PCR) amplification, TA cloning of certain products, and automated sequencing as described below.

E. coli groEL: Cultures of Luria broth were inoculated with single colonies of freshly streaked ECOR isolates and incubated for 18 hr at 37° and 250 rpm. Genomic DNA was extracted using the DNeasy tissue kit (QIAGEN, Chatsworth, CA). We used PCR to amplify a 2.1-kb region of the *groE* operon with *E. coli*-specific primers designed for this study: ECgrES-42F (5′-AAACCACGTAAGCTCCGGCG-3′) and EcgrEL+35R (5′-ACCCCCAGACATTTCTGCC-3′). PCR reactions were performed at 25 μl and contained one-tenth volume of diluted DNA, PCR buffer [Fisher or Promega (Madison, WI)], 2.5 mM $MgCl_2$ (Promega), 1.0 mM dNTPs (Invitrogen, San Diego), 0.4 pmol/μl each primer, and 0.04 units of Taq polymerase (Fisher or Promega) and were brought to volume using sterile $_{dd}H_2O$. All PCR reactions were performed in a PTC-200 gradient thermocycler (MJ Research, Watertown,

MA) using initial denaturation of 94° for 2 hr, 35 cycles of 95° for 20 sec, 61° for 50 sec, 72° for 1 min, followed by a final extension at 72° for 7 min. *E. coli* PCR products were confirmed on agarose gels and cloned using the TOPO TA cloning kit and Top 10 One Shot chemically competent cells (Invitrogen) according to manufacturer's instructions. Clones were purified using Qiaquick PCR purification kit (QIAGEN) and were quantified by gel electrophoresis and spectrophotometry.

Buchnera-Uroleucon groEL: A region of *groES* and *groEL* of Buchnera was amplified from aphid DNA samples prepared in previous studies (Moran *et al.* 1999; Funk *et al.* 2001). Buchnera-specific PCR primers were designed to span a 2-kb region of the *groE* operon: uroGroES1F (5′-GAAAATTCGTC CGTTGCATG-3′) and uroG1640R (5′-ATCATTCCGCCCA TACC-3′). PCR reactions were performed as above, but with a reaction volume of 50 μl and an annealing temperature of 55°. PCR products were confirmed on agarose gels prior to purification using the Qiaquick kit (QIAGEN).

TA clones and PCR products of *groEL* genes were sequenced using appropriate primers on an ABI 3700 automated sequencer using Big Dye v3.0 (Applied Biosystems, Foster City, CA). Internal sequencing primers in both forward and reverse orientations were designed on the basis of the external reads. Sequences were assembled and edited using PHRED, PHRAP, and CONSED. All DNA assemblies were checked by eye and any ambiguous base calls were changed to N. Edited *groEL* sequences totaled 1644 bp for *E. coli* and 1569 bp for Buchnera. Bacterial isolates sampled and GenBank accession numbers are given in Table 1.

Data analysis: Sequences were aligned using both MacClade 4.04 (Maddison and Maddison 2000) and Se-Al v2.0a11 (Rambaut 2002) and edited by eye. Alignments for all data sets were unambiguous. Estimates of nucleotide variation were calculated using DNASP (Rozas and Rozas 1999). These included π, the average pairwise nucleotide diversity, and θ_w, the number of segregating sites for haploid genomes. Both π and θ_w are estimates of the neutral parameter ($\theta = 2N_e\mu$ for haploid, maternally inherited genomes, where N_e is the female effective population size). In addition, we calculated the absolute number of synonymous and nonsynonymous polymorphisms and used these to estimate *K*, the average pairwise divergence between two species. We applied multiple tests of neutrality of sequence evolution, including Tajima's *D* (Tajima 1989), Fu and Li's *D** (Fu and Li 1993), Fu and Li's *F** (Fu and Li 1993), and Fu's *Fs* (Fu 1997). Each of these statistics tests the prediction that two estimators of θ (*e.g.*, π and θ_w) should be equivalent in an equilibrium population that is evolving neutrally (Kreitman 2000).

We applied the McDonald-Kreitman test (MK test; McDonald and Kreitman 1991) and calculated the neutrality index (NI; Rand and Kann 1996) to compare the ratios of synonymous to nonsynonymous mutations within Buchnera-*U. ambrosiae* and between this species and Buchnera-*U. rudbeckiae.* The null hypothesis of neutrality predicts that the two ratios will be equal. Buchnera-*U. rudbeckiae* was used for comparison because this aphid host is closely related to *U. ambrosiae* (Funk *et al.* 2001; Wernegreen and Moran 2001) and was used as the outgroup in the previous study (Funk *et al.* 2001). The same tests were performed for *E. coli groEL*, using *Salmonella typhimurium* (GenBank accession no. U01039) as an outgroup (Brenner 1984; Dauga 2002).

Ratios of nonsynonymous (d_N) to synonymous (d_S) substitution rates provide an index for the strength and nature of selection at a given locus. We used the program *codeml* from the PAML package (Yang 2000) to estimate site-specific d_S and d_N for the Uroleucon interspecific data set of *groEL* and the *leuBC*, *trpEG*, and *dnaN* data sets examined previously (Wernegreen *et al.* 2001; Table 1). Parameters were optimized across the phylogeny of Buchnera *groEL* (data not shown), which is consistent with published phylogenies of Buchnera-Uroleucon (Clark *et al.* 1999; Wernegreen and Moran 2001) and the host (Moran *et al.* 1999). Parameter estimates were calculated using two nested likelihood models of sequence evolution. Model 0 assumes a single d_N/d_S (ω) across all sites in a gene, while model 3 allows ω to vary among codon sites, with three site classes available. (Neither model allows variation in ω among branches in the phylogeny.) The significance of differences in the likelihoods of the two models was evaluated with the likelihood ratio test (Huelsenbeck and Bull 1996). When interpreting d_N/d_S, ω values >1 are generally considered evidence for positive selection, while ω values <1 suggest purifying selection (Nielsen 2001). The power of site-specific ω estimates is particularly sensitive to the taxon sample size, as ω values can be overestimated for small samples such as the 10 species used in this study (Suzuki and Nei 2002). This does not seriously compromise its use here, however, since we are primarily interested in the presence and relative strength of selection among Buchnera genes (all of which would be similarly affected by such overestimates), rather than in quantifying it in absolute terms.

RESULTS

Buchnera

Intraspecific analysis of Buchnera-*U. ambrosiae*: The sample of 21 Buchnera-*U. ambrosiae groEL* sequences represented only five distinct haplotypes and 12 segregating sites, 10 of which were singletons (Tables 2 and 3). Buchnera *groEL* showed low nucleotide variation relative to other genes in Buchnera and to *E. coli groEL*. For example, nucleotide diversity per site (π_{tot}) was ~10-fold lower (0.10 for Buchnera) compared to that for *E. coli* (0.96; Table 2). Tests of neutrality in Buchnera *groEL* indicated an excess of rare alleles, with significantly negative values for Tajima's *D* for both silent and replacement sites and for Fu and Li's *D** and *F** (Table 4). The NI (Table 4) and MK test (Table 5) revealed a higher nonsynonymous to synonymous ratio for polymorphism than for divergence, and the MK test showed a significant deviation from the neutral expectation ($G = 5.1$, $P = 0.024$).

Interspecific analysis of d_N/d_S: The relatively low estimate of d_N/d_S (or ω) at Buchnera *groEL* compared to those at other Buchnera genes implies low rates of nonsynonymous substitution due to strong purifying selection. The ω estimate in model 0 (a single ω value for all sites) was 10–25 times lower for *groEL* than for other loci (Table 6). The higher d_N/d_S observed at *trpEG* and *leuABC* corroborated previous results showing accelerated nonsynonymous substitutions at these amino acid biosynthetic genes in Buchnera-Uroleucon (Wernegreen *et al.* 2001). Likelihood estimates of site-specific substitution rates (model 3) fit the data better than model 0 does for every gene (Table 6), indicating significant variation in ω site classes. A proportion of sites

TABLE 2

Summary of haplotypes and nucleotide variation across genes within populations of Buchnera-*U. ambrosiae* and *E. coli*

	N	Alleles	bp	η(*s*) %	π_{tot}	π_{non}	π_{syn}	π_{non}/π_{syn}	θ_{tot}	θ_{non}	θ_{syn}	CAI
						Buchnera						
groEL	21	5	1569	86	0.10	0.03	0.32	0.10	0.25	0.09	0.75	
dnaN[a]	21	7	1107	73	0.15	0.13	0.20	0.68	0.28	0.25	0.36	
leuBC[a]	21	6	1674	76	0.18	0.08	0.56	0.14	0.37	0.17	1.07	
trpEG[a]	21	5	1200	78	0.20	0.09	0.60	0.16	0.42	0.21	1.21	
						E. coli						
groEL	9	9	1644	50	0.96	0.16	3.35	0.05	1.07	0.31	3.33	0.77
gapA[b]	13	9	924		0.26	0.09	0.77	0.12	0.45	0.19	1.12	0.86
mdh[c]	20	13	864		1.19	0.16	3.83	0.04	1.34	0.22	4.40	0.58
gutB[d]	11	9	369		1.41	0.55	3.99	0.14	1.47	0.61	3.79	0.35
celC[d]	11	10	348		1.22	0.20	4.93	0.06	1.48	0.38	5.25	0.34
pabB[b]	11	6	1008		2.07	0.87	6.09	0.14	1.73	0.86	4.69	0.33
putP[e]	12	11	1893		2.06	0.33	7.89	0.04	2.26	0.51	7.81	0.28

N, sample size; alleles, number of unique haplotypes; bp, number of base pairs; η(s) %, percentage of segregating sites that are singleton alleles; π, nucleotide diversity per site (%) for all, nonsynonymous, and synonymous sites; θ, θ per site (%), from the total number of mutations, for all, nonsynonymous, and synonymous sites. Parameter values for *gapA, mdh, gutB, celC, pabB,* and *putP* were obtained from the literature, with associated references shown. CAI, codon adaptation index, values for *E. coli groEL* were calculated using CodonW (version 1.3 for UNIX, J. Peden; http://www.molbiol.ox.ac.uk/cu/).

[a] Funk *et al.* (2001).

[b] Guttman and Dykhuizen (1994).

[c] Boyd *et al.* (1994).

[d] Hall and Sharp (1992).

[e] Nelson and Selander (1992).

in *dnaN*, *leuBC*, and *trpEG* showed $\omega > 1$. In contrast, the highest ω estimated at *groEL* was still quite low (maximum $\omega = 0.1355$) and represented a small fraction (5.7%) of the total sites. This very low ω at *groEL* indicates strong purifying selection against amino acid changes and provides no evidence of positive selection (*i.e.*, $\omega > 1$).

E. coli

Each of the nine *E. coli* isolates represented a unique haplotype at *groEL* because, as in other population genetic studies of *E. coli* (see above), we selected isolates that span the known genetic diversity of the ECOR strain collection. Fifty percent of segregating sites were singletons and, as mentioned above, *E. coli* showed much higher levels of nucleotide diversity than did Buchnera at *groEL* (Table 2). Compared to other genes in *E. coli*, however, *groEL* showed low nucleotide diversity and extreme codon bias (Table 2). Tests of neutrality based on mutation spectra were nonsignificant in *E. coli* (Table 4), except for Tajima's *D* estimate for replacement mutations. Nevertheless, the relatively high NI (3.6; Table 4) and a significant MK test result ($G = 4.4$, $P = 0.036$; Table 5) indicate elevated ratios of nonsynonymous to synonymous polymorphism relative to divergence.

DISCUSSION

Molecular evolutionary rates in Buchnera are elevated at both synonymous and nonsynonymous sites, but the rate acceleration is greater at nonsynonymous sites (Moran 1996; Wernegreen and Moran 1999). In addition, endosymbionts such as Buchnera experience substitutions in the 16S rDNA gene that destabilize the secondary structure of the 16S rRNA molecule and further suggest the accumulation of deleterious changes by genetic drift (Lambert and Moran 1998). Previous intraspecific analyses (Wernegreen and Moran 1999; Funk *et al.* 2001; Abbot and Moran 2002) are completely consistent with the hypothesis that genetic drift underlies this observed rate increase. The present study extends these investigations and evaluates whether drift offers an explanation that is sufficiently general and powerful to account for variation at an overexpressed chaperonin, *groEL*.

Evolution of *groEL*—comparisons within Buchnera: Consistent with its functional importance in the symbiosis, we observed low d_N/d_S at Buchnera *groEL* compared to other Buchnera genes. Likelihood estimates of substitution rates between Buchnera species reveal only a small fraction (5.7%) of sites with ω ratios as high as 0.1355, in contrast to $\omega > 1$ for 2.6 and 7.7% of sites

TABLE 3

Polymorphic sites in Buchnera *groEL*

		Nucleotide site											
		0	0	0	0	0	0	1	1	1	1	1	1
		3	6	7	9	9	9	2	2	4	4	5	5
		0	0	3	4	8	9	8	9	2	6	3	6
Allele	*N*	0	4	2	2	1	0	8	0	5	9	7	3
Common	17	T	C	C	A	A	T	C	T	C	A	C	C
Georgia	1	C	·	T	·	·	A	A	C	·	·	T	T
Utah	1	·	T	A	·	C	·	·	·	·	G	·	·
Minnesota "B"	1	C	·	·	G	·	C	·	·	T	·	·	·
Ohio	1	C	·	·	·	·	C	·	·	·	·	·	·
Ti/Tv:		i	i	i/v	i	v	i/v	v	i	i	i	i	i
Codon position:		3	1	3	3	3	3	1	3	3	2	1	3
Amino acid:		V	P	G	L	K	T		R	N	D	L	V
AA polymorphism:			S			N			S		G		

·, same base as first sequence.

in the biosynthetic genes *leuBC* and *trpEG*, respectively (Table 6). A miniscule fraction of sites at *dnaN* (0.9%) showed $\omega > 1$. The action of positive selection at *leuBC* and *trpEG* is unclear, given the relatively small taxon sample available (Suzuki and Nei 2002). However, this comparison highlights the variable selective pressures experienced by different Buchnera loci and the exposure of *groEL* to comparatively strong purifying selection.

Our population genetic analysis of *groEL* adds to the growing evidence that strong effects of genetic drift in small endosymbiont populations explain unusual patterns of genetic variation in Buchnera. Our pertinent findings from Buchnera-*U. ambrosiae* include low levels of synonymous polymorphism, the apparent accumulation of slightly deleterious mutations suggested by MK tests, and an excess of young, rare alleles and singletons that is reflected in significant values of Tajima's *D* and Fu and Li's *D** and *F**. All these observations are consistent with the expected effects of drift under the repeated bottlenecking caused by bacterial transmission and aphid demographics. Such bottlenecks result in (1) a loss of allelic diversity; (2) a high proportion of extant alleles that have had insufficient time to rise to appreciable frequencies (Tajima 1989); and (3) a genome-wide decrease in the efficacy of selection, so that an increasing proportion of mutations fall into the nearly neutral category (Ohta 1992) and are observed as nonsynonymous polymorphisms.

Such slightly deleterious amino acid changes would be quickly removed in large populations where selection is more effective, but may fluctuate under genetic drift in small populations, thus contributing to elevated polymorphism (Ohta 1992). However, even these slightly deleterious nonsynonymous mutations are likely to be eliminated by selection prior to fixation (McDonald and Kreitman 1991; Brookfield and Sharp 1994), such that ratios of nonsynonymous to synonymous changes within species should exceed those between species. The common observation of this pattern in mitochondrial genes, for example, has recently been interpreted as indicating the unexpectedly high frequency of slightly deleterious alleles in the mitochon-

TABLE 4

Tests of neutral evolution

Taxon gene	NI	Tajima's D_{total}	Tajima's D_{non}	Tajima's D_{syn}	Fu and Li's *D**	Fu and Li's *F**	Fu's *Fs*
				Buchnera			
groEL	12.0	−2.160**	−1.873*	−1.979*	−2.912*	−3.128*	0.104
dnaN	4.5	−1.671			−2.019*	−2.035*	−1.960*
leuBC	0.9	−1.823*			−2.495*	−2.432*	1.107
trpEG	1.3	−1.924*			−2.454*	−2.414*	1.327
				E. coli			
groEL	3.6	−0.5278	−1.798*	−0.062	−0.297	−0.396	−1.838

*$P < 0.05$; **$P < 0.01$.

TABLE 5

McDonald-Kreitman tests comparing nonsynonymous and synonymous nucleotide variation at *groEL* within and between bacterial species

	Nonsynonymous changes	Synonymous changes	Ratio of nonsynonymous to synonymous changes	*G*	*P*
Buchnera-*U. ambrosiae* (divergence estimated using Buchnera-*U. rudbeckiae*)					
Polymorphism	4	10	0.40	5.1	0.024*
Divergence	1	26	0.04		
E. coli (divergence estimated using *S. typhimurium*)					
Polymorphism	9	38	0.24	4.4	0.036*
Divergence	6	81	0.07		

Columns G and P present test statistics and probability values, respectively, for *G*-tests of independence (Sokal and Rohlf 1991). *$P < 0.05$.

drial genome (Rand *et al.* 1994, 2000; Rand and Kann 1996). The large NI value of Buchnera *groEL* relative to other, less conserved, Buchnera genes also supports previous findings of greater ratios of nonsynonymous to synonymous polymorphism than divergence in more conserved genes (Rand and Kann 1996; Hasegawa *et al.* 1998).

Many explanatory alternatives to drift exist, but none are completely compatible with the sum of our findings. These alternatives are summarized here for the sake of completeness. First, although excess nonsynonymous polymorphism might be explained by relaxed selection, this mechanism should yield similar increases in nonsynonymous divergence, which is not observed. This dis-

TABLE 6

Maximum-likelihood estimation of synonymous and nonsynonymous substitution rates ($\omega = d_N/d_S$) of Buchnera genes

	ω, p	$-\ln L$	d.f.	$2\Delta l$[e]	*P*
	groEL (chaperonin), 10 *Uroleucon* spp., *K* (K2P)[a]: 0.061				
M0[b]	$\omega = 0.0057$	−3893.4453			
M3[c]	$\omega_0 = 0.00001$, $\omega_1 = 0.00001$, $\omega_2 = 0.1355$[d]	−3875.7003			
	$p_0 = 0.72234$, $p_1 = 0.22097$, $p_2 = 0.0567$		4	34.9	<0.001*
	dnaN (DNA replication), 9 *Uroleucon* spp., *K* (K2P): 0.132				
M0[b]	$\omega = 0.1464$	−3950.3818			
M3[c]	$\omega_0 = 0.01838$, $\omega_1 = 0.27676$, $\omega_2 = 1.4389$	−3902.4879			
	$p_0 = 0.54626$, $p_1 = 0.44447$, $p_2 = 0.0093$		4	95.7	<0.001*
	leuBC (biosynthetic), 7 *Uroleucon* spp., *K* (K2P): 0.136				
M0	$\omega = 0.0637$	−5322.9099			
M3	$\omega_0 = 0.01028$, $\omega_1 = 0.18046$, $\omega_2 = 1.0566$	−5196.1612			
	$p_0 = 0.70806$, $p_1 = 0.26620$, $p_2 = 0.0257$		4	204.8	<0.001*
	trpEG (biosynthetic), 7 *Uroleucon* spp., *K* (K2P): 0.258				
M0[b]	$\omega = 0.1401$	−4981.5798			
M3[c]	$\omega_0 = 0.02263$, $\omega_1 = 0.36473$, $\omega_2 = 1.1397$	−4826.5317			
	$p_0 = 0.61874$, $p_1 = 0.30383$, $p_2 = 0.0774$		4	310.1	<0.001*

[a] Mean pairwise divergence across all synonymous and nonsynonymous sites, using a Kimura two-parameter correction for multiple hits.

[b] M0, model 0 of *codeml*: ω (d_N/d_S) held constant across all lineages and all amino acid sites.

[c] M3, model 3 of *codeml*: ω free to vary among sites among three site classes, but held constant across lineages.

[d] ω_n, p_n: ω value for each site class (0, 1, or 2) with proportion (p_n) of sites in each gene that have the respective ω estimate (ω_n).

[e] Likelihood-ratio tests to compare the nested models of codon evolution, model 0 and model 3, as described in materials and methods.

crepancy might be a consequence of a recent relaxation of selection that is restricted to the focal study species (here, Buchnera-*U. ambrosiae*) and has not affected the outgroup lineage (here, *U. rudbeckiae*; Nachman *et al.* 1996). However, this hypothesis is inconsistent with the general rate of acceleration observed across Buchnera lineages associated with diverse aphid host taxa (Clark *et al.* 1999).

Second, although a recent selective sweep can also explain low synonymous polymorphism and left-skewed allele distributions (Tajima 1989), it cannot explain the excess of nonsynonymous intraspecific polymorphisms observed in the MK tests (Table 5).

Third, balancing selection (Polley and Conway 2001) may explain excess nonsynonymous polymorphism, but also predicts an excess of alleles at intermediate frequency rather than the excess of rare Buchnera alleles observed here and previously (Funk *et al.* 2001; Mira and Moran 2002).

Fourth, it has been proposed that the elevated substitution rate in Buchnera might entirely reflect increased mutation rates across the genome (Itoh *et al.* 2002). However, increased mutation pressure alone cannot explain the elevated d_N/d_S documented extensively for Buchnera (Moran 1996; Brynnel *et al.* 1998; Clark *et al.* 1999; Wernegreen and Moran 1999). Increased mutation rates should affect both nonsynonymous and synonymous sites equally and thus leave their ratio unchanged. Furthermore, elevated mutation rate cannot explain our observations of low synonymous polymorphism levels, skewed allele distributions, and significant MK test results.

Evolution of *groEL*—Buchnera *vs. E. coli*: Previous studies have compared patterns of sequence evolution in Buchnera and *E. coli*, due to their close phylogenetic relationship and extreme differences in life histories and population sizes (Clark *et al.* 1999; Wernegreen and Moran 1999). The effective population size of *E. coli* has been estimated at $\sim 2 \times 10^8$ (Hartl *et al.* 1994) and $\sim 2.5 \times 10^9$ (Ochman and Wilson 1987), while N_e of Buchnera is estimated to be $\sim 10^7$ for both Buchnera-*U. ambrosiae* (Funk *et al.* 2001) and Buchnera-*P. obesinymphae* (Abbot and Moran 2002). Unlike Buchnera, *E. coli* experiences limited recombination among strains and is globally distributed across diverse hosts. As discussed above, we sampled *E. coli* to deliberately span distinct genetic (MLEE) groups within the ECOR collection, as done in other *E. coli* population genetic studies (Hall and Sharp 1992; Nelson and Selander 1992; Boyd *et al.* 1994; Guttman and Dykhuizen 1994).

This sample allows us to compare overall levels of genetic variation between Buchnera and *E. coli* at *groEL* and to compare this chaperonin with other loci previously sampled from each species. At *groEL*, *E. coli* shows 5-fold higher levels of nonsynonymous polymorphism than Buchnera does ($\pi_{non} = 0.16$ and 0.03, respectively) and 10-fold higher levels of synonymous polymorphism ($\pi_{syn} = 3.35$ and 0.32), consistent with the predicted negative relationship between N_e and nucleotide polymorphism (Rand and Kann 1996). Further, π_{non}/π_{syn} is higher in Buchnera (0.10) *groEL* than in *E. coli* (0.05), consistent with decreased synonymous polymorphism and/or increased (slightly deleterious) nonsynonymous polymorphism in this bottlenecked endosymbiont. For both species, *groEL* is relatively conserved compared to other genes (Table 2). The low nonsynonymous divergence between *E. coli* and *S. typhimurium* at *groEL* ($K_A = 0.007$) compared to other loci sampled (mean $K_A = 0.039$ for 67 pairwise comparisons) indicates exceptionally strong purifying selection at this chaperonin (Sharp 1991). In *E. coli*, *groEL* shows extreme codon bias (0.77 codon adaptation index; Sharp and Li 1987a), consistent with its high expression level and demonstrated functional importance, and the large N_e of *E. coli*.

Contrary to expected patterns of sequence variation in large populations, *E. coli groEL*, like that of Buchnera, showed an excess of nonsynonymous polymorphism, as indicated by the significant MK test (Table 5). Like Buchnera, *E. coli* also exhibited a significant excess of rare alleles at replacement sites relative to the neutral expectation. However, the clonal and subdivided population structure of *E. coli* (Milkman 1973; Whittam *et al.* 1983) and our own nonrandom selection of genetically divergent and ecologically diverse strains for analysis may partially explain these patterns. For example, this sampling scheme may have predisposed us to find nonsynonymous mutations that had been fixed in local populations by either drift or divergent selection. Indeed, all of the nine nonsynonymous mutations in our sample of *E. coli groEL* are singletons unique to six isolates representing major ECOR divisions. In addition, selection on codon usage at high expression genes in *E. coli* may have influenced synonymous variation and thus affected the ratios of nonsynonymous to synonymous changes (Sharp and Li 1987b). Thus, any tentative explanations for the unexpected MK and Tajima's *D* results will require further analysis of additional genes and of closely related isolates within the ECOR groups. However, potential inflation of nonsynonymous polymorphism in *E. coli* would actually bias against the conclusions we draw from our comparison of π_{non}/π_{syn} in Buchnera and *E. coli*. That is, if our sampling strategy overestimated nonsynonymous polymorphism in *E. coli*, then π_{non}/π_{syn} would be elevated in *E. coli*. Despite this potential bias, π_{non}/π_{syn} is nonetheless greater in Buchnera than in *E. coli*, consistent with the effects of a decreased N_e and repeated bottlenecks.

In sum, our study documents patterns of nucleotide variation that are highly consistent with an important role for genetic drift in the nearly neutral molecular evolution of a highly constrained Buchnera locus. Our population genetic approach allows us to further demonstrate these patterns to be inconsistent with explanations based on alternative evolutionary mechanisms.

These results further support the hypothesis that population bottlenecks play a generally important role in the molecular evolution of bacterial endosymbionts.

We are grateful to Seth Bordenstein, Adam Lazarus, and two anonymous reviewers for comments on the manuscript and Roger Milkman for helpful discussion. We thank Jonas Sandström and Nancy Moran for collecting the Uroleucon samples used in the interspecific analysis and Paul Baumann for DNA extractions of several of these isolates. This work was made possible by support to J.J.W. from the National Institutes of Health (R01 GM62626-01), the National Science Foundation (DEB 0089455), the National Aeronautics and Space Administration Astrobiology Institute (NCC2-1054), and the Josephine Bay Paul and C. Michael Paul Foundation.

LITERATURE CITED

Abbot, P., and N. A. Moran, 2002 Extremely low levels of genetic polymorphism in endosymbionts (*Buchnera*) of aphids (*Pemphigus*). Mol. Ecol. **11:** 2649–2660.

Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba *et al.*, 2002 Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia.* Nat. Genet. **32:** 402–407.

Aksoy, S., 2000 Tsetse—a haven for microorganisms. Parasitol. Today **16:** 114–118.

Andersson, J. O., and S. G. Andersson, 1999 Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. **9:** 664–671.

Baumann, L., P. Baumann and M. A. Clark, 1996 Levels of *Buchnera aphidicola* chaperonin *groEL* during growth of the aphid *Schizaphis graminum.* Curr. Microbiol. **32:** 279–285.

Bergthorsson, U., and H. Ochman, 1995 Heterogeneity of genome sizes among natural isolates of *Escherichia coli.* J. Bacteriol. **177:** 5784–5789.

Bochdareva, E. S., N. M. Lissen and A. S. Girshovich, 1988 Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. Nature **336:** 254–257.

Boyd, E. F., K. Nelson, F. S. Wang, T. S. Whittam and R. K. Selander, 1994 Molecular genetic basis for allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica.* Proc. Natl. Acad. Sci. USA **91:** 1280–1284.

Brenner, D., 1984 *Enterobacteriaceae.* Williams & Wilkins, Baltimore.

Brookfield, J. F. Y., and P. M. Sharp, 1994 Neutralism and selection face up to DNA data. Trends Genet. **10:** 109–111.

Brynnel, E. U., C. G. Kurland, N. A. Moran and S. G. 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.* Interscience Publishers, New York.

Clark, M. A., N. A. Moran and P. Baumann, 1999 Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol. Biol. Evol. **16:** 1586–1598.

Clark, M. A., L. Baumann, M. L. Thao, N. A. Moran and P. Baumann, 2001 Degenerative minimalism in the genome of a psyllid endosymbiont. J. Bacteriol. **183:** 1853–1861.

Cohan, F. M., 2002 What are bacterial species? Annu. Rev. Microbiol. **56:** 457–487.

Dauga, C., 2002 Evolution of the *gyrB* gene and the molecular phylogeny of Enterobacteriaceae: a model molecule for molecular systematic studies. Int. J. Syst. Evol. Microbiol. **52:** 531–547.

Fares, M. A., E. Barrio, B. Sabater-Munoz and A. Moya, 2002a The evolution of the heat-shock protein *groEL* from *Buchnera*, the primary endosymbiont of aphids, is governed by positive selection. Mol. Biol. Evol. **19:** 1162–1170.

Fares, M. A., M. X. Ruiz-Gonzalez, A. Moya, S. F. Elena and E. Barrio, 2002b Endosymbiotic bacteria: *groEL* buffers against deleterious mutations. Nature **417:** 398.

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.

Fu, Y. X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147:** 915–925.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Funk, D. J., L. Helbling, J. J. Wernegreen and N. A. Moran, 2000 Intraspecific phylogenetic congruence among multiple symbiont genomes. Proc. R. Soc. Lond. B Biol. Sci. **267:** 2517–2521.

Funk, D. J., J. J. Wernegreen and N. A. Moran, 2001 Intraspecific variation in symbiont genomes: bottlenecks and the aphid-Buchnera association. Genetics **157:** 477–489.

Gil, R., B. Sabater-Munoz, A. Latorre, F. J. Silva and A. Moya, 2002 Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. Proc. Natl. Acad. Sci. USA **99:** 4454–4458.

Guttman, D. S., and D. E. Dykhuizen, 1994 Detecting selective sweeps in naturally occurring *Escherichia coli.* Genetics **138:** 993–1003.

Hales, D. F., J. Tomiuk, K. Wohrmann and P. Sunnucks, 1997 Evolutionary and genetic aspects of aphid biology: a review. Eur. J. Entomol. **94:** 1–55.

Hall, B. G., and P. M. Sharp, 1992 Molecular population genetics of *Escherichia coli*: DNA sequence diversity at the *celC, crr*, and *gutB* loci of natural isolates. Mol. Biol. Evol. **9:** 654–665.

Hara, E., T. Fukatsu, K. Kakeda, M. Kengaku, C. Ohtaka *et al.*, 1990 The predominant protein in an aphid endosymbiont is homologous to an *E. coli* heat shock protein. Symbiosis **8:** 271–283.

Hartl, D. L., E. N. Moriyama and S. A. Sawyer, 1994 Selection intensity for codon bias. Genetics **138:** 227–234.

Hasegawa, M., Y. Cao and Z. Yang, 1998 Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol. Biol. Evol. **15:** 1499–1505.

Herzer, P. J., S. Inouye, M. Inouye and T. S. Whittam, 1990 Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of Escherichia coli. J. Bacteriol. **172:** 6175–6181.

Huelsenbeck, J., and J. Bull, 1996 A likelihood ratio test to detect conflicting phylogenetic signal. Syst. Biol. **45:** 92–98.

Ishikawa, H., 1984 Characterization of the protein species synthesized *in vivo* and *in vitro* by an aphid endosymbiont. Insect Biochem. Mol. Biol. **14:** 417–425.

Itoh, T., W. Martin and M. Nei, 2002 Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. Proc. Natl. Acad. Sci. USA **99:** 12944–12948.

Kreitman, M., 2000 Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet. **1:** 539–559.

Lambert, J. D., and N. A. Moran, 1998 Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **95:** 4458–4462.

Maddison, D., and W. Maddison, 2000 *MacClade: Analysis of Phylogeny and Character Evolution.* Sinauer Associates, Sunderland, MA.

Matsumoto, K., M. Morioka and H. Ishikawa, 1999 Phosphocarrier proteins in an intracellular symbiotic bacterium of aphids. J. Biochem. **126:** 578–583.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila.* Nature **351:** 652–654.

Milkman, R., 1973 Electrophoretic variation in *Escherichia coli* from natural sources. Science **182:** 1024–1026.

Mira, A., and N. A. Moran, 2002 Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microbiol. Ecol. **44:** 137–143.

Moran, N. A., 1996 Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93:** 2873–2878.

Moran, N., M. Kaplan, M. Gelsey, T. Murphy and E. Scholes, 1999 Phylogenetics and evolution of the aphid genus *Uroleucon* based on mitochondrial and nuclear DNA sequences. Syst. Entomol. **24:** 85–93.

Morioka, M., H. Muraoka and H. Ishikawa, 1993 Chaperonin produced by an intracellular symbiont is an energy-coupling protein with phosphotransferase activity. J. Biochem. **114:** 246–250.

Morioka, M., H. Muraoka, K. Yamamoto and H. Ishikawa, 1994 An endosymbiont chaperonin is a novel type of histidine protein kinase. J. Biochem. **116:** 1075–1081.

Munson, M. A., P. Baumann, M. A. Clark, L. Baumann, N. A. Moran

et al., 1991 Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. J. Bacteriol. **173:** 6321–6324.

Nachman, M. W., W. M. Brown, M. Stoneking and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953–963.

Nelson, K., and R. K. Selander, 1992 Evolutionary genetics of proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. **174:** 6886–6895.

Nielsen, R., 2001 Statistical tests of selective neutrality in the age of genomics. Heredity **86:** 641–647.

Ochman, H., and R. K. Selander, 1984 Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:** 690–693.

Ochman, H., and A. C. Wilson, 1987 Evolutionary history of enteric bacteria, pp. 1649–1654 in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, edited by H. E. Umbarger. American Society for Microbiology, Washington, DC.

Ohta, T., 1973 Slightly deleterious mutant substitutions in evolution. Nature **246:** 96–98.

Ohta, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23:** 263–286.

Palacios, C., and J. J. Wernegreen, 2002 A strong effect of AT mutational bias on amino acid usage in *Buchnera* is mitigated at high expression genes. Mol. Biol. Evol. **19:** 1575–1584.

Peek, A. S., R. C. Vrijenhoek and B. S. Gaut, 1998 Accelerated evolutionary rate in sulfur-oxidizing endosymbiotic bacteria associated with the mode of symbiont transmission. Mol. Biol. Evol. **15:** 1514–1523.

Polley, S. D., and D. J. Conway, 2001 Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. Genetics **158:** 1505–1512.

Rambaut, A., 2002 *Se-Al Sequence Alignment Editor*. Oxford University Press, London/New York/Oxford.

Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol. Biol. Evol. **13:** 735–748.

Rand, D. M., M. Dorfsman and L. M. Kann, 1994 Neutral and nonneutral evolution of Drosophila mitochondrial DNA. Genetics **138:** 741–756.

Rand, D. M., D. M. Weinreich and B. O. Cezairliyan, 2000 Neutrality tests of conservative-radical amino acid changes in nuclear- and mitochondrially-encoded proteins. Gene **261:** 115–125.

Rouhbakhsh, D., M. A. Clark, L. Baumann, N. A. Moran and P. Baumann, 1997 Evolution of the tryptophan biosynthetic pathway in *Buchnera* (aphid endosymbionts): studies of plasmid-associated *trpEG* within the genus *Uroleucon*. Mol. Phylogenet. Evol. **8:** 167–176.

Rozas, J., and R. Rozas, 1999 DNASP, version 3: an integrated program for molecular population genetics and molecular evolutionary analyses. Bioinformatics **15:** 174–175.

Sharp, P. M., 1991 Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J. Mol. Evol. **33:** 23–33.

Sharp, P. M., and W. H. Li, 1987a The Codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:** 1281–1295.

Sharp, P. M., and W. H. Li, 1987b The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4:** 222–230.

Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki and H. Ishikawa, 2000 Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature **407:** 81–86.

Sokal, R. R., and F. J. Rohlf, 1991 *Biometry*, Ed. 3. W. H. Freeman, New York.

Suzuki, Y., and M. Nei, 2002 Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **19:** 1865–1869.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tamas, I., L. Klasson, B. Canback, A. K. Naslund, A. S. Eriksson *et al.*, 2002 50 million years of genomic stasis in endosymbiotic bacteria. Science **296:** 2376–2379.

Wernegreen, J. J., and N. A. Moran, 1999 Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. Mol. Biol. Evol. **16:** 83–97.

Wernegreen, J. J., and N. A. Moran, 2001 Vertical transmission of biosynthetic plasmids in aphid endosymbionts (*Buchnera*). J. Bacteriol. **183:** 785–790.

Wernegreen, J. J., A. O. Richardson and N. A. Moran, 2001 Parallel acceleration of evolutionary rates in symbiont genes underlying host nutrition. Mol. Phylogenet. Evol. **19:** 479–485.

Wernegreen, J. J., A. B. Lazarus and P. H. Degnan, 2002 Small genome of *Candidatus* Blochmannia, the bacterial endosymbiont of *Camponotus*, implies irreversible specialization to an intracellular lifestyle. Microbiology **148:** 2551–2556.

Whittam, T. S., H. Ochman and R. K. Selander, 1983 Multilocus genetic structure in natural populations of Escherichia coli. Proc. Natl. Acad. Sci. USA **80:** 1751–1755.

Yang, Z., 2000 *PAML: Phylogenetic Analysis by Maximum Likelihood*. University College, London.

Communicating editor: Z. Yang