Sequence Evolution in Bacterial Endosymbionts Having Extreme Base Compositions

Marta A. Clark, Nancy A. Moran,† and Paul Baumann**

*Microbiology Section, University of California, Davis; and †Department of Ecology and Evolutionary Biology, University of Arizona

A major limitation on ability to reconstruct bacterial evolution is the lack of dated ancestors that might be used to evaluate and calibrate molecular clocks. Vertically transmitted symbionts that have cospeciated with animal hosts offer a firm basis for calibrating sequence evolution in bacteria, since fossils of the hosts can be used to date divergence events. Sequences for a functionally diverse set of genes have been obtained for bacterial endosymbionts (*Buchnera*) from two pairs of aphid host species, each pair diverging 50–70 MYA. Using these dates and estimated numbers of *Buchnera* generations per year, we calculated rates of base substitution for neutral and selected sites of protein-coding genes and overall rates for rRNA genes. *Buchnera* shows homogeneity among loci with regard to synonymous rate. The *Buchnera* synonymous rate is about twice that for low-codon-bias genes of *Escherichia coli–Salmonella typhimurium* on an absolute timescale, and fourfold higher on a generational timescale. Nonsynonymous substitutions show a greater rate disparity in favor of *Buchnera*, a result consistent with a genomewide decrease in selection efficiency in *Buchnera*. Ratios of synonymous to nonsynonymous substitutions differ for the two pairs of *Buchnera*, indicating that selection efficiency varies among lineages. Like numerous other intracellular bacteria, such as *Rickettsia* and *Wolbachia*, *Buchnera* has accumulated amino acids with codons rich in A or T. Phylogenetic reconstruction of amino acid replacements indicates that replacements yielding increased A+T predominated early in the evolution of *Buchnera*, with the trend slowing or stopping during the last 50 Myr. This suggests that base composition in *Buchnera* has approached a limit enforced by selective constraint acting on protein function.

Introduction

Prokaryotes constitute the bulk of present and past biotic diversity. While the widespread availability of DNA sequences has allowed major advances in the understanding of prokaryote phylogenetic relationships (Woese 1987; Pace 1997; Maidak et al. 1999), we have little firm evidence on which to reconstruct the timing of events in bacterial evolution. This is largely because bacterial groups do not possess distinctive morphologies that allow their origins to be dated with fossil records, as is routinely possible for many groups of animals and plants.

Since the extent of differences in DNA sequences correlates with time since divergence and since sequences are easily obtained for pairs of living bacterial taxa, molecular clocks might help to elucidate the timing of events during bacterial diversification. However, to calibrate rates of sequence evolution, the dates of some ancestors must be reconstructed using independent evidence; the calibrated rate can then be used to estimate divergence dates of modern taxa for which no other evidence is available. In some recent papers (Doolittle et al. 1996; Feng, Cho, and Doolittle 1997), absolute rates of amino acid sequence evolution are presumed to be similar among diverse lineages of eukaryotes and prokaryotes, a situation which would allow the routine application of molecular clocks calibrated using the fossil records of animals. However, little evidence supports this assumption, and some evidence contradicts it (Li 1993*b*; Ayala 1997). Thus, independent dating evidence is needed for as many bacterial ancestors as possible. A variety of approaches have been used to determine dates for bacterial divergence events, but most depend on uncertain assumptions (e.g., Ochman and Wilson 1987). The best supported bacterial divergence dates are those reconstructed for endosymbionts that have undergone synchronous cospeciation with their animal hosts. Dates for ancestral hosts, derived from fossil evidence, can be extended to the corresponding ancestors of the bacterial endosymbionts. Such dating has been proposed for *Buchnera aphidicola*, the bacterial endosymbionts of aphids (Insecta: Hemiptera: Aphididae) (Munson et al. 1991; Moran et al. 1993), and for a variety of other animal endosymbionts (e.g., Bandi et al. 1995; Schröder et al. 1996; Moran and Telang 1998; Peek et al. 1998; Spaulding and von Dohlen 1998; Chen, Li, and Aksoy 1999). Calibrating rates of sequence evolution using these groups could provide anchors for dating divergences in other groups of bacteria.

A second major requirement for the valid use of molecular clocks is that rates of change be demonstrated to be roughly constant among lineages and over time. When divergences are not too great, neutral sites may be more clocklike because they are expected to evolve at a rate that is constant both among lineages over time and among regions of the genome. This is because absolute rates of substitution of neutral alleles are affected only by the mutation rate and generation time, and not by variation in population size or by site-specific levels of selective constraint (Kimura 1968). Most of a bacterial genome consists of low-expression protein-coding genes, in which silent sites may be selectively neutral or close to neutral. If so, silent sites of most genes have the potential to serve as molecular chronometers for estimating the divergence times of many groups of bacteria. However, silent sites evolve too quickly to be use-

Key words: aphids, *Buchnera*, endosymbiosis, evolutionary rate, molecular clock, mutational bias.

Address for correspondence and reprints: Nancy A. Moran, Department of Ecology and Evolutionary Biology, Unversity of Arizona, Tucson, Arizona 85721. E-mail: nmoran@u.arizona.edu.

Mol. Biol. Evol. 16(11):1586–1598. 1999

ful for estimating numbers of substitutions across deep divergences; rRNA sequences, replacement sites, or amino acid sequences must be used in these cases. Thus, to determine the utility of molecular clocks for prokaryotes, the extent of rate homogeneity among genes for different categories of sites and genes should be established.

Evidence to date for *Buchnera*, the endosymbionts of aphids, suggests that their intracellular habitat prevents recombination or gene movement between symbionts from different aphids, even for plasmid-encoded genes (Baumann, Moran, and Baumann 1997; Rouhbakhsh et al. 1996, 1997). Thus, divergences at different endosymbiont loci have occurred over the same time interval, beginning when host lineages split. In contrast, loci of free-living bacteria may have heterogeneous histories due to transfer of genes among lineages (e.g., Lawrence and Ochman 1997, 1998) or due to recombination (e.g., Bisercic, Feutrier, and Reeves 1991; Guttman and Dykhuizen 1994).

In this paper, we present analyses of mostly newly determined DNA sequences from *Buchnera* of four aphid species. The loci analyzed include genes for enzymes involved in biosynthesis of amino acids for host nutrition, as well as protein-coding genes for other functions and rRNA genes. The aphid hosts belong to two pairs, from each of two aphid families. Fossil-based evidence supports a similar date (50–70 MYA) for the ancestor of each pair, allowing absolute rates of substitution to be calibrated. We address whether rates of synonymous and nonsynonymous substitutions are similar among lineages and whether rates of synonymous substitutions are similar among loci. *Buchnera* lineages diverged over a time frame comparable to that for the divergence of *Escherichia coli* and *Salmonella typhimurium*, and we compare rates and patterns of sequence evolution in *Buchnera* versus these enterics. Finally, *Buchnera* is one of numerous intracellular bacteria that show strong base-compositional bias favoring A+T; we examine the phylogenetic pattern of amino acid replacements connected with the evolution of A+T bias.

Materials and Methods

Determination of Ancestral Dates

The phylogeny of the study taxa is given in figure 1. *Buchnera* strains were presumed to have diverged synchronously with aphid hosts, an assumption supported by their concordant phylogenies (Moran et al. 1993). Aphid collections were from the same localities and sites reported in Moran et al. (1993). Host divergence dates (fig. 1) were estimated as follows.

Melaphis rhois and *Schlechtendalia chinensis* are in the family Pemphigidae, subfamily Fordinae, subtribe Melaphidina (Blackman and Eastop 1994). The minimum age for the split between *M. rhois* and *S. chinensis* (fig. 1) is based on their highly derived association with host plants in the genus *Rhus*, on which members of the Melaphidina form complex galls. This host plant association is presumed to have originated in a common ancestor before the Asian and American ranges of *Rhus* were highly separated, since the aphids are now present on both continents (Moran 1989). The disjunction in the host plant ranges occurred as northern range limits moved south from Beringia as climates cooled in the early Tertiary. This was about 50 MYA, as indicated by fossil *Rhus* leaves (Moran 1989). The maximum age of 70 Myr is set prior to the first fossil records of the genus *Rhus* (see Moran et al. 1993).

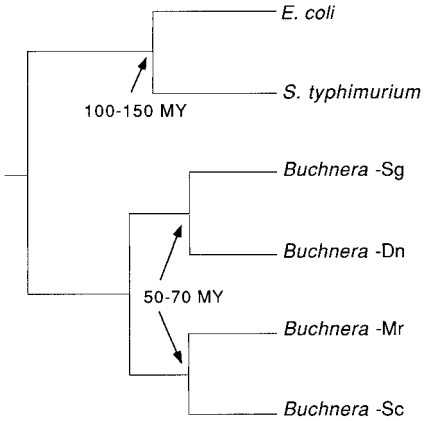

FIG. 1.—Phylogenetic relationships and estimated dates of divergence of *Buchnera* study taxa and enterics.

The common ancestor of *Schizaphis graminum* and *Diuraphis noxia* corresponds to the ancestor of the subtribes Rhopalosiphina and Macrosiphina within the family Aphididae. Although this family was probably present in the upper Cretaceous (80 MYA) as *Aphidocallis caudata* Kononova (Heie 1987; O. E. Heie, personal communication), the first fossils assignable to the two groups are from lower Tertiary. *Diatomyzus* in the diatomaceous earth formation of Denmark (50 My) probably belongs to Rhopalosiphina, and *Pseudoamphorophora* in Baltic amber (38–45 My) probably belongs to Macrosiphina (Heie 1987; O. E. Heie, personal communication). In addition, the shared ancestor of the Rhopalosiphina and Macrosiphina probably occurred on a plant in the Rosaceae, for which definitive fossils are not present until the early Tertiary. Thus, the ancestor of *S. graminum* and *D. noxia* must have been in the early Tertiary or, conservatively, the very late Cretaceous, placing the date at 50–70 MYA. Although some systematic schemes place *S. graminum* and *D. noxia* in separate tribes or subfamilies rather than in separate subtribes (Blackman and Eastop 1984), the same reasoning and estimated dates would apply for this divergence.

Escherichia coli and *Salmonella enterica* var. *typhimurium* are members of the Enterobacteriaceae. Their divergence is estimated at 100–150 MYA based on the approximate time of the increase in abundance of mammals during the Cretaceous. *Escherichia coli* occurs primarily in mammalian hosts and is believed to have diverged from *Salmonella* when mammals became abundant hosts (Ochman and Wilson 1987). We note that a later or earlier divergence cannot be absolutely excluded, because the divergence of *E. coli* could have occurred before or after its colonization of its usual group of modern hosts.

Table 1
Function and Location of Genes Studied, Grouped by Linkage Group (See Fig. 2)

Gene	Function	Genome Location in *Buchnera*
ilvDC	Amino acid biosynthesis	Chromosome
ilvIH	Amino acid biosynthesis	Chromosome
hisGD	Amino acid biosynthesis	Chromosome
trpDC(F)BA	Amino acid biosynthesis	Chromosome
trpEG	Amino acid biosynthesis	*trpEG* plasmid in *Buchnera*-Sg and *Buchnera*-Dn, chromosome in *Buchnera*-Mr and *Buchnera*-Sc
leuABCD	Amino acid biosynthesis	*leu* plasmid in *Buchnera*-Sg and *Buchnera*-Dn (not sequenced for *Buchnera*-Mr and *Buchnera*-Sc)
ORF1	Unknown	*leu* plasmid in *Buchnera*-Sg and *Buchnera*-Dn (not sequenced for *Buchnera*-Mr and *Buchnera*-Sc)
repA1, repA2	Plasmid replication	*leu* plasmid in *Buchnera*-Sg and *Buchnera*-Dn (not sequenced for *Buchnera*-Mr and *Buchnera*-Sc)
ftsZ	Cell division	Chromosome
gnd	Energy metabolism	Chromosome
atpAGD	Energy metabolism	Chromosome
16S rDNA	Ribosomal RNA	Chromosome
23S rDNA	Ribosomal RNA	Chromosome

Estimation of Numbers of Generations per Year in Nature

The number of *Buchnera* replications per year can be estimated as the product of the number of replications per aphid generation and the number of aphid generations per year. Based on a quantitative PCR study of the *Buchnera* of *S. graminum* (hereafter *Buchnera*-Sg), a 1-day-old female aphid has 0.2×10^6 copies of the *Buchnera* chromosome, while a mature, reproducing female aphid has 5.6×10^6 copies (Baumann and Baumann 1994). There is little degeneration of *Buchnera* cells during the growth of a young aphid, implying that the increase in numbers can serve as an index of the average number of replications. Thus, there are approximately five or six doublings during the life span of a single aphid generation. Depending on the aphid life cycle, the annual number of generations is 3–10 per year (Moran 1992). The Melaphidina, containing *M. rhois* and *S. chinensis*, share similar life cycles that contain about five generations per year (Takada 1991; unpublished data). The Aphididae, containing *S. graminum* and *D. noxia*, have faster generation times, with 5–12 overlapping generations each year. We used eight as a likely average for the lineages ancestral to *S. graminum* and *D. noxia*. Thus, *Buchnera* has approximately 30–50 doublings per year, with the *Buchnera* of Melaphidina at the low end of this range and the *Buchnera* of Aphididae at the high end.

Enteric bacteria are estimated to have about 100 generations per year in nature, based on studies of *E. coli* replication rates in naturally occurring hosts and other habitats (Gibbons and Kapsimalis 1967; Savageau 1983).

Choice of Genes

Buchnera is unusual in having a very reduced genome (Charles and Ishikawa 1999) but still retaining genes for amino acid biosynthesis (Baumann et al. 1995, 1997); this feature presumably reflects the functional contribution of the symbiont to its host. The set of genes chosen for study was divided between loci that function in the overproduction of amino acids for the nutritional benefit of the host aphids and ''housekeeping'' genes that function in the growth and replication of the bacterial cells (table 1). Some of the loci involved in amino acid biosynthesis are located on plasmids in *Buchnera* of the Aphididae (including *Buchnera*-Sg and *Buchnera*-Dn and their common ancestor): *leuABCD* is on a plasmid of the IncFII type (Bracho et al. 1995; van Ham, Moya, and Latorre 1997; Silva et al. 1998; Baumann et al. 1999), and *trpEG* is on a plasmid consisting of tandem repeats of these loci and an origin of replication containing dnaA boxes (Lai, Baumann, and Baumann 1994). In both cases, molecular and phylogenetic evidence supports a history of strictly vertical transmission of the plasmids, with the plasmid-borne biosynthetic genes originating from the chromosome (Baumann et al. 1999; Clark, Baumann, and Moran 1999). The homologous *trp* genes are not plasmid-borne in *Buchnera*-Mr and *Buchnera*-Sc (Lai, Baumann, and Moran 1995). In addition, for *Buchnera*-Sg and *Buchnera*-Dn, the three other genes from the *leu* plasmid were included: two closely related loci involved in plasmid replication (*repA1* and *repA2*) and a gene (ORF1) that is homologous to an *E. coli* gene of unknown function.

Laboratory Methods

Most of the laboratory methods used in this study have been described in detail in our past publications (Lai, Baumann, and Baumann 1994; Lai, Baumann, and Moran 1996; Clark, Baumann, and Baumann 1998*a*); only an outline will be presented here. These procedures include purification of whole aphid DNA (containing both *Buchnera* and host genomes), restriction enzyme and Southern blot analyses, amplification of DNA by the polymerase chain reaction (PCR), and nucleotide (nt) sequence determination. PCR products were treated with restriction enzymes which digested the sites on the synthetic oligonucleotide primers and cloned into p-BCKS/SK (Stratagene, La Jolla, Calif.) as previously

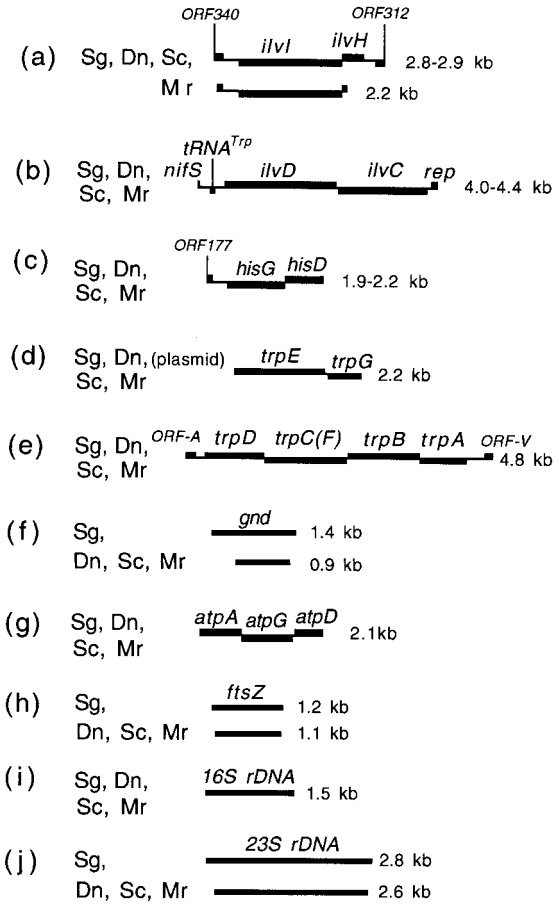

FIG. 2.—Genetic maps of the DNA fragments containing the genes compared in this study.

described. Unless otherwise noted, standard molecular biology methods were used (Sambrook, Fritsch, and Maniatis 1989).

ilvIH-*Containing DNA Fragments*

Previously, we obtained a 0.7-kb *ilvI* probe derived from *Buchnera*-Sg (Thao and Baumann 1998). Restriction enzyme and Southern blot analyses were performed using this probe with *Buchnera*-Dn and *Buchnera*-Sc DNA. The results indicated the presence of a 9.9-kb *Eco*RI DNA fragment from *Buchnera*-Dn and a 7.8-kb *Eco*RI fragment from *Buchnera*-Sc. Following restriction enzyme digestion and electrophoresis, these fragments were electroeluted and ligated into appropriately digested λZAP (Stratagene), and the 0.7-kb *ilvI* probe was used for the detection of recombinants (Lai, Baumann, and Moran 1996; Clark, Baumann, and Baumann 1998*a*). The nt sequence was determined from the end of ORF340 in both fragments and to the beginning of ORF312 in the *Buchnera*-Sc fragment (fig. 2*a*). The *Buchnera*-Dn fragment lacked the last 58 amino acids of IlvI. Based on the *Buchnera*-Dn *ilvI* sequence and the amino acid sequence of ORF312 in *Buchnera*-Sg, *E. coli* (1786270), and *Haemophilus influenzae* (1175483), the following oligonucleotides were designed: (*Bam*HI) 5′-CGC GGA TCC GGT ATG GTT AAA CAA TGG C-3′ and (*Sal*I) 5′-CGC GTC GAC GWC CWC CCA TWC CAA AWG TWC CAT C-3pr. The 1.0-kb DNA fragment was amplified from *Buchnera*-Dn DNA by PCR and sequenced to the beginning of the ORF312 gene.

Based on the sequences of ORF340 and *ilvH* of *Buchnera*-Sg, *Buchnera*-Dn, and *Buchnera*-Sc, the following oligonucleotide primers were designed: (*Bam*HI) 5′-GAC GGA TCC GAY GSW TGG GAT WSM GSA TTA MTK ATH WTR GG-3′ and (*Eco*RI) 5′-GAC GAA TTC CTA YDA CTC KTG AYA RWG CWC CWG ATT C-3′. These oligonucleotides correspond to amino acids 290–300 of ORF340 and amino acids 21–12 of IlvH of *Buchnera*-Sg (Thao and Baumann 1998). Using these primers, *Buchnera*-Mr DNA, and PCR, a 2.2-kb DNA fragment was amplified (fig. 2*a*).

ilvDC-*Containing DNA Fragments*

Based on the nt sequence of *Buchnera*-Sg *ilvDC* (Clark, Baumann, and Baumann 1998*b*), the following oligonucleotide primers were designed: 5′-GAG TGC AGA TAA AGG TGC-3′ (amino acids 601–607 of IlvD) and 5′-CTC TCA CTT CTG TTC CAG G-3′ (amino acids 162–156 of IlvC). Using *Buchnera*-Sg DNA, a 0.6-kb DNA fragment was amplified. Restriction enzyme and Southern blot analyses of *Buchnera*-Dn DNA indicated the presence of a 7.3-kb *Eco*RI DNA fragment which hybridized with the probe. This fragment was cloned into λZAP (Stratagene) as described above.

Based on the amino acid sequences of NifS and Rep from *Buchnera*-Sg, *Buchnera*-Dn, *E. coli* (1788879, 1790212), and *H. influenzae* (2492875, 1172905), the following oligonucleotide primers were designed: (*Bam*HI) 5′-CGC GGA TCC ATW CCW CKT TCA TGW CCW CCW CCA TG-3′ and (*Sal*I) 5′-CGC GTC GAC CWG CCC CWG CWA RWA TWA RAC AGG GWC C-3′). These primers correspond to amino acids 239–231 of *Buchnera*-Sg NifS and amino acids 25–16 of Rep (Clark, Baumann, and Baumann 1998*a*). Using these primers and DNA from *Buchnera*-Sc and *Buchnera*-Mr, 4.7-kb DNA fragments were amplified by PCR. For all recombinants, only the nt sequence of the region from *nifS* to *rep* was determined (fig. 2*b*).

hisGD-*Containing DNA Fragments*

Based on the amino acid sequences of AroC and HisD from *Buchnera*-Sg, *E. coli* (1788669, 1788331), and *H. influenzae* (1168504, 1170289), the following oligonucleotide primers were designed: (*Sac*I) 5′-CGC GAG CTC CWG CDG CWA CWC KCA TWG CWG TTT CWC KWG C-3′ and (*Bam*HI) 5′-CGC GGA TCC GCA TTW CCW GGW CCA AAW ATT TTA TCW ACT TTW GG-3′. These primers correspond to amino acids 138–128 of *Buchnera*-Sg AroC and amino acids 213–202 of HisD (Clark, Baumann, and Baumann 1998*b*). Using these primers and DNA from *Buchnera*-Dn and *Buchnera*-Sc, 3.0–3.3-kb DNA fragments were amplified by PCR. In the case of *Buchnera*-Mr, the *Bam*HI site on the oligonucleotide primer was replaced by *Eco*RI, since the former restriction site was present on the amplified fragment. In all three fragments, the

sequence from the end of *ORF177* to *hisD* was determined (fig. 2*c*).

Buchnera-*Mr* trpEG- *and* trpD(C)FBA-*Containing DNA Fragments*

We previously obtained the sequences of *trpEG* and *trpD(C)FBA* for *Buchnera*-Sg, *Buchnera*-Dn, and *Buchnera*-Sc (fig. 2*d*; Lai, Baumann, and Baumann 1994; Lai, Baumann, and Moran 1995; Baumann, Baumann, and Moran 1998). The methods used for cloning a 4.0-kb *Buchnera*-Mr *Eco*RI DNA fragment containing *trpEG* were identical to those previously described for *Buchnera*-Sc *trpEG* (Lai, Baumann, and Moran 1995). Using a probe for *trpB,* an 8.0-kb *Buchnera*-Mr *Eco*RI DNA fragment was cloned as described. Sequence determination indicated that one of the *Eco*RI sites was within *trpB.* Based on the amino acid sequences of *Buchnera*-Sg, *Buchnera*-Dn, and *Buchnera*-Sc, the following oligonucleotide primer was designed: (*Kpn*I, *Bam*HI) 5′-GGC GGT ACC GGA TCC GCW GAA ACW GGW GCW GGW CAA CAT GG-3′ (amino acids 107–115 of *Buchnera*-Sg *trpB*). In addition, based on the amino acid sequences of *Buchnera*-Sg, *E. coli,* and *H. influenzae* hypothetical protein P14 (1787506 and 1175670 for the last two), the following oligonucleotide primer was designed: (*Pst*I) 5′-GCG CTG CAG CCA WCC WCC AAA WAT ATC WCC ATT WGC-3′ (amino acids 45–37 of *Buchnera*-Sg P14). Using these two primers, a 2.1-kb fragment was obtained by PCR, digested with *Kpn*I and *Pst*I, and cloned into the plasmid vector. Only the sequence indicated in figure 2*e* was obtained.

gnd-*Containing DNA Fragments*

Based on the amino acid sequences of Gnd from *Buchnera*-Sg, *E. coli* (1788341), *H. influenzae* (1168235), and *Bacillus subtilis* (2636555), the following oligonucleotide primers were madedesigned: (*Bam*HI) 5′-CGC GGA TCC GGW CCW WSW ATW ATG CCW GGW GG-3′′ and (*Apa*I) 5′-CGC GGG CCC GTA TGW GCW CCA AAA TAA TCW CKT TGW GCT TG-3′. These primers correspond to amino acids 137–144 and 453–443 of *Buchnera*-Sg Gnd. Using these primers and DNA from *Buchnera*-Dn, *Buchnera*-Sc, and *Buchnera*-Mr, 0.9-kb DNA fragments were amplified by PCR (fig. 2*f*).

atpA-atpG-atpD-*Containing DNA Fragments*

Based on comparisons of amino acid sequences of AtpA and AtpD from *Buchnera*-Sg, *E. coli* (1790172, 1790170, and *H. influenzae* (1168566, 2506208), the following oligonucleotide primers were designed: (*Bam*HI) 5′-CGC GGA TCC WCC WGG WMG WGA AGC WTT TCC WGG WG-3′ and (*Sac*I) 5′-CGC GAG CTC CWG CWC CWC CAA AWA RWC CWA CTT TWC CWC C-3′. These primers correspond to amino acids 280–289 of *Buchnera*-Sg AtpA and amino acids 159–149 of AtpD. Amplification by PCR resulted in 2.1-kb DNA fragments. In the case of *Buchnera*-Dn, the *Bam*HI site of the first oligonucleotide primer was replaced by *Pst*I.

ftsZ-*Containing DNA Fragments*

Based on comparisons of FtsZ from *Buchnera*-Sg, *E. coli* (1786284), *H. influenzae* (1169767), and *Pseudomonas aeruginosa* (1346052), the following oligonucleotide primer was designed: (*Kpn*I, *Sac*I) 5′-CGC GGT ACC GAG CTC GGW GTW GGW GGW GGW GGW GG-3′. This primer corresponds to amino acids 17–23 of *Buchnera*-Sg FtsZ (Baumann and Baumann 1998). The oligonucleotide primer given in Baumann and Baumann (1998) corresponding to amino acids 379–374 of *Buchnera*-Sg was also used (*Bam*HI, *Sal*I). The resulting PCR product was about 1.1 kb in size (fig. 2*h*).

23S rDNA

Buchnera-Dn, *Buchnera*-Sc, and *Buchnera*-Mr DNA fragments containing most of the 23S rDNA gene were obtained by PCR using the oligonucleotide primers B1 and C2 described by Rouhbakhsh and Baumann (1995).

Nucleotide Sequence Accession Numbers

The accession numbers of the new *Buchnera* nt sequences, as well as those of additional sequences used in the analyses, are given in table 2. Some sequences for *S. typhimurium* were obtained from BLAST similarity searches of sequence contigs of the *S. typhimurium* genome project, available through the Genome Sequencing Center, Washington University (http://genome.wustl.edu/gsc/bacterial/bacterial_blast_server.html).

Sequence Analyses

Divergence Estimation

For each coding gene, sequences were translated and then aligned using Pileup of the GCG package (Genetics Computer Group 1997). DNA alignments were then fit to the amino acid alignments. Alignments for rDNA genes were obtained using Pileup on the DNA sequences. For both protein-coding and rDNA genes, alignments were largely unambiguous. Divergences for silent and replacement sites of coding genes were calculated using the ''Diverge'' command in GCG, which implements the algorithm of Li (1993*a*) with modifications of Pamilo and Bianchi (1993). Divergences for rDNA genes were estimated based on the Kimura (1980) two-parameter model of substitution, using the ''Distances'' command of GCG.

In order to examine the effect of A+T bias on amino acid replacements in *Buchnera* over its evolutionary history, amino acid changes were mapped onto the known phylogeny containing the four *Buchnera* taxa, the two enterics, and the outgroup *H. influenzae.* For the set of 14 chromosomally encoded protein-coding genes available for all four of the *Buchnera* taxa (*ilvDC, ilvI, hisGD, trpDC(F)BA, gnd, atpAGD,* and *ftsZ*), unambiguous amino acid changes were mapped onto each branch based on parsimony using MacClade 3.0 (Maddison and Maddison 1992). Parsimony was chosen because the level of homoplasy for the amino acid tree was low, indicating that the number of multiple changes at a site was small. The consistency index (number of changes

Table 2
GenBank Accession Numbers of the Sequences Used in this Study

Gene(s)	Sg	Dn	Sc	Mr	*Escherichia coli*	*Salmonella typhimurium*
ilvI	AF060492	*AF129501	*AF129502	*AF129503	AE000690	^
ilvH	AF060492	*AF129501	*AF129502		AE000118	X55456
ilvD-ilvC	AF008210	*AF130812	*AF130813	*AF130814	AE000453, AE000454	^
hisG-hisD	AF067228	*AF129281	*AF129282	*AF129283	AE000293	J01804
trpE-trpG	Z21938	L46769	U09184	*AF130815	AE000224	V01378
trpD-trpC(F)-trpB-trpA	Z19055	AF038565	U09185	*AF132318	AE000224	V01376-8, M30285, M30286
gnd	AF067228	*AF129278	*AF129279	*AF129280	AE000294	U14495
atpA-atpG-atpD	AF008210	*AF129402	*AF129403	*AF129404	AE000450	^
ftsZ	AF012886	*AF130816	*AF130817	*AF130818	AE000119	^
leuA-leuB-leuC-leuD-repA1-ORF1-*repA2*	AF041836	AF041837			AE000117, AE000382	X51583, X53376, X51476, U14495
16S rDNA	M63246	M63251	Z19056	M63255	AE000460	U90316
23S rDNA	U09230	*AF129498	*AF129499	*AF129500	AE000461	U65886

NOTE.—Asterisks indicate sequences obtained in this study. Carets indicate sequences for *S. typhimurium* obtained from BLAST searches of sequence contigs available through the Genome Sequencing Center, Washington University (see text).

if no homoplasy/numbers of changes in actual tree, as calculated in PAUP [Swofford 1999]) for the amino acid tree was 0.94 for the overall tree and 0.98 for the tree containing only *Buchnera*. Every mapped amino acid change was categorized as increasing A+T, decreasing A+T, or having no effect on A+T by the following procedure: Amino acids were first grouped according to the number of A+T in the corresponding codons: 0–1, 0–2, 1–2, 2–3, 2 (methionine only), or 1 (tryptophan only). Any change that results in a higher or lower mean number of amino acids in the codon family was counted as a change increasing or decreasing A+T, respectively; for example, a change from proline (0–1 A+T) to histidine (1–2 A+T) was counted as a change increasing A+T. The difference between the number of changes increasing A+T and the number of changes decreasing A+T was calculated for each branch and mapped onto the phylogeny. Changes involving arginine (0–2 A+T) were not counted as changing AT content, since the direction of the effect could not be reconstructed.

To examine possible effects of codon bias on silent divergences, codon adaptation indices (CAIs; Sharp and Li 1987*b*) were calculated for the all *E. coli* genes for which *E. coli–S. typhimurium* divergences were obtained (some genes were not available for *S. typhimurium*). CAI values were based on *E. coli* codon use and were calculated using the Molecular Evolutionary Analysis Package (November 1997 version) of E. Moriyama.

Results and Discussion

Rates of Silent Substitution in *Buchnera* Versus Enterics

Divergences for all coding genes are shown in table 3, and average rates of divergence are shown in table 4. Silent divergence values (K_s) are remarkably consistent among loci for both *Buchnera* comparisons. For *Buchnera*-Mr– *Buchnera*-Sc, K_s values for all 16 coding loci are within one standard deviation of 0.73, the K_s value obtained across loci weighted by the number of synonymous sites per locus; this implies that the genomewide divergence is consistent across loci. The plasmid-borne *trpE* and *trpG* in Aphididae have already been noted to show higher silent rates of substitution than other loci (Rouhbakhsh et al. 1997), and this pattern is observed in our data for *Buchnera*-Sg–*Buchnera*-Dn. If we consider all other coding genes, only one out of the total of 22 loci shows a K_s value more than one standard deviation from 0.92, the weighted K_s value for all loci excluding *trpEG* (no more than expected by chance). This implies that the mutation rate is the same for different loci within a given *Buchnera* lineage, with the exception of the plasmid-borne *trpEG* of *Buchnera*-Sg–*Buchnera*-Dn. In *Buchnera*-Mr–*Buchnera*-Sc, *trpEG* is not plasmid-borne and does not show a higher synonymous divergence (table 3).

In contrast, synonymous divergences are more heterogeneous among loci for *E. coli–S. typhimurium* than for *Buchnera* pairs. For *E. coli–S. typhimurium,* 9 of 16 loci are more than one standard deviation from the overall K_s value of 0.84. Two factors that may contribute to this contrast are the presence of purifying selection on synonymous sites in enterics (Sharp and Li 1987*a*; Eyre-walker and Bulmer 1995) but not in *Buchnera* (Brynnel et al. 1998; Wernegreen and Moran 1999), and recombination in enterics but not in *Buchnera*. Previous studies have shown that silent divergences for *E. coli–S. typhimurium* decline with increasing CAI (e.g., Sharp and Li 1987*a*; Sharp 1991), indicating the presence of purifying selection at synonymous sites. To examine the role of selection on codon use for this set of loci, we plotted K_s values with respect to CAI (based on *E. coli*) for all three pairs of taxa (fig. 3). As in previous studies, K_s values for the *E. coli–S. typhimurium* comparison show a general decline with increasing CAI. An exception to this trend is the locus *gnd,* which has a K_s value higher than expected on the basis of its high CAI. This locus has previously been shown to undergo unusual rates of intragenic recombination in enterics due to its linkage with genes for surface antigens under diversifying selection (Bisercic, Feutrier, and Reeves 1991; Nelson and Selander 1994). The relationship between CAI and K_s implies that the average K_s value for *E. coli–S. typhimurium* (table 3) is lower than the divergence expected due to mutation at neutral sites. The K_s values

Table 3
Synonymous (K_s) and Nonsynonymous (K_a) Divergences (as Substitutions per Site) for Pairwise Comparisons Among *Buchnera*-Sg–*Buchnera*-Dn, *Buchnera*-Mr–*Buchnera*-Sc, and *Escherichia coli–Salmonella typhimurium*

		Buchnera-Sg–*Buchnera*-Dn					*Buchnera*-Mr–*Buchnera*-Sc					*E. coli–S. typhimurium*				
Gene	Codon	K_s	SD	K_a	SD	K_s/K_a	K_s	SD	K_a	SD	K_s/K_a	K_s	SD	K_a	SD	K_s/K_a
ilvD[a]	617	1.04	0.21	0.15	0.02	6.9	0.76	0.12	0.13	0.02	5.8	0.70	0.12	0.12	0.03	5.8
ilvC	491	0.95	0.23	0.16	0.02	5.9	0.74	0.13	0.13	0.02	5.7	0.40	0.06	0.02	0.01	20.0
ilvI	573	0.73	0.12	0.11	0.02	6.6	0.73	0.44	0.08	0.02	9.1	1.03	0.17	0.06	0.02	17.2
ilvH	159	0.80	0.20	0.11	0.03	7.3	n/a[b]					1.06	0.23	0.02	0.01	53.0
hisD	201	1.31	0.79	0.21	0.04	6.2	0.73	0.17	0.31	0.04	2.4	1.01	0.19	0.07	0.07	14.4
hisG	299	0.70	0.63	0.07	0.02	10.0	0.69	0.12	0.07	0.02	9.9	0.59	0.09	0.03	0.01	19.7
trpA	268	0.92	0.23	0.23	0.03	4.0	0.66	0.15	0.18	0.03	3.7	1.44	0.27	0.09	0.02	16.0
trpB	397	0.86	0.22	0.09	0.02	9.6	0.78	0.13	0.10	0.02	7.8	0.85	0.11	0.02	0.01	42.5
trpC(F)	461	1.01	0.21	0.21	0.03	4.8	0.75	0.12	0.21	0.02	3.6	1.11	0.15	0.08	0.01	13.9
trpD	334	0.82	0.18	0.12	0.03	3.9	0.79	0.16	0.19	0.03	4.2	0.80	0.11	0.02	0.01	40.0
trpE	519	1.56	0.47	0.18	0.02	8.7	0.69	0.10	0.13	0.02	5.3	0.71	0.10	0.03	0.01	23.7
trpG	201	1.12	0.25	0.14	0.03	8.0	0.71	0.20	0.20	0.03	3.6	n/a[c]				
leuA	518	0.91	0.20	0.08	0.01	11.4	n/a[b]					0.93	0.12	0.04	0.01	23.3
leuB	365	1.16	0.47	0.17	0.02	6.8	n/a[b]					0.85	0.15	0.02	0.02	42.5
leuC	468	0.90	0.19	0.13	0.02	6.9	n/a[b]					0.71	0.09	0.04	0.01	17.8
leuD	207	1.02	0.36	0.26	0.04	3.9	n/a[b]					0.82	0.15	0.04	0.02	20.5
repA1	279	0.89	0.19	0.10	0.02	8.9	n/a[b]					n/a[c]				
repA2	279	0.79	0.21	0.13	0.03	6.1	n/a[b]					n/a[c]				
ORF1[a]	167	0.97	0.31	0.04	0.01	24.3	n/a[b]					1.14	0.27	0.04	0.02	28.5
gnd	297	1.11	0.32	0.18	0.03	6.2	0.74	0.14	0.15	0.03	4.9	0.99	0.17	0.02	0.01	49.5
atpA	224	0.74	0.19	0.14	0.03	5.3	0.66	0.15	0.17	0.03	3.9	n/a[c]				
atpG[a]	291	0.90	0.24	0.19	0.03	4.7	0.71	0.17	0.27	0.03	2.6	0.26	0.06	0.02	0.01	13.0
atpD[a]	148	0.84	0.25	0.11	0.03	7.6	0.54	0.15	0.13	0.03	4.2	0.27	0.06	0.03	0.01	9.0
ftsZ	352	0.78	0.16	0.06	0.01	13.0	0.84	0.14	0.15	0.02	5.6	0.71	0.10	0.03	0.01	23.7
Average		0.95	0.20	0.14	0.06	7.0	0.72	0.07	0.16	0.06	5.1[d]	0.78[d]	0.25	0.04	0.03	24.7

[a] Shorter sequence for *E. coli–S. typhimurium* comparison.
[b] Sequences not obtained for *Buchnera*-Mr–*Buchnera*-Sc.
[c] Sequence not available for *S. typhimurium.*
[d] Only genes available for *Buchnera*-Mr and *Buchnera*-Sc included in average.

that correspond to loci with low CAIs (<0.33) were used as representative of the intrinsic mutation rate for comparisons with *Buchnera.* In contrast to the comparisons for *E. coli–S. typhimurium, K_s* values for the pairs of *Buchnera* show no relationship to CAI as calculated for *E. coli* (fig. 3). Since the variation among K_s values for different loci is not greater than that expected due to chance, and since *Buchnera* lacks codon bias beyond that resulting from biased base composition (Wernegreen and Moran 1999), we conclude that selection on codon use is not affecting K_s values in *Buchnera.*

When the two pairs of *Buchnera* are compared, *Buchnera*-Sg–*Buchnera*-Dn, shows greater divergence than does *Buchnera*-Mr–*Buchnera*-Sc at 12 of 14 loci (sign test, $P < 0.01$). Thus, if the divergence dates are the same for the two pairs, rates of neutral substitution averaged 30% higher for *Buchnera*-Sg–*Buchnera*-Dn. A possible basis for the difference is the greater number of generations elapsed for *Buchnera*-Sg–*Buchnera*-Dn than for *Buchnera*-Mr–*Buchnera*-Sc due to typical life cycle differences between the two families of aphids. Factoring in the approximate generation times gives similar synonymous-rate estimates for the two pairs (table 4).

On average, synonymous divergences for *E. coli–S. typhimurium* are similar to those for the comparisons

Table 4
Absolute and Per-Generation Rates of Substitution Among *Buchnera*-Sg–*Buchnera*-Dn, *Buchnera*-Mr–*Buchnera*-Sc, and *Escherichia coli–Salmonella typhimurium* for Synonymous Sites and Nonsynonymous Sites Averaged Over Protein-Coding Loci and for 16S rDNA and 23S rDNA

	Buchnera-Sg–*Buchnera*-Dn		*Buchnera*-Mr–*Buchnera*-Sc		*E. coli–S. typhimurium*	
	Absolute (s/s/100 Myr)	Generation (s/s/billion generations)	Absolute (s/s/100 Myr)	Generation (s/s/billion generations)	Absolute (s/s/100 Myr)	Generation (s/s/billion generations)
K_s	0.68–0.95	0.14–0.19	0.51–0.72	0.17–0.24	0.36–0.54[a]	0.04–0.05
K_a	0.10–0.14	0.02–0.03	0.11–0.16	0.04–0.05	0.01–0.02	0.001–0.002
16S rRNA	0.019–0.027	0.004–0.005	0.032–0.045	0.011–0.015	0.012–0.018	0.001–0.002
23S rRNA	0.030–0.042	0.006–0.008	0.038–0.054	0.013–0.018	0.007–0.011	0.001–0.001

Note.—Upper and lower bounds correspond to the lower and upper estimates of the dates of divergence for each pair. Generations per year were set at 50, 30, and 100 for *Buchnera*-Sg–*Buchnera*-Dn, *Buchnera*-Mr–*Buchnera*-Sc, and *E. coli–S. typhimurium,* respectively (see text). ss = substitutions per site.
[a] Based on average K_s for the four loci with the lowest CAI value.

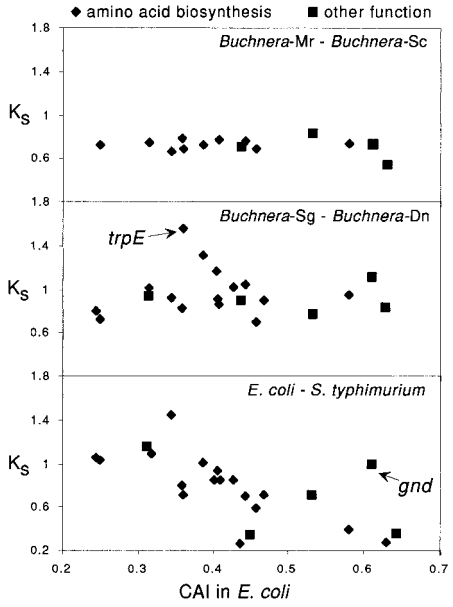

FIG. 3.—Codon adapation index (CAI) versus K_s for *Escherichia coli–Salmonella typhimurium* and the two pairs of *Buchnera.* In *Buchnera*-Sg–*Buchnera*-Dn, *trpE* is on a plasmid with a higher mutation rate.

within *Buchnera,* with the average K_s lying between the values for for the two *Buchnera* pairs. However, in the enterics, the K_s for loci with low codon bias (1.06) is more representative of the neutral divergence and is higher than the average for either pair of *Buchnera* (0.95 and 0.72). If we accept the date of 100–150 MYA for the divergence of *E. coli* from *S. typhimurium* (Ochman and Wilson 1987), this implies that rates of neutral substitution on an absolute timescale are lower for enterics than for the *Buchnera* pairs (table 4) and that the underlying mutation rate is about twofold higher in *Buchnera*. The rate difference is more pronounced on a per-generation basis. If the enterics have about 100 divisions per year (Gibbons and Kapsimalis 1967; Savageau 1983), then their divergence has occurred over about 1×10^{10} generations, whereas those for *Buchnera*-Sg–*Buchnera*-Dn and *Buchnera*-Mr–*Buchnera*-Sc have occurred over about 3×10^9 and 1.5×10^9 generations, respectively. Based on comparison with the *E. coli* loci with lowest codon bias, *Buchnera* shows an average mutation rate that is approximately fourfold higher per generation (table 4). This difference would be greater if lineages leading to *E. coli* and *S. typhimurium* underwent more than 100 generations per year.

A possible source of error in the *Buchnera* K_s values is the strong bias in base composition at silent codon positions in *Buchnera*. The Li (1993*a*) method used to calculate K_s values incorporates a Kimura (1980) two-parameter model to estimate divergences at silent sites; this model may not give accurate divergence estimates under strong bias in base composition. To examine the possible magnitude of error, we compared different methods of estimating divergence for third codon positions (a set that shows large overlap with the set of silent sites). Using algorithms incorporated into PAUP (version 4.0b2; Swofford 1999), values from the Kimura (1980) two-parameter model were contrasted with values from maximum- likelihood models (HKY85) that incorporated empirical base compositions within enterics or within *Buchnera*. For *Buchnera* and *E. coli–S. typhimurium*, respectively, the Kimura (1980) two-parameter model gave divergence estimates that were 60%–79% and 91%–109% of the estimates based on maximum likelihood and empirical base frequencies (based on the low-CAI set of loci in *E. coli*). Thus, the K_s values calculated for *Buchnera* may underestimate the actual divergence rate at silent sites, implying that the difference in actual substitution rate may be greater than that presented in table 4.

Nonsynonymous Divergences in *Buchnera* and Enterics

For all loci, K_s/K_a ratios are much lower for *Buchnera* than for enterics. This difference is consistent with previous observations, some of which were based on deeper divergences that were near saturation for fewer loci (Moran 1996; Wernegreen and Moran 1999). Whereas *E. coli–S. typhimurium* have an average K_s/K_a of 20.2 for these loci, the two *Buchnera* pairs have average K_s/K_a values of 7.0 and 5.1 (table 3). The simplest explanation for this pattern, which is consistent across all loci, is that *Buchnera* show an elevated rate of nonsynonymous substitution due to decreased effectiveness of purifying selection. Under this view, the excess nonsynonymous substitutions represent deleterious mutations or a composite of deleterious and compensatory mutations.

Nonsynonymous (or amino acid) divergences (K_a) are similar for the two pairs of *Buchnera,* with averages of 0.14 and 0.16 (*t*-test; $P = 0.43$). The ratio of synonymous to nonsynonymous substitutions (K_s/K_a) is consistently less for *Buchnera*-Mr–*Buchnera*-Sc (average $K_s/K_a = 5.1$) than for *Buchnera*-Sg–*Buchnera*-Dn (average $K_s/K_a = 7.0$), with 14 of 16 loci showing smaller values for *Buchnera*-Mr–*Buchnera*-Sc (sign test; $P < 0.01$). Assuming that synonymous sites reflect mutation rate and time (or number of generations) since divergence, an implication is that *Buchnera*-Mr–*Buchnera*-Sc are less constrained by selection than are *Buchnera*-Sg–*Buchnera*-Dn. Since this difference applies across loci with a variety of functions, the most plausible explanation is that the *Buchnera* in melaphidines have smaller effective population size and more genetic drift, either due to smaller host populations or due to smaller bottlenecks at the time of infection (unpublished data).

Ribosomal DNA Divergences

Table 5 shows divergences for 16S rDNA and 23S rDNA for the two pairs of *Buchnera* and for *E. coli–S. typhimurium.* Despite having been diverged for more time and many more generations, *E. coli–S. typhimurium* show rDNA divergences that are the same as or lower than those for the two pairs of *Buchnera*. This pattern resembles that for K_a values for protein-coding genes and is expected on the basis of much more effective selective constraint in the enterics due to their great-

Table 5
Divergences in rRNA Genes (as substitutions per site) for Pairwise Comparisons of *Buchnera*-Sg–*Buchnera*-Dn, *Buchnera*-Mr–*Buchnera*-Sc, and *Escherichia coli–Salmonella typhimurium*

	No. of Nucleotides	*Buchnera*-Sg–*Buchnera*-Dn		*Buchnera*-Mr–*Buchnera*-Sc		*E. coli–S. typhimurium*	
		K	SD	*K*	SD	*K*	SD
16S rRNA	1,530	0.027	0.004	0.045	0.005	0.033	0.004
23S rRNA	2,359	0.042	0.004	0.054	0.005	0.021	0.003

er effective population sizes. Since most sites in rRNA genes are under strong selective constraint, differences in rDNA divergences are expected to resemble differences for K_a values rather than neutral (K_s) differences. When the two pairs of *Buchnera* are compared, divergences are greater for both genes in *Buchnera*-Mr–*Buchnera*-Sc, with 16S and 23S values 70% and 28% greater, respectively. This parallels differences in K_a values for the two pairs of *Buchnera*, and it contrasts with the pattern for K_s values, for which *Buchnera*-Mr–*Buchnera*-Sc show smaller average divergences. The rate comparisons for rDNA genes add to evidence that selective constraints are less effective in *Buchnera*-Mr–*Buchnera*-Sc than in *Buchnera*-Sg–*Buchnera*-Dn.

Patterns of Amino Acid Substitution in *Buchnera* Versus Enterics

Buchnera has an AT-rich genome (Ohtaka and Ishikawa 1993; Baumann et al. 1997; Clark, Baumann, and Baumann 1998*a*), as do many intracellular bacteria. The bias in nucleotide composition affects the amino acid composition of polypeptides, resulting in a proportional increase in abundance of amino acids for which corresponding codons contain more A or T in the DNA sequence (Moran 1996). Our sequences confirm this amino acid bias for all four *Buchnera* taxa for the 14 chromosomally encoded genes available for all of the *Buchnera* (fig. 4). This pattern resembles that previously noted for the *trp* genes alone in two *Buchnera* taxa (Moran 1996); the current data show that the effect is genomewide. We examined the pattern of this bias in amino acid substitution across the evolution of this AT-rich group using a phylogenetic framework. By categorizing the amino acid substitutions according to their effects on AT content and mapping these substitutions on a phylogeny containing the four *Buchnera, E. coli,* and *S. typhimurium* (with *H. influenzae* as an outgroup), we can observe the effect of base composition on amino

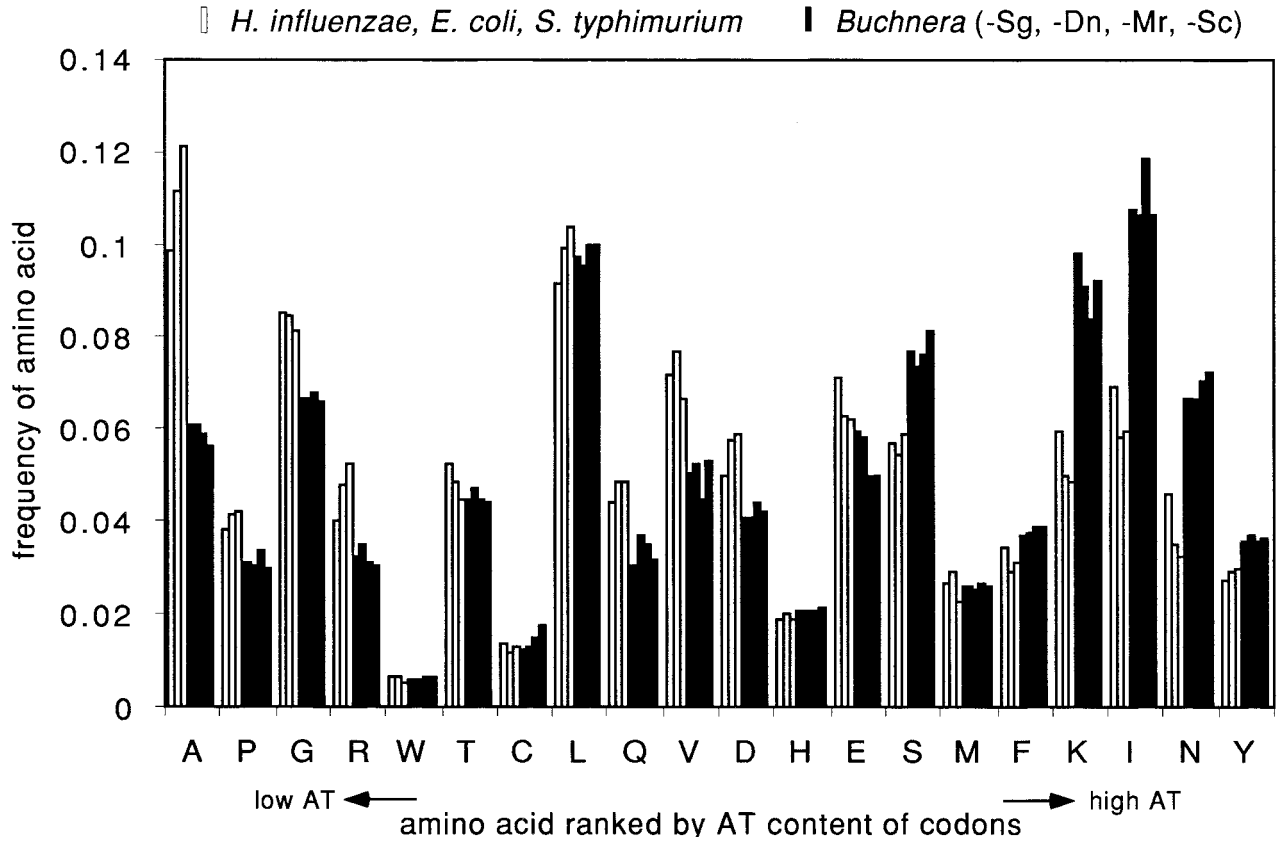

FIG. 4.—Amino acid compositions for *Haemophilus influenzae, Escherichia coli, Salmonella typhimurium,* and four *Buchnera* taxa, based on the 14 chromosomal coding genes available for all four *Buchnera* taxa (*ilvDC, ilvI, hisGD, trpDC(F)BA, gnd, atpAGD,* and *ftsZ*). Amino acids are grouped according to the number of A+T in the corresponding codon family.

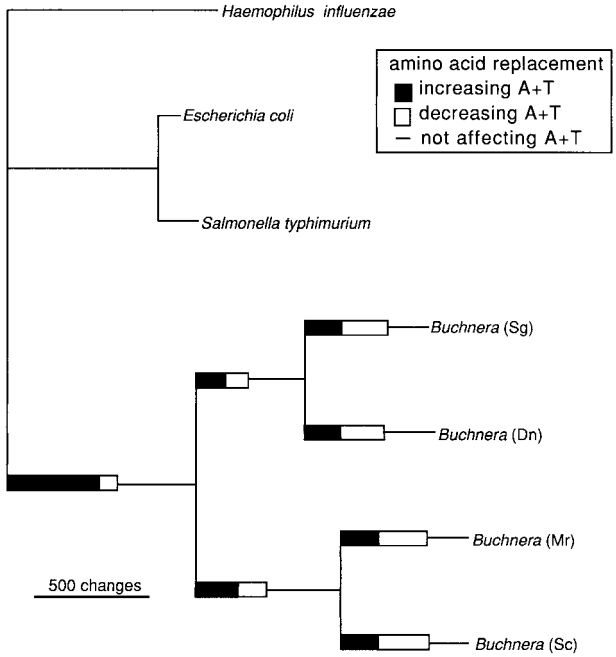

FIG. 5.—Pattern of amino acid replacement in relation to DNA base composition during the evolution of *Buchnera*. The tree is based on the 14 chromosomal genes available for all four *Buchnera* taxa (a total of 4,890 amino acid residues in the alignment).

acid substitutions at different stages in the evolution of *Buchnera* (fig. 5). The basal branch of *Buchnera* shows a very large predominance of amino acid replacements that result in more A+T in the corresponding codons, with 83% of the replacements that affect base composition favoring increased A+T. The intermediate branches in *Buchnera* are less biased in amino acid replacements, with 57% and 60% of the replacements resulting in more A+T. Replacements in the four terminal branches in the *Buchnera* tree are not biased in favor of A+T; these branches show a slight bias in favor of amino acid replacements that increase GC content (43%–46% for each of the four branches). Because homoplasy in the overall amino acid tree is low (the consistency index is 0.98 for the *Buchnera* portion of the tree), possible errors in mapping based on parsimony cannot be extensive enough to explain the observed pattern. Thus, early evolution of this symbiont group showed AT-biased mutation effecting an extremely biased pattern of amino acid replacement, but later replacements were less biased (fig. 5).

General Conclusions Regarding Rates of Evolution

Nonsynonymous substitutions in protein-coding DNA and substitutions in rDNA are substantially elevated in *Buchnera* as compared with enterics, resulting in much smaller K_s/K_a ratios for *Buchnera*. This is consistent with a reduced effect of purifying selection, either due to smaller population size causing more genetic drift or due to relaxation of selection. Of these two explanations, the latter is difficult to reconcile with the fact that effects of purifying selection are reduced at all loci. Loci examined in this study include amino acid biosynthetic genes that overproduce nutrients needed for growth of the hosts (*ilvIH, ilvDC, hisGD, trpDC(F)AB, leuABCD,* and *trpEG*), as well as ''housekeeping'' genes that function in the growth and replication of the bacterial cells (*ftsZ, gnd, atpAGD,* 16S rDNA, and 23S rDNA). Among the first set of genes, *trp* and *leu* loci have undergone plasmid-associated amplification as an adaptation to better provision hosts. The fact that all genes in this functionally diverse set show similar reductions in purifying selection supports small population size rather than reduced selective constraint as the basis of increased rate of substitution at sites under selection. The two pairs of *Buchnera* differ with regard to K_s/K_a ratios, implying that the extent of genetic drift or relaxation of selection is different for different lineages.

Brynnel et al. (1998) measured the synonymous and nonsynonymous rates for *tuf* genes in *Buchnera* across two of the same divergences, and they compared these with the divergences for *E. coli–S. typhimurium.* (Their ancestor A is the same as the ancestor of *Buchnera*-Sg–*Buchnera*-Dn of our study; their ancestor C is the same as the ancestor of *Buchnera*-Mr–*Buchnera*-Sc in our study.) Synonymous rates for *tuf* are slightly lower than those for loci in our study. As for genes in our study, synonymous divergence for *tuf* is higher for the *Buchnera* in Aphididae than for the *Buchnera* of Melaphidina, but this trend is reversed for nonsynonymous divergence, implying more genetic drift or less selective constraint in *Buchnera* of Melaphidina.

The primary exception to the homogeneous rates of silent evolution seen among *Buchnera* loci are *trpE* and *trpG* in the comparison between *Buchnera* of the Aphididae (*Buchnera*-Sg–*Buchnera*-Dn). In this pair, but not in *Buchnera*-Mr–*Buchnera*-Sc, *trpEG* is amplified as tandem repeats on plasmids (Lai, Baumann, and Baumann 1994; Lai, Baumann, and Moran 1995). The higher synonymous rate must reflect a higher mutation rate for this particular plasmid; similar elevation in synonymous rate has been observed for plasmid-borne *trpEG* in other Aphididae (Rouhbakhsh et al. 1996, 1997). A plausible explanation for this difference is that replication of the *trpEG* plasmid depends on a different polymerase; this possibility is consistent with the observation that the *trpEG* plasmid has a putative origin of replication that includes dnaA boxes, in contrast to the *leuABCD* plasmid.

Using *Buchnera* Rates to Estimate the Age of *Wolbachia*

The bacterium *Wolbachia pipientis* is a reproductive parasite that is widespread in arthropods and some other invertebrate hosts (Werren, Zhang, and Guo 1995; Bandi et al. 1998). Like *Buchnera,* it is intracellular and maternally inherited and has a reduced genome. For *ftsZ,* the only protein-coding gene available for many species, *Wolbachia* shows relatively fast rates of substitution based on nonsynonymous-sites ratios (K_s/K_a for *ftsZ* is 5.3–6.7 for divergences of major clades of *Wolbachia* compared with 5.6 and 13 for *Buchnera* and 23.7 for *E. coli–S.typhimurium*; table 3). In addition, *Wolbachia,* like *Buchnera,* shows an accelerated rate of substitution

in 16S rRNA relative to related nonsymbiotic bacteria (Moran 1996). Applying the *Buchnera* rate of synonymous substitution to K_s values calculated among the major clades of *Wolbachia,* we estimate that the basal divergence for the A and B clades of *Wolbachia* (mostly arthropod associates; Werren, Zhang, and Guo 1995) would be 32 MYA and that the divergence for the deepest ancestor (including nematode associates) of known *Wolbachia* would be 50–55 MYA. These dates are about half the previous estimates (Werren, Zhang, and Guo 1995; Bandi et al. 1998) using the *E. coli–S. typhimurium* synonymous rate (Ochman and Wilson 1987; consistent with table 4). The *Buchnera* rate is preferable for two reasons: the calibration is based on more firmly established evidence, and *Wolbachia,* as an intracellular parasite, is more similar to *Buchnera* in its biology and patterns of sequence evolution than to the enterics. An implication is that *Wolbachia* represents a relatively recent diversification that invaded insects and other animal groups during the Tertiary, long after the diversification of the hosts themselves.

AT Bias and Patterns of Amino Acid Substitution

The genomic base composition of *Buchnera,* at about 28% GC, is near the extreme for bacteria overall but is typical of a phylogenetically diverse set of intracellular bacteria including the mycoplasmas and spiroplasmas, endosymbionts of other insects, and rickettsiae (e.g., Winkler and Wood 1988; Andersson and Sharp 1996). In addition to major effects on codon use, this mutational bias favoring A+T affects amino acid frequencies in proteins, causing increased abundance of amino acids with more A+T in the corresponding codons (Sueoka 1988; Osawa et al. 1990; Moran 1996; Lobry 1997). Our analysis is the first to examine the pattern of amino acid replacement during the evolution of an AT-rich clade. Most of the shift toward amino acids with AT-rich codons occurred early in the evolution of *Buchnera,* with much of it preceding the split between melaphidines and aphidines, about 200 MYA. This result, combined with the phylogenetic placement of *Buchnera,* which is nested in the largely nonbiased gamma-3 subdivision of the *Proteobacteria,* indicates that an AT-rich lineage can evolve relatively quickly from an ancestor with intermediate base composition. Part of the reason for the decrease in the shift to A+T is that the strength of mutational bias changes over the evolution of a lineage; as A+T increases, fewer G or C sites remain to mutate. As in other AT-rich bacterial genomes, the AT bias in *Buchnera* is strongest at third codon positions and in spacer regions and weakest in first and second codon positions and rRNA genes (Muto and Osawa 1987; Moran 1996). This implies that the plateau in AT content is not due to mutational equilibrium alone, but to a balance between mutation and selective constraint.

Acknowledgments

This material is based on work supported by National Science Foundation Awards MCB-9807145 and DEB-9817795 (P.B.) and DEB-9815413 (N.A.M.) and by the the University of California Experiment Station (P.B.). We thank S. Elwyn, H. Ochman, and J. Wernegreen for comments on the analyses and manuscript.

LITERATURE CITED

ANDERSSON, S. G. E., and P. M. SHARP. 1996. Codon usage and base composition in *Rickettsia prowazekii.* J. Mol. Evol. **42**:525–536.

AYALA, F. J. 1997. Vagaries of the molecular clock. Proc. Natl. Acad. Sci. USA **94**:7776–7783.

BANDI, C., T. J. C. ANDERSON, C. GENCHI, and M. L. BLAXTER. 1998. Phylogeny of *Wolbachia* in filarial nematodes. Proc. R. Soc. Lond. B Biol. Sci **265**:2407–2413.

BANDI, C., M. SIRONI, G. DAMIANI, L. MAGRASSI, C. A. NALEPA, U. LAUDANI, and L. SACCHI. 1995. The establishment of intracellular symbiosis in an ancestor of cockroaches and termites. Proc. R. Soc. Lond. B Biol. Sci. **259**:293–299.

BAUMANN, L. and P. BAUMANN. 1994. Growth kinetics of the endosymbiont *Buchnera aphidicola* in the aphid *Schizaphis graminum.* Appl. Environ. Microbiol. **60**:3440–3443.

———. 1998. Characterization of *ftsZ,* the cell division gene of *Buchnera aphidicola* (endosymbiont of aphids) and detection of the product. Curr. Microbiol. **36**:85–89.

BAUMANN, P., L. BAUMANN, C.-Y. LAI, D. ROUHBAKHSH, N. A. MORAN, and M. A. CLARK. 1995. Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. Annu. Rev. Microbiol. **49**: 55–94.

BAUMANN, L., P. BAUMANN, and N. A. MORAN. 1998. The endosymbiont (*Buchnera*) of the aphid *Diuraphis noxia,* contains all the genes of the tryptophan biosynthetic pathway. Curr. Microbiol. **37**:58–59.

BAUMANN, L., P. BAUMANN, N. A. MORAN, J. SANDSTRÖM, and M. L. THAO. 1999. Genetic characterization of plasmids containing genes encoding enzymes of leucine biosynthesis in endosymbionts (*Buchnera*) of aphids. J. Mol. Evol. **48**: 77–85.

BAUMANN, L., M. A. CLARK, D. ROUHBAKHSH, P. BAUMANN, N. A. MORAN, and D. J. VOEGTLIN. 1997. Endosymbionts (*Buchnera*) of the aphid *Uroleucon sonchi* contain plasmids with *trpEG* and remnants of *trpE* pseudogenes. Curr. Microbiol. **35**:18–21.

BAUMANN, P., N. A. MORAN, and L. BAUMANN. 1997. The evolution and genetics of aphid endosymbionts. Bioscience **47**:12–20.

BISERCIC, M., J. Y. FEUTRIER, and P. K. REEVES. 1991. Nucleotide sequences of the *gnd* genes from 9 natural isolates of *Escherichia coli.* Evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. J. Bacteriol. **173**:3894–3900.

BLACKMAN, R. L., and V. F. EASTOP. 1984. Aphids on the world's crops. John Wiley Interscience, New York.

———. 1994. Aphids on the world's trees. CAB International, Wallingford, England.

BRACHO, A. M., D. MARTÍNEZ-TORRES, A. MOYA, and A. LATORRE. 1995. Discovery and molecular characterization of a plasmid localized in *Buchnera* sp., a bacterial endosymbiont of the aphid *Rhopalosiphum padi.* J. Mol. Evol. **41**: 67–73.

BRYNNEL, E. U., C. G. KURLAND, N. A. MORAN, and S. G. E. ANDERSSON. 1998. Evolutionary rates of *tuf* genes in endosymbionts of aphids. Mol. Biol. Evol. **15**:574–582.

CHARLES, H., and H. ISHIKAWA. 1999. Physical and genetic map of the genome of *Buchnera,* the primary endosymbiont

of the pea aphid *Acyrthosiphon pisum.* J. Mol. Evol. **48**: 142–150.

CHEN, X. A., S. LI, and S. AKSOY. 1999. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia.* J. Mol. Evol. **48**: 49–58.

CLARK, M. A., L. BAUMANN, and P. BAUMANN. 1998*a.* Sequence analysis of a 34.7 kb DNA segment from the genome of *Buchnera aphidicola* (endosymbiont of aphids) containing *groEL, dnaA,* the *atp* operon, *gidA* and *rho.* Curr. Microbiol. **36**:158–163.

———. 1998*b. Buchnera aphidicola* (aphid endosymbiont) contains genes encoding enzymes of histidine biosynthesis. Curr. Microbiol. **37**:356–358.

CLARK, M. A., P. BAUMANN, and N. A. MORAN. 1999. *Buchnera* plasmid-associated *trpEG* probably originated from a chromosomal location between *hslU* and *fpr.* Curr. Microbiol. **38**:309–311.

DOOLITTLE, R. F., D. F. FENG, S. TSANG, G. CHO, and E. LITTLE. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271**: 470–477.

EYREWALKER, A., and M. BULMER. 1995. Synonymous substitution rates in enterobacteria. Genetics **140**:1407–1412.

FENG, D. F., G. CHO, and R. F. DOOLITTLE. 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA **94**:13028–13033.

GENETICS COMPUTER GROUP. 1997. Program manual for the Wisconsin package. Version 9.1. Genetics Computer Group, Madison, Wis.

GIBBONS, R. J., and B. KAPSIMALIS. 1967. Estimates of the overall rate of growth of the intestinal microflora of hamsters, guinea pigs, and mice. J. Bacteriol. **93**:510–512.

GUTTMAN, D. S., and D. E. DYKHUIZEN. 1994. Clonal divergence in Escherichia coli as a result of recombination not mutation. Science **266**:1380–1383.

HEIE, O. E. 1987. Paleontology and phylogeny. Pp. 367–391 *in* A. K. MINKS and P. HARREWIJN, eds. World crop pests, Vol. 2A. Aphids: their biology, natural enemies and control. Elsevier, Amsterdam.

KIMURA, M. 1968. Evolutionary rate at the molecular level. Nature **217**:624–626.

———. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

LAI, C.-Y., L. BAUMANN, and P. BAUMANN. 1994. Amplification of *trpEG*: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. Proc. Natl. Acad. Sci. USA **91**:3819–3823.

LAI, C.-Y., L. BAUMANN, and N. A. MORAN. 1995. Genetics of the tryptophan biosynthetic pathway of the prokaryotic endosymbiont (*Buchnera*) of the aphid *Schlechtendalia chinensis.* Insect Mol. Biol. **4**:47–59.

———. 1996. The endosymbiont (*Buchnera* sp.) of the aphid *Diuraphis noxia* contains plasmids consisting of *trpEG* and tandem repeats of *trpEG* pseudogenes. Appl. Environ. Microbiol. **62**:332–339.

LAWRENCE, J. G., and H. OCHMAN. 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. **44**:383–397.

———. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95**:9413–9417.

LI, W.-H. 1993*a.* Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**: 96–99.

———. 1993*b.* So, what about the molecular clock hypothesis? Opin. Genet. Dev. **3**:896–901.

LOBRY, J. R. 1997. Influence of genomic G+C content on average amino-acid composition of proteins from 59 bacterial species. Gene **205**:309–316.

MADDISON, W. P., and D. M. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3. Sinauer, Sunderland, Mass.

MAIDAK, B. L., J. R. COLE, C. T. PARKER et al. (14 co-authors). 1999. A new version of the RDP (Ribosomal Database Project). Nucleic Acids Res. **27**:171–173.

MORAN, N. A. 1989. A 48-million-year-old aphid-host plant association and complex life cycle: biogeographic evidence. Science **245**:173–175.

———. 1992. The evolution of life cycles in aphids. Annu. Rev Entomol. **37**:321–348.

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

MORAN, N. A., M. A. MUNSON, P. BAUMANN, and H. ISHIKAWA. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. B Biol. Sci. **253**:167–171.

MORAN, N. A., and A. TELANG. 1998. The evolution of bacteriocyte-associated endosymbionts in insects. Bioscience **48**:295–304.

MUNSON, M. A., P. BAUMANN, M. A. CLARK, L. BAUMANN, N. A. MORAN, D. J. VOEGTLIN, and B. C. CAMPBELL. 1991. Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. J. Bacteriol. **173**:6321–6324.

MUTO, A., and S. OSAWA. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA **84**:166–169.

NELSON, K., and R. K. SELANDER. 1994. Intergenic transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. Proc. Natl. Acad. Sci. USA **91**:10227–10231.

OCHMAN, H., and A. C. WILSON. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26**:74–83.

OHTAKA, C., and H. ISHIKAWA. 1993. Accumulation of adenine and thymine in a groE-homologous operon of an intracellular symbiont. J. Mol. Evol. **36**:121–126.

OSAWA, S., A. MUTO, T. OHAMA, Y. ANDACHI, R. TANAKA, and F. YAMAO. 1990. Prokaryotic genetic code. Experientia **46**:1097–1106.

PACE, N. R. 1997. A molecular view of microbial diversity and the biosphere. Science **276**:734–740.

PAMILO, P., and N. O. BIANCHI. 1993. Evolution of the *zfx* and *zfy* genes—rates and interdependence between the genes. Mol. Biol. Evol. **10**:271–281.

PEEK, A. S., R. A. FELDMAN, R. A. LUTZ, and R. C. VRIJENHOEK. 1998. Cospeciation of chemoautotrophic bacteria and deep sea clams. Proc. Natl. Acad. Sci. USA **95**:9962–9966.

ROUHBAKHSH, D., and P. BAUMANN. 1995. Characterization of a putative 23S-5S rRNA operon of *Buchnera aphidicola* (endosymbiont of aphids) unlinked to the 16S rRNA-encoding gene. Gene **155**:107–112.

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.

ROUHBAKHSH, D., C.-Y. LAI, C. D. VON DOHLEN, M. A. CLARK, L. BAUMANN, P. BAUMANN, N. A. MORAN, and D. J. VOEGTLIN. 1996. The tryptophan biosynthetic pathway of

aphid endosymbionts (*Buchnera*): genetics and evolution of plasmid-associated anthranilate synthase (*trpEG*) within the Aphididae. J. Mol. Evol. **42**:414–421.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SAVAGEAU, M. A. 1983. Escherichia coli habitats, cell types, and mechanisms of gene control. Am. Nat. **122**:732–744.

SCHRÖDER, D., H. DEPPISCH, M. OBERMEYER, G. KROHNE, E. STACKEBRANDT, B. HÖLLDOBLER, W. GOEBEL, and R. GROSS. 1996. Intracellular endosymbiotic bacteria of *Camponotus* species (carpenter ants): systematics, evolution and ultrastructural characterization. Mol. Microbiol. **21**:479–490.

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. 1987*a*. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4**:222–230.

———. 1987*b*. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**:1281–1295.

SILVA, F., R. C. H. J. VAN HAM, B. SABATER, and A. LATORRE. 1998. Structure and evolution of leucine plasmids carried by the endosymbiont (*Buchnera aphidicola*) from aphids of the family Aphididae. FEMS Microbiol. Lett. **168**:43–49.

SMITH, J. M., and N. H. SMITH. 1996. Site-specific codon bias in bacteria. Genetics **142**:1037–1043.

SPAULDING, A. W., and C. D. VON DOHLEN. 1998. Phylogenetic characterization and molecular evolution of bacterial endosymbionts in psyllids (Hemiptera: Sternorrhyncha). Mol. Biol. Evol. **15**:1506–1513.

SUEOKA, N. 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85**:2653–2657.

SWOFFORD, D. L. 1999. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

TAKADA, H. 1991. Does the sexual female of *Schlechtendalia chinensis* (Bell) (Homoptera: Pemphigidae) ''viviparously'' produce the fundatrix? Appl. Entomol. Zool. **26**:117–121.

THAO, M. L., and P. BAUMANN. 1998. Sequence analysis of a DNA fragment from *Buchnera aphidicola* (aphid endosymbiont) containing the genes *dapD-htrA-ilvI-ilvH-ftsL-ftsI-murE*. Curr. Microbiol. **37**:214–216.

VAN HAM, R. C. H. J., A. MOYA, and A. LATORRE. 1997. Putative evolutionary origin of plasmids carrying the genes involved in leucine biosynthesis in *Buchnera aphidicola*. J. Bacteriol. **179**:4768–4777.

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

WERREN, J. H., W. ZHANG, and L. R. GUO. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. Proc. R. Soc. Lond. B Biol. Sci. **261**:55–71.

WINKLER, H. H., and D. O. WOOD. 1988. Codon usage in selected AT-rich bacteria. Biochimie **70**:977–986.

WOESE, C. R. 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271.

JULIAN ADAMS, reviewing editor

Accepted August 5, 1999