Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 2873–2878, April 1996
Evolution

Accelerated evolution and Muller's rachet in endosymbiotic bacteria

(*Buchnera*/DNA sequence/mutation/nearly neutral theory/population size)

NANCY A. MORAN

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721

Communicated by Michael T. Clegg, University of California, Riverside, CA, November 27, 1995 (received for review August 3, 1995)

ABSTRACT **Many bacteria live only within animal cells and infect hosts through cytoplasmic inheritance. These endosymbiotic lineages show distinctive population structure, with small population size and effectively no recombination. As a result, endosymbionts are expected to accumulate mildly deleterious mutations. If these constitute a substantial proportion of new mutations, endosymbionts will show (*i*) faster sequence evolution and (*ii*) a possible shift in base composition reflecting mutational bias. Analyses of 16S rDNA of five independently derived endosymbiont clades show, in every case, faster evolution in endosymbionts than in free-living relatives. For aphid endosymbionts (genus *Buchnera*), coding genes exhibit accelerated evolution and unusually low ratios of synonymous to nonsynonymous substitutions compared to ratios for the same genes for enterics. This concentration of the rate increase in nonsynonymous substitutions is expected under the hypothesis of increased fixation of deleterious mutations. Polypeptides for all *Buchnera* genes analyzed have accumulated amino acids with codon families rich in A+T, supporting the hypothesis that substitutions are deleterious in terms of polypeptide function. These observations are best explained as the result of Muller's ratchet within small asexual populations, combined with mutational bias. In light of this explanation, two observations reported earlier for *Buchnera*, the apparent loss of a repair gene and the overproduction of a chaperonin, may reflect compensatory evolution. An alternative hypothesis, involving selection on genomic base composition, is contradicted by the observation that the speedup is concentrated at nonsynonymous sites.**

The rate of evolution for completely neutral mutations depends only on the mutation rate per individual (1). In contrast, for mutations affecting fitness, substitution rates also depend on population structure and the strength and direction of selection. In particular, rates of substitution for mildly deleterious mutations are greater in small populations and in strictly clonal ones (2–9). If a substantial proportion of mutations are mildly deleterious, small and/or asexual populations should show increased rates of sequence evolution (8).

Prokaryotes show a wide range of population structures. Population sizes for most are believed to be extremely large, with effective population sizes estimated at 10^9 for *Escherichia coli* (10, 11). Although natural populations of bacteria appear to be primarily clonal (e.g., see refs. 11–15), recent findings indicate evolutionarily significant levels of recombination (16, 17). In contrast to typical, free-living prokaryotes are endosymbiotic bacteria that live only within cells of eukaryotes and that are transmitted between hosts through maternal, cytoplasmic inheritance (18–20). Phylogenetic analyses based primarily on 16S rDNA sequences indicate that these endosymbiotic lineages are derived from a variety of bacterial subdivisions (21–24). Concordance between endosymbiont and host phylogenies indicates that at least some endosymbiotic infections are ancient. For example, *Buchnera aphidicola* appears to have infected its aphid hosts 100–250 million years (MY) ago (25, 26); some other endosymbiotic infections may be still older (27). Compared to free-living relatives, cytoplasmically inherited endosymbionts have tiny populations; in most, a bottleneck occurs each host generation when progeny are inoculated. They are effectively clonal since no recombination can occur between lineages sequestered in different hosts and since horizontal transfer between hosts appears to be either absent or extremely rare (22).

Endosymbiotic bacteria with this population structure are expected to accumulate mildly deleterious mutations. If such accumulation is occurring, several predictions can be made. First, endosymbiont sequences should evolve faster than those of free-living related lineages. Second, this faster evolution should occur at sites under selection but not at neutral sites, since the latter evolve independently of population structure. Third, since selective constraint will be relatively ineffective, any mutational bias, in which some bases are favored as the endpoints of mutations, should have a greater effect on base composition of the genome, including sites subject to selection.

In this paper, I examine support for the above predictions. First, relative rates of substitution are examined for 16S rDNA of five independently evolved endosymbiotic lineages. For *Buchnera*, the one case in which coding sequences are available, I address whether other genes evolve faster, as predicted. Second, the prediction that the rate increase is concentrated in sites subject to selection is tested by comparing ratios of synonymous to nonsynonymous substitutions for *Buchnera* and for *E. coli*/*Salmonella typhimurium*. Third, I present evidence that mutational bias has affected DNA sequence evolution in *Buchnera* and that the bias reflects evolution that is deleterious at the level of polypeptide function.

MATERIALS AND METHODS

Relative Rates of Substitution in Cytoplasmically Inherited Bacteria. Tests of rate increases were applied to 16S rDNA in five instances of cytoplasmically inherited bacteria. These were the mutualistic endosymbionts of aphids [*B. aphidicola* (26, 28)], whiteflies (21), mealybugs (23), and tsetse flies (29), and also *Wolbachia pipientis*, a cytoplasmically transmitted bacterium that causes reproductive abnormalities in arthropods (24, 30). Each was derived independently from free-living members of the Proteobacteria, and each infection appears to be ancient (22, 24, 25, 29).

Relative rates tests (31, 32) were used to test whether endosymbionts evolved faster than related, free-living bacteria. Each test depends on identifying an endosymbiont, a closely related free-living taxon and a more distant, reference taxon. Power and resolution are improved if the three taxa are as close as possible while fulfilling this phylogenetic requirement (31). Choice of taxa was governed by availability of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: MY, million years.

complete or nearly complete sequence, strong support for phylogenetic assumptions based on previous analyses of 16S rDNA sequences (21, 24–26, 29, 33), and closeness of relationship. In addition, the "Similarity_Rank" command in the ribosomal data base project (34) revealed a close relationship between *Oceanospirillum japonicum* and whitefly endosymbionts; this was the only previously undocumented relationship used in a test. Otherwise, sequences of related taxa were obtained from GenBank/EMBL. Only sites shared by all test taxa were included. Alignments were performed using PILEUP of GCG (35); divergence levels were low enough that alignments were largely unambiguous. Distances were calculated with GCG, using the Kimura two-parameter correction for multiple substitutions at a site.

Many tests are possible for each endosymbiont group since several sequences are available for all three taxon choices. Two tests were performed for each group, with all three taxa used in each test different. The two tests are not independent, because they analyze evolutionary distances across overlapping branches of the phylogeny. However, the paired tests help to minimize the possibility of a misleading result due to recent evolution in one test lineage.

GenBank/EMBL accession nos. for 16S rDNA are as follows: *Buchnera* from the aphid *Schizaphis graminum* [hereafter *Buchnera(Sg)*], M63246; *Buchnera* from *Schlechtendalia chinensis* [hereafter *Buchnera(Sc)*], Z19056; whitefly endosymbionts from *Siphonius phillyreae* and *Bemesia tabaci*, Z11927 and Z11925; mealybug endosymbionts from *Pseudococcus longispinus* and *Dysmicoccus neobrevipes*, M68889 and M68888; tsetse endosymbionts from *Glossina brevipalpis* and *G. austeni*, L37341 and L37340; and *W. pipientis* from *Culex pipiens* and *Trichogramma deion* X61768 and L02884. Free-living bacteria used as sister and reference taxa were *Aeromonas trota* (X60415), *Agrobacterium tumefaciens* (M11223), *Erythrobacter longus* (M96744), *E. coli* (J01695), *Legionella* sp. (X60080), *Nitrosococcus oceanus* (M96395), *Nitrosolobus multiformis* (M96401), *Proteus vulgaris* (X07652), *Pseudomonas aeruginosa* (X06684), *Pseudomonas testosteroni* (M11224), *Rhodopseudomonas globiformis* (M59066), and *Vibrio parahaemolyticus* (X74721).

Relative Rates of Nondegenerate Substitutions in Coding Genes of *B. aphidicola*. To determine whether rate increases extend to other genes, relative rates tests were applied to eight coding genes of *Buchnera*. These consisted of five genes coding for tryptophan biosynthetic enzymes: *trpa, trpb, trpc(f), trpd*, and *trpe* plus *dnaA, rpoD*, and *groEL*. *dnaA* is involved in the initiation of chromosome replication, *rpoD* encodes a subunit of RNA polymerase, and *groEL* encodes the large heat shock protein, a chaperonin. The five *trp* sequences are available for *Buchnera(Sc)* and *Buchnera(Sg)* (36, 37). These diverged at the base of the clade of extant aphids 100–250 MY ago according to fossil and molecular clock evidence (25, 38). For *trp* genes, two tests were performed using different taxa. Sequences of *dnaA* and *rpoD* are available for *Buchnera(Sg)* (39, 40); *groEL* sequence is available for *Buchnera* of *Acyrthosiphon pisum* (41).

For comparisons between *Buchnera* and even the closest free-living bacteria, synonymous sites are saturated for all genes and 2-fold degenerate sites are saturated for some genes. Thus, tests were performed only on nondegenerate sites (at which any substitution results in an amino acid change). Under the hypothesis of accumulation of deleterious mutations, these sites should be affected. Nondegenerate sites were defined using the method and program of Li (42). DNA alignments were made to conform to alignments of inferred amino acid sequences. Tests were otherwise similar to those for 16S rDNA, although choice of free-living taxa was more limited by sequence availability.

GenBank/EMBL accession nos. for coding genes are as follows: *Buchnera(Sg)* (*trp*, Z19055, Z21938; *dnaA*, M80817; *rpoD*, M90644), *Buchnera(Sc)* (*trp*, U09184–U09185), *Buchnera(Ap)* (*groEL*, S88668), *E. coli* (*trp*, V00364, V00366–V00368, V00372; *dnaA*, J01602; *rpoD*, J01687; *groEL*, X07850), *Ps. aeruginosa* (*trp*, M15826, M33814; *rpoD*, D90118; *groEL*, M63957), *Ps. putida* (*dnaA*, X14791), *S. typhimurium* (*trp*, M30285–M30286, V01376–V01378), *V. parahaemolyticus* (*trp*, X17149).

Relative Frequencies of Synonymous to Nonsynonymous Substitutions in *Buchnera*. Neutral mutations are expected to accumulate at the same rate in endosymbionts and free-living bacteria, whereas deleterious mutations should accumulate faster in endosymbionts due to small population size and absence of recombination. Thus, d_S/d_N (distance at synonymous sites/distance at nonsynonymous sites) is predicted to be smaller for *Buchnera(Sg)–Buchnera(Sc)* than for *E. coli–S. typhimurium*. All evolutionary change separating the two *Buchnera* occurred in endosymbiotic lineages, whereas all change separating *E. coli* and *S. typhimurium* occurred in free-living lineages. These enterics are in the clade most closely related to *Buchnera* (26). Molecular and fossil evidence suggests similar divergence times for the two pairs, with 140 MY hypothesized for the split between *E. coli–S. typhimurium* (9, 43) and 100–250 MY estimated for *Buchnera(Sg)–Buchnera(Sc)* (25). Thus d_S values are expected to be roughly similar for the two pairs; d_N values should be greater for *Buchnera*. Estimates of d_S and d_N were obtained using the program MEGA (44).

Additionally, d_S and d_N for *Buchnera(Sg)–Buchnera(Sc)* were obtained for partial sequences of *argS* (encoding arginyl-tRNA synthetase, which charges $tRNA^{arg}$), the only other coding sequence available for more than one *Buchnera* (ref. 45, accession nos. L18927 and L18932).

Evolution of Base Composition in Endosymbionts. A distinctive aspect of the *Buchnera* genome is its high [A+T]; the

Table 1. Relative rates tests for 16S rDNA of five groups of endosymbionts compared to related free-living bacteria

Taxon 1	Taxon 2	Taxon 3	K_{12}	K_{13}	K_{23}	$K_{13}-K_{23}$	z	K_{01}/K_{02}
Buchnera(Sg)	*E. coli*	*Ps. aeruginosa*	0.112	0.202	0.171	0.031 ± 0.00021	2.11*	1.75
Buchnera(Sc)	*A. trota*	*A. tumefaciens*	0.149	0.274	0.231	0.043 ± 0.00039	2.40**	1.82
Mealybug(Dn)	*R. globiformis*	*A. tumefaciens*	0.214	0.296	0.244	0.052 ± 0.00039	2.65**	1.64
Mealybug(Pl)	*N. multiformis*	*E. coli*	0.221	0.267	0.203	0.064 ± 0.00036	3.37***	1.81
Whitefly(Sp)	*O. japonicum*	*P. vulgaris*	0.167	0.199	0.156	0.043 ± 0.00023	2.78**	1.68
Whitefly(Bt)	*E. coli*	*N. oceanus*	0.230	0.236	0.183	0.053 ± 0.00031	3.37***	1.81
Tsetse(Gb)	*E. coli*	*N. oceanus*	0.150	0.231	0.182	0.049 ± 0.00033	2.70**	1.97
Tsetse(Ga)	*P. vulgaris*	*Ps. testosteroni*	0.189	0.288	0.236	0.052 ± 0.00045	2.47**	1.76
Wolbachia(Cp)	*A. tumefaciens*	*N. oceanus*	0.214	0.272	0.200	0.072 ± 0.00039	3.70***	2.03
Wolbachia(Td)	*E. longus*	*Ps. testosteroni*	0.248	0.310	0.234	0.076 ± 0.00041	3.72***	1.88

See text for details on tests and selection of taxa and for species name and GenBank/EMBL accession nos. In each test, taxon 1 is the endosymbiont, taxon 2 is a related free-living bacterium, taxon 0 is the most recent common ancestor of 1 and 2, and taxon 3 is a more distantly related reference taxon. K_{ij} is the estimate of substitutions per site between taxon i and taxon j. K_{01} and K_{02} were calculated as $K_{01} = (K_{13} - K_{23} + K_{12})/2$; $K_{02} = K_{12} - K_{01}$ (32). z scores were calculated as described (31). All tests support the hypothesis that endosymbionts evolve faster (one-tailed tests with H_o: $K_{01} </= K_{02}$ and H_1: $K_{01} > K_{02}$). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Table 2. Relative rates tests for substitution at nondegenerate sites in coding genes of *Buchnera* versus free-living relatives

Gene	No. of sites†	Taxon 1	Taxon 2	Taxon 3	K_{12}	K_{13}	K_{23}	$K_{13}-K_{23} \pm$ SD	z	K_{01}/K_{02}
trpa	508	*Buchnera(Sg)*	*E. coli*	*Ps. aeruginosa*	0.405	0.825	0.666	0.159 ± 0.075	2.11*	2.30
trpa	509	*Buchnera(Sc)*	*S. typhimurium*	*V. parahaemolyticus*	0.460	0.543	0.300	0.243 ± 0.049	5.00***	3.25
trpb	754	*Buchnera(Sg)*	*E. coli*	*Ps. aeruginosa*	0.171	0.391	0.361	0.030 ± 0.032	0.96	1.43
trpb	754	*Buchnera(Sc)*	*S. typhimurium*	*V. parahaemolyticus*	0.215	0.249	0.137	0.112 ± 0.022	4.88***	3.19
trpc(f)	852	*Buchnera(Sg)*	*E. coli*	*V. parahaemolyticus*‡	0.441	0.490	0.353	0.137 ± 0.036	3.75***	1.91
trpc(f)	860	*Buchnera(Sc)*	*S. typhimurium*	*V. parahaemolyticus*	0.439	0.486	0.352	0.134 ± 0.036	3.67***	1.88
trpd	631	*Buchnera(Sg)*	*E. coli*	*Ps. aeruginosa*	0.499	0.901	0.715	0.186 ± 0.086	2.17*	2.19
trpd	629	*Buchnera(Sc)*	*S. typhimurium*	*V. parahaemolyticus*	0.481	0.498	0.270	0.228 ± 0.042	5.41***	2.80
trpe	901	*Buchnera(Sg)*	*E. coli*	*Ps. aeruginosa*	0.346	0.715	0.673	0.042 ± 0.050	0.84	1.28
trpe	980	*Buchnera(Sc)*	*S. typhimurium*	*V. parahaemolyticus*	0.381	0.430	0.369	0.074 ± 0.086	1.95*	1.39
dnaA	781	*Buchnera(Sg)*	*E. coli*	*Ps. putida*	0.137	0.275	0.217	0.058 ± 0.024	2.44**	2.46
rpoD	1190	*Buchnera(Sg)*	*E. coli*	*Ps. aeruginosa*	0.118	0.287	0.255	0.032 ± 0.017	1.62	1.75
groEL	1071	*Buchnera(Ap)*	*E. coli*	*Ps. aeruginosa*	0.086	0.177	0.141	0.036 ± 0.014	2.34**	2.39

See text for GenBank/EMBL accession nos. Calculations are the same as for Table 1. All tests support faster evolution in *Buchnera* (one-tailed tests with H_o: $K_{01} </= K_{02}$ and H_1: $K_{01} > K_{02}$). *, $P < 0.05$, **, $P < 0.01$; ***, $P < 0.001$.
†Number of nondegenerate sites defined as in Li (42).
‡*Ps. aeruginosa* lacks fused *trpc(f)* and so was not used as a reference taxon.

GC% of *Buchnera(Sg)* is 27% compared to 50% for *E. coli* (46–48). Several authors have noted that, among sites, effects of mutational bias on substitutions are inversely related to the strength of selection (49–52). Among lineages, effects of mutational bias on substitutions are inversely related to the effectiveness of selection. In species with typically large populations, selection even at weakly selected sites can be very effective, as evidenced by selection on translational efficiency in highly expressed genes of *E. coli* (53, 54). Evolution in such lineages should be relatively unaffected by mutational bias. In endosymbionts, selection will be less effective at overriding effects of mutational bias, possibly resulting in a shift in genomic base composition.

Bias is expected to be greatest at relatively neutral sites but also to have affected selected sites. The first prediction was tested by examining %GC for different codon positions and also codon usage in relation to base composition. The second was tested by determining whether the bias toward A+T results in substantial evolution at the polypeptide level within *Buchnera*. Such an observation would imply that these substitutions are deleterious, at least at the level of polypeptide function. Calculations were performed using GCG (ref. 35; Codonpreference command).

RESULTS

Relative Rates of Substitution in Cytoplasmically Inherited Bacteria. For all five cases examined, rates of evolution of 16S rDNA sequences were significantly higher in endosymbionts than in related bacteria (Table 1). Rates are 1.7- to 2.7-fold greater for the endosymbiont lineage than the free-living sister lineage. Since endosymbiotic existence began after the shared ancestor, these values are minimum estimates of the actual rate increases.

Relative Rates of Nondegenerate Substitutions in Coding Genes of *B. aphidicola*. For all eight genes, rates of substitution at nondegenerate sites were greater in lineages leading to extant *Buchnera* relative to those leading to *E. coli* or *S. typhimurium* (Table 2). Rate differences were statistically significant for almost all tests, with *Buchnera* sequences evolving 1.4–3.2 times faster.

Synonymous and Nonsynonymous Substitutions in *Buchnera* Versus Enterics. Values for d_S/d_N are unusually low for *Buchnera(Sg)–Buchnera(Sc)* and consistently smaller than for *E. coli–S. typhimurium* (Table 3). Values for d_S are similar for the five *trp* genes, consistent with the premise that synonymous substitutions are essentially neutral and thus subject to similar levels of selective constraint. Values for d_S are also similar for the two taxon pairs, consistent with previous evidence for roughly equal divergence times. Thus, the lower d_S/d_N for *Buchnera* is due to its elevated rates of nonsynonymous substitutions.

These observations support the hypothesis that the speedup in *Buchnera* involves accumulation of deleterious mutations. They are inconsistent with the hypothesis that differences in generation time or mutation rate underlie the rate difference, since either of these should affect synonymous sites as much or more than nonsynonymous sites.

Higher d_N also will be observed if positive selection leads to amino acid substitutions, as in instances in which there is a change in function of the polypeptide (e.g., see refs. 19 and 55–58). This possibility can be excluded here. First, the change affects all genes; positive darwinian selection should be restricted to one or a few genes coding for the selected polypep-

Table 3. Distances based on synonymous (d_S) and nonsynonymous (d_N) substitutions for *trp* genes of two *Buchnera* and of *E. coli–S. typhimurium* and for *argS* of the two *Buchnera*

Gene (no. of codons)	Comparison	d_S	d_N	d_S/d_N
trpa(266)	*E. coli–S. typhimurium*	0.704 ± 0.032	0.083 ± 0.011	8.48
	Buchnera(Sg)–Buchnera(Sc)	0.570 ± 0.037	0.281 ± 0.018	2.03
trpb(397)	*E. coli–S. typhimurium*	0.581 ± 0.029	0.022 ± 0.005	26.77
	Buchnera(Sg)–Buchnera(Sc)	0.578 ± 0.031	0.167 ± 0.012	3.47
trpc(f)(469)	*E. coli–S. typhimurium*	0.629 ± 0.038	0.036 ± 0.018	17.5
	Buchnera(Sg)–Buchnera(Sc)	0.590 ± 0.035	0.273 ± 0.008	2.16
trpd(337)	*E. coli–S. typhimurium*	0.556 ± 0.031	0.016 ± 0.005	36.61
	Buchnera(Sg)–Buchnera(Sc)	0.567 ± 0.034	0.269 ± 0.016	2.11
trpe(520)	*E. coli–S. typhimurium*	0.577 ± 0.025	0.071 ± 0.008	8.13
	Buchnera(Sg)–Buchnera(Sc)	0.689 ± 0.026	0.264 ± 0.013	2.61
argS(131)	*Buchnera(Sg)–Buchnera(Sc)*	0.533 ± 0.056	0.370 ± 0.029	1.44

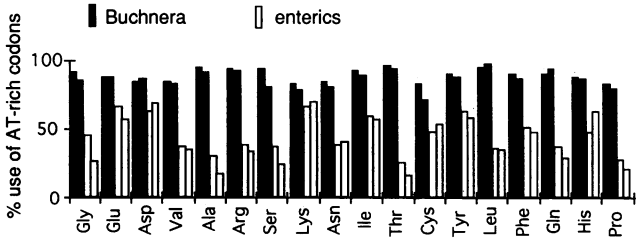

FIG. 1. Use of A+T-rich codons in *Buchnera* (solid bars) and representative free-living members of the γ-3 subgroup of the Proteobacteria (open bars). For each amino acid, codon families were categorized based on A+T content, and the frequency of codons with high versus low [A+T] was calculated. The four taxa (left to right) are *Buchnera(Sg), Buchnera(Sc), E. coli*, and *S. typhimurium*.

tides. Second, analyses of base composition indicate that most nucleotide sequence evolution is independent of or in opposition to selection at the polypeptide level. These analyses are presented in the next section.

Evolution of Base Composition in Endosymbionts. For a given amino acid, *Buchnera* sequences favor those codons that are richer in A+T (Fig. 1). Similarly, previous results for other genes within *Buchnera* indicate greater biases in codon third positions (41, 46, 47). Thus, selective constraint acting on polypeptide function appears to limit compositional bias. These observations are consistent with a role of mutational bias in determining overall genomic base composition (49–52).

More significant for the hypothesis that endosymbionts accumulate deleterious mutations is the finding that substitutions favoring A+T are not restricted to silent sites. For all five genes analyzed, the majority of amino acid differences between *E. coli* and *Buchnera(Sg)* are ones in which the amino acid in *Buchnera* corresponds to a codon family with higher [A+T] than the *E. coli* amino acid (Table 4). Likewise, amino acid frequencies within the *trp* sequences of *Buchnera* strongly favor those with codon families with high [A+T] (Fig. 2). These observations support the hypothesis that a substantial proportion of nucleotide substitutions in *Buchnera* are deleterious at the level of polypeptide function.

DISCUSSION

The results above suggest that sequence evolution in *Buchnera* is governed by two unusual circumstances. The first is the increased rate of fixation of mildly deleterious mutations, as a consequence of smaller populations and no recombination. This condition appears to be general in cytoplasmically inherited, endosymbiotic bacteria, as supported by the relative rates tests on 16S rDNA (Table 1). The second circumstance is some process leading to an increase in [A+T]. A similar pattern of codon usage bias to that observed in *Buchnera* is found in some other prokaryotes (50, 53, 59, 60) and in animal mitochondria (49). This has been attributed to mutational bias and contrasted with the adaptive codon bias documented for *E. coli*, in which highly expressed genes favor codons that are more

Table 4. Number of amino acid differences between *E. coli* and *Buchnera(Sg)* that result in a codon family in *Buchnera(Sg)* with greater [A+T], lower [A+T], or no change in [A+T] in genes coding for tryptophan biosynthetic enzymes

	No. of amino acids	Total no. of substitutions	More A+T	Less A+T	Same A+T
trpa	267	58	31	14	13
trpb	397	59	32	7	20
trpc(f)	471	198	145	20	33
trpd	337	50	25	9	16
trpe	563	87	59	12	16
Total	2035	452	292	62	98

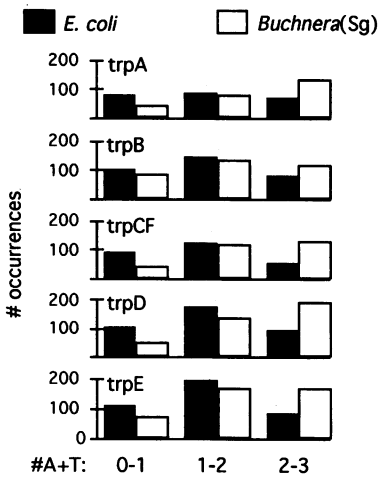

FIG. 2. Use of amino acids corresponding to codon families with different [A+T] content in *trp* genes of *Buchnera(Sg)* and *E. coli*. Amino acids are grouped according to number of A+T in the codon family: 0–1 (alanine, glycine, proline), 1–2 (asparagine, cysteine, glutamic acid, glutamine, histidine, serine, threonine, valine), and 2–3 (aspartic acid, isoleucine, lysine, phenylalanine, tyrosine). Amino acids not fitting one of these categories (arginine, leucine, methionine, tryptophan) are omitted here.

efficiently translated (53). If the bias in *Buchnera* toward A+T results strictly from mutational bias, then its major role in polypeptide evolution implies the accumulation of deleterious mutations. Many amino acid residues that have been conserved by selection from deep nodes of the phylogeny of the Proteobacteria are replaced in *Buchnera* with amino acids that allow greater A+T. In organisms with large and/or recombining populations, mutational bias has little effect on DNA sequences because it is overridden by selection, even at sites that are close to neutral (53). However, if selection is ineffective, as in small, asexual populations, mutational bias will affect substitution, at both neutral and weakly selected sites. Findings for *Buchnera* fit expectations for the combined effects of Muller's ratchet and mutational bias.

An alternative explanation for the bias in base composition is that selection acts on genomic base composition and favors greater [A+T]. Such selection could result from nucleotide availability during replication, faster replication of A+T-rich genomes, or some other aspect of DNA polymerization. Under this explanation, selection acting on DNA composition opposes and overrides selection on polypeptide function. Although this kind of selection cannot be ruled out entirely for *Buchnera*, selection on genomic [A+T] alone cannot explain the full suite of patterns in *Buchnera*. If the speedup in overall rate of substitution were due solely to selection on genomic base composition, the rate increase should be concentrated at synonymous rather than nonsynonymous sites. The opposite is observed. Although the accumulation of A+T is limited by site-specific functional constraints, as indicated by the codon bias (ref. 47; Fig. 1), the increase in evolutionary rates is concentrated at sites subject to selection (Table 3). Additionally, any effect of a single nucleotide on overall genomic base composition is tiny [on the order of $1/(3 \times 10^{-8})$, based on the genome size of *E. coli* (61)], suggesting that any selective differential associated with having A or T versus C or G at a particular position must be small. It would be remarkable if this weak selection often could override selection on polypeptide function to cause amino acid replacements at selectively constrained sites throughout the genome.

If endosymbionts in general are subject to Muller's ratchet, we expect their sequence evolution to be more often affected by mutational biases. In fact, base compositional biases are also apparent for whitefly endosymbionts, tsetse endosymbionts,

Table 5. Numbers of base differences for pairwise comparisons of 16S rDNA sequences of cytoplasmically inherited bacteria and representative relatives

Endosymbiont	Related lineage	Number of differences with endosymbiont having: More A+T	Less A+T	Same A+T	Total differences
Buchnera(Sg)	*E. coli*	97	26	19	142
Whitefly(Bt)	*E. coli*	142	49	61	252
Mealybug(Dn)	*Ps. testosteroni*	91	101	84	276
Tsetse(Gb)	*E. coli*	102	31	40	173
Wolbachia(Cp)	*A. tumefaciens*	136	40	64	240

Differences are subdivided into three categories: those in which the endosymbiont has A or T and the free-living lineage has C or G, those in which the endosymbiont has C or G and the free-living lineage has A or T, and those in which the difference is between A and T or between C and G. Selected relatives had intermediate base compositions typical for their subdivision.

and *Wolbachia* (Table 5). The exceptions are the mealybug endosymbionts, which show a base composition for this sequence near 50%GC, close to that of free-living relatives. The lack of bias is consistent with accumulation of deleterious mutations in this clade since the direction and degree of mutational bias are expected to be organism specific. Indeed, the mealybug endosymbionts, which show accelerated evolution (Table 1) but no apparent shift in base composition (Table 5), are further evidence that the rate increase is not due to selection on genomic base composition.

An alternative hypothesis for the rate increase observed at sites subject to selection in *Buchnera* is that selection in general is weaker in endosymbionts, which may be buffered against the range of environments encountered by free-living bacteria. Generally weaker selection would have similar consequences for sequence evolution as small population size and asexuality. However, relaxed selection at all genes seems unlikely. For example, in the case of *Buchnera*, bacterial activities are essential to hosts (19), and selection on the host/endosymbiont association has caused adaptive evolution within *Buchnera* (36, 46). The *trp* genes in particular are known to function as an essential part of the association because aphids require tryptophan synthesized by *Buchnera*. Becoming endosymbiotic might affect the strength and direction of selection on different genes and sites within genes but would seem unlikely to simply weaken selection at all genes. The decreased role of selection at all loci analyzed is more readily explained as an effect of population structure. Changes in population structure will affect all loci; changes in selective coefficients are expected to vary among loci.

Implications for Universal Rate in Bacteria. Previous rate calculations for 16S rDNA of *Buchnera*, calibrated with dates based on fossils of the cospeciating aphid hosts, were about 1.5–2 times greater than those proposed earlier (9, 43) for a variety of free-living bacteria, including enterics (25, 28). The analyses of Table 1 imply that rates calculated for free-living bacteria may be quite accurate, with rates observed in *Buchnera* being somewhat atypical.

The fact that some DNA sequence evolution appears to be constant on a scale of absolute time rather than generations has been considered evidence that many substitutions are slightly deleterious rather than strictly neutral (8, 62, 63). Under this argument, the tendency of rates to scale to absolute time has been explained as a consequence of a negative association between population size and generation time: organisms with small populations evolve faster due to fixation of slightly deleterious mutations but evolve more slowly because they tend to be large and to have long generation times. But any relationship between population size and generation time has many exceptions. Endosymbionts, such as *Buchnera*, have relatively small populations and short generations. As exceptions, they provide unusually good support for Ohta's nearly neutral theory (review in ref. 8), which holds that a large proportion of mutations are mildly deleterious.

Compensatory Processes. Theoretical models indicate that mutational load in small asexual populations can lead to extinction (3–5, 64). *Buchnera* has persisted over 100 MY without becoming extinct, although it has been lost from a few aphid lineages (e.g., see ref. 65). It is possible that adaptive evolution on the part of hosts, which retain sexual recombination and large populations, compensates for deleterious evolution in their endosymbionts. In addition, compensatory processes may occur in endosymbionts themselves.

In lineages subject to Muller's ratchet, selection can favor increased severity of deleterious mutations (3, 10). Lack of repair genes in organelles has been noted as supportive of this theoretical finding (3); however, organelles provide rather weak evidence since they have lost most genes (66). The apparent absence of *recF* in *Buchnera(Sg)* (39) provides stronger support, since it is missing from the position expected on the basis of knowledge of *E. coli*, with presence of >50 other coding genes now confirmed in *Buchnera(Sg)* (46). An alternative interpretation of the apparent loss of *recF* is that its deletion itself constitutes a mildly deleterious mutation.

During normal existence within hosts, *Buchnera* produces unusually high amounts of the stress protein, or chaperonin, GroEL (48, 67). High levels of GroEL also occur in endosymbionts of tsetse flies (68). In typical prokaryotes, *groEL* is expressed at low levels except in response to environmental stress, such as heat. GroEL mediates the folding of a variety of polypeptides into their functional, folded forms (69–71). Constitutive high expression of *groEL* in *Buchnera* (72) could be a compensatory mechanism that allows enzymes to retain functional conformation despite decreased stability caused by the accumulation of many amino acid substitutions.

Endosymbiosis is an evolutionary innovation without which many animal groups would not exist (20). The patterns described for *Buchnera* raise the possibility that, once "captured," endosymbionts undergo long-term deterioration due to accumulation of mutations, possibly limiting long-term fitness of hosts.

P. Baumann and E. Waters provided stimulating discussions and useful suggestions throughout this study. For ideas and advice, I also thank A. Colwell, C. Kurland, B. Normark, J. Seger, B. Sullender, and C. von Dohlen. This work was supported by the National Science Foundation (Grant DEB-9306495).

1. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, New York).
2. Felsenstein, J. (1974) *Genetics* **78,** 737–756.
3. Gabriel, W., Lynch, M. & Burger, R. (1993) *Evolution* **47,** 1744–1757.
4. Kondrashov, A. S. (1988) *Nature (London)* **334,** 435–440.
5. Lynch, M., Burger, R., Butcher, D. & Gabriel, W. (1993) *J. Hered.* **84,** 339–344.
6. Muller, H. J. (1964) *Mutat. Res.* **1,** 2–9.
7. Ohta, T. (1987) *J. Mol. Evol.* **26,** 1–6.
8. Ohta, T. (1993) *Annu. Rev. Ecol. Syst.* **23,** 263–286.

9. Ochman, H. & Wilson, A. C. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. Neidhardt, F. (Am. Soc. for Microbiol., Washington, DC), pp. 1649–1654.
10. Stephan, W., Chao, L. & Smale, J. G. (1993) *Genet. Res.* **61,** 225–231.
11. Selander, R. K., Caugant, D. A. & Whittam, T. S. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. Neidhardt, F. (Am. Soc. for Microbiol., Washington, DC), pp. 1625–1648.
12. Maynard Smith, J. (1993) *J. Hered.* **84,** 326–327.
13. Mckane, M. & Milkman, R. (1995) *Genetics* **139,** 35–43.
14. Milkman, R. & Bridges, M. M. (1993) *Genetics* **133,** 455–468.
15. Whittam, T. S. (1992) *Curr. Biol.* **2,** 676–678.
16. Dykhuizen, D. E. & Green, L. (1991) *J. Bacteriol.* **173,** 7257–7268.
17. Guttman, D. S. & Dykhuizen, D. E. (1994) *Science* **266,** 1380–1383.
18. Buchner, P. (1965) *Endosymbiosis of Animals with Plant Microorganisms* (Interscience, New York).
19. Houk, E. J. & Griffiths, G. W. (1980) *Annu. Rev. Entomol.* **25,** 161–187.
20. Saffo, M. B. (1992) *Am. Zool.* **32,** 557–565.
21. Clark, M. A., Baumann, L., Munson, M. A., Baumann, P., Campbell, B. C., Duffus, J. E., Osborne, L. S. & Moran, N. A. (1992) *Curr. Microbiol.* **25,** 119–123.
22. Moran, N. A. & Baumann, P. (1994) *Trends Ecol. Evol.* **9,** 15–20.
23. Munson, M., Baumann, P. & Moran, N. (1992) *Mol. Phylog. Evol.* **1,** 26–30.
24. Ohtaka, C. & Ishikawa, H. (1993) *J. Mol. Evol.* **36,** 121–126.
25. Moran, N. A., Munson, M. A., Baumann, P. & Ishikawa, H. (1993) *Proc. R. Soc. London B.* **253,** 167–171.
26. Munson, M. A., Baumann, P., Clark, M. A., Baumann, L., Moran, N. A., Voegtlin, D. J. & Campbell, B. C. (1991) *J. Bacteriol.* **173,** 6321–6324.
27. Distel, D., Felbeck, H. & Cavanaugh, C. M. (1994) *J. Mol. Evol.* **38,** 533–542.
28. Moran, N. A., von Dohlen, C. D. & Baumann, P. (1995) *J. Mol. Evol.* **41,** 727–731.
29. Aksoy, S., Pourhosseini, A. A. & Chow, A. (1995) *Insect Mol. Biol.* **4,** 15–22.
30. Breeuwer, J. A. J., Stouthamer, R., Barns, S. M., Pelletier, D. A., Weisburg, W. G. & Werren, J. H. (1992) *Insect Mol. Biol.* **1,** 25–36.
31. Muse, S. V. & Weir, B. S. (1992) *Genetics* **132,** 269–276.
32. Wu, C.-I. & Li, W.-H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1741–1745.
33. Woese, C. R. (1987) *Microbiol. Rev.* **52,** 221–271.
34. Larsen, N., Olsen, G. J., Maidak, B. L., McCaughey, M. J., Overbeek, R., Macke, T. J., Marsh, T. L. & Woese, C. R. (1993) *Nucleic Acids Res.* **21,** 3021–3023.
35. Program Manual for the Wisconsin Package (1994) (Genetics Computer Group, Madison, WI), Version 8.
36. Lai, C.-Y., Baumann, P. & Baumann, L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3819–3823.
37. Lai, C.-Y., Baumann, P. & Moran, N. A. (1995) *Insect Mol. Biol.* **4,** 47–59.
38. Heie, O. E. (1987) in *World Crop Pests, Aphids*, eds. Minks, A. K. & Harrewijn, P. (Elsevier, Amsterdam), Vol. 2A, pp. 367–391.
39. Lai, C.-Y. & Baumann, P. (1992) *Gene* **113,** 175–181.
40. Lai, C.-Y. & Baumann, P. (1992) *Gene* **119,** 113–118.
41. O'Neill, S. L., Giordano, R., Colbert, A., Karr, T. L. & Robertson, H. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 2699–2702.
42. Li, W.-H. (1993) *J. Mol. Evol.* **36,** 96–99.
43. Ochman, H. & Wilson, A. C. (1987) *J. Mol. Evol.* **26,** 74–86.
44. Kumar, S., Tamura, K. & Nei, M. (1993) MEGA, Molecular Evolutionary Genetic Analysis (Pennsylvania State Univ., University Park, PA), Version 1.01.
45. Munson, M. A., Baumann, L. & Baumann, P. (1993) *Gene* **137,** 171–178.
46. Baumann, P., Baumann, L., Lai, C.-Y., Rouhbakhsh, D., Moran, N. A. & Clark, M. A. (1995) *Annu. Rev. Microbiol.* **41,** 55–94.
47. Clark, M. A., Baumann, L. & Baumann, P. (1992) *Curr. Microbiol.* **25,** 283–290.
48. Ishikawa, H. (1987) *J. Mol. Evol.* **24,** 205–211.
49. Jermiin, L. S., Graur, D., Lowe, R. M. & Crozier, R. H. (1994) *J. Mol. Evol.* **39,** 160–173.
50. Muto, A. & Osawa, S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 166–169.
51. Sueoka, N. (1961) *Proc. Natl. Acad. Sci. USA* **47,** 1141–1149.
52. Sueoka, N. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2653–2657.
53. Sharp, P. M. (1989) in *Evolution and Animal Breeding*, eds. Hill, W. G. & MacKay, T. F. C. (CAB, London), pp. 23–32.
54. Sharp, P. M. & Li, W.-H. (1986) *J. Mol. Evol.* **24,** 28–38.
55. Hughes, A. L. & Hughes, A. K. (1993) *J. Mol. Evol.* **36,** 249–254.
56. Hughes, A. L. (1993) *Mol. Biol. Evol.* **10,** 1343–1359.
57. Jolles, J., Polles, P., Bowman, B. H., Prager, E. M., Stewart, C.-B. & Wilson, A. C. (1989) *J. Mol. Evol.* **28,** 528–535.
58. Swanson, W. J. & Vacquier, V. D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 4957–4961.
59. Andersson, S. & Kurland, C. (1990) *Microbiol. Rev.* **54,** 198–219.
60. Winkler, H. H. & Wood, D. O. (1988) *Biochimie* **70,** 977–986.
61. Riley, M. & Krawiec, S. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. Neidhardt, F. (Am. Soc. for Microbiol., Washington, DC), pp. 967–981.
62. Chao, L. & Carr, D. E. (1993) *Evolution* **47,** 688–690.
63. Kimura, M. (1989) in *Evolution and Animal Breeding*, eds. Hill, W. G. & MacKay, T. F. C. (CAB, London), pp. 13–16.
64. Kondrashov, A. S. (1994) *Genetics* **137,** 311–318.
65. Fukatsu, T. & Ishikawa, H. (1992) *J. Insect Physiol.* **38,** 765–773.
66. Gillham, N. W. (1994) *Organelle Genes and Genomes* (Oxford Univ. Press, New York), pp. 92–108.
67. Ishikawa, H. (1989) *Int. Rev. Cytol.* **116,** 1–45.
68. Aksoy, S. (1995) *Insect Mol. Biol.* **4,** 23–29.
69. Becker, J. & Craig, E. A. (1994) *Eur. J. Biochem.* **219,** 11–23.
70. Hartl, F. U. (1994) *Nature (London)* **371,** 557–558.
71. Hendrik, J. P. & Hartl, F. U. (1993) *Annu. Rev. Biochem.* **62,** 349–384.
72. Baumann, P., Baumann, L. & Clark, M. A. (1996) *Curr. Microbiol.* **32,** 1–7.