# Intraspecific Variation in Symbiont Genomes: Bottlenecks and the Aphid-Buchnera Association

**Daniel J. Funk, Jennifer J. Wernegreen and Nancy A. Moran**

*Department of Ecology and Evolutionary Biology, and Center for Insect Science, University of Arizona, Tucson, Arizona 85721*

Manuscript received April 17, 2000
Accepted for publication October 10, 2000

ABSTRACT

Buchnera are maternally transmitted bacterial endosymbionts that synthesize amino acids that are limiting in the diet of their aphid hosts. Previous studies demonstrated accelerated sequence evolution in Buchnera compared to free-living bacteria, especially for nonsynonymous substitutions. Two mechanisms may explain this acceleration: relaxed purifying selection and increased fixation of slightly deleterious alleles under drift. Here, we test the divergent predictions of these hypotheses for intraspecific polymorphism using Buchnera associated with natural populations of the ragweed aphid, *Uroleucon ambrosiae*. Contrary to expectations under relaxed selection, *U. ambrosiae* from across the United States yielded strikingly low sequence diversity at three Buchnera loci (*dnaN*, *trpBC*, *trpEG*), revealing polymorphism three orders of magnitude lower than in enteric bacteria. An excess of nonsynonymous polymorphism and of rare alleles was also observed. Local sampling of additional *dnaN* sequences revealed similar patterns of polymorphism and no evidence of food plant-associated genetic structure. Aphid mitochondrial sequences further suggested that host bottlenecks and large-scale dispersal may contribute to genetic homogenization of aphids and symbionts. Together, our results support reduced $N_e$ as a primary cause of accelerated sequence evolution in Buchnera. However, our study cannot rule out the possibility that mechanisms other than bottlenecks also contribute to reduced $N_e$ at aphid and endosymbiont loci.

OVER the last decade, molecular and phylogenetic investigations of the endosymbiotic bacterium *Buchnera aphidicola* and its aphid hosts (Baumann *et al.* 1997; Moran and Telang 1998) have provided compelling evidence for two phenomena: (1) a strictly congruent phylogenetic association between aphid and Buchnera loci (Munson *et al.* 1991; Moran *et al.* 1993; Clark *et al.* 2000; Funk *et al.* 2000); and (2) an elevated rate of DNA sequence evolution in Buchnera that can largely be attributed to increased fixation of nonsynonymous mutations (Moran 1996; Rouhbakhsh *et al.* 1997; Brynnel *et al.* 1998; Lambert and Moran 1998; Clark *et al.* 1999; Wernegreen and Moran 1999).

Potential explanations for these patterns are suggested by the biology of this endosymbiosis (Buchner 1965; Baumann *et al.* 1997). Buchnera belongs to the gamma division of the Proteobacteria, which also contains *Escherichia coli*. These bacteria are usually housed in specialized cells (bacteriocytes) within the body cavity of aphids (Buchner 1965). However, a small number of Buchnera infect each developing aphid egg or embryo, thus ensuring maternal transmission into the next host generation. In combination with fossil evidence, the strict phylogenetic congruence of Buchnera and aphids suggests that their association dates to the origin of the Aphidoidea ~200 mya (Moran *et al.* 1993). This ancient mutualism is also obligate and neither aphid nor bacterium can survive and reproduce without its symbiont. Despite its small genome (Charles and Ishikawa 1999), Buchnera retains genes for biosynthesis of essential amino acids (Baumann *et al.* 1997) that are limiting in the phloem sap diet of their hosts (reviewed in Douglas 1998; Sandström and Moran 1999). And in many Buchnera, genes encoding enzymes for leucine and tryptophan biosynthesis have been amplified on two types of plasmid, apparently as adaptations for increasing amino acid production and thus improving host nutrition (Lai *et al.* 1994; Bracho *et al.* 1995; van Ham *et al.* 1997; Baumann *et al.* 1999).

Reduced $N_e$ offers a potential explanation for the increased molecular evolutionary rates observed in Buchnera and other maternally inherited endosymbiotic bacteria relative to those of their free-living relatives (Moran 1996; Lambert and Moran 1998; Peek *et al.* 1998; Spaulding and von Dohlen 1998; Clark *et al.* 1999; Wernegreen and Moran 1999). If a substantial number of new mutations are slightly deleterious, reduced $N_e$ will cause more mutations to drift to fixation, resulting in an overall increase in rates of substitution (Ohta 1992). For symbiont mutations deleterious at both the level of the host and that of the symbiont, fixation is dependent both on the number of infective cells transmitted between generations and on the $N_e$ of female hosts in the population (Bergstrom and Pritchard 1998; Rispe and Moran 2000). This can

*Corresponding author:* Daniel J. Funk, Department of Biology, Vanderbilt University, Nashville, TN 37235.
E-mail: daniel.j.funk@vanderbilt.edu

be understood by noting that if few bacterial cells are transmitted to progeny, bacterial polymorphisms will be lost within individual aphids, whereas if hosts have small $N_e$ due to strong selection, small census population sizes, or population bottlenecks, polymorphisms will also be lost from host populations. Thus, the $N_e$ of symbionts will be dependent on the size, dynamics, and genetic structure of host populations. For these reasons, the $N_e$ of animal symbionts will likely be smaller than that of free-living bacterial taxa (Rispe and Moran 2000). The likely absence of recombination in vertically transmitted endosymbionts is also a contributing factor, since asexuality increases the rate of fixation of deleterious mutations (Muller 1964; Lynch 1996, 1997).

An alternative explanation for elevated rates of nonsynonymous substitution is positive selection for amino acid changes in endosymbionts. Positive selection, however, is not expected to show the consistent effects across all loci observed in Buchnera (Moran 1996; Wernegreen and Moran 1999) or to result in the decreased stability of gene products observed for rRNA genes in Buchnera and other endosymbionts (Lambert and Moran 1998). By contrast, demographic history and population structure are expected to affect all loci and to promote slightly deleterious substitution.

One other viable hypothesis, however, is consistent with the genome-wide effects observed in various intracellular bacteria: relaxation of purifying selection across loci. Bacterial symbionts and chronic pathogens show low maximum replication rates and may inhabit relatively stable habitats compared to most free-living bacteria, which undergo occasional periods of exponential growth and may experience more variable environments. Any reduced selection intensity due to the symbiotic lifestyle might eliminate the fitness benefits associated with maximizing replication rate or maintaining resistance to extreme conditions and thus permit the fixation of nonsynonymous mutations that would be removed by selection in other bacteria.

These two competing hypotheses—decreased $N_e$ *vs.* relaxed selection—cannot be distinguished on the basis of interspecific sequence divergence patterns, since a decrease in either $N_e$ or $s$ affects substitution rates similarly. However, the two hypotheses have different predictions for intraspecific polymorphism patterns. Under relaxed selection, a greater proportion of sites is effectively neutral, resulting in elevated polymorphism. By contrast, polymorphism is expected to be low under decreased $N_e$ due to increased loss of alleles under drift. The distribution of allele frequencies is also expected to differ under the two hypotheses, with an excess of rare alleles expected only under decreased $N_e$. However, because certain patterns are expected whether $N_e$ is reduced due to bottlenecks or to other mechanisms such as selective sweeps, quantifying their relative contribution to these patterns is difficult.

In this article we evaluate the "bottlenecks" hypothesis of increased genetic drift in the first study of DNA sequence variation among Buchnera [hereafter, referred to as "Buchnera (Ua)"] within a species of aphid host. To this end, we sequenced regions from aphid mitochondrial DNA, the Buchnera chromosome, and two plasmids from individuals of the ragweed aphid, *Uroleucon ambrosiae*, from 20 widely distributed U.S. localities. If reduced $N_e$ has played a role in symbiont evolution, expected patterns include: (1) low levels of overall polymorphism compared to other bacteria and insects, due to the loss of neutral alleles under drift (Kimura 1983; McDonald and Kreitman 1991); (2) higher ratios of nonsynonymous to synonymous variation for within-species *vs.* between-species comparisons, as predicted if (slightly deleterious) mutations are sometimes eliminated by selection before fixation (*e.g.*, Hasegawa *et al.* 1998); and (3) an excess of young or rare mutations compared to the strictly neutral expectation, reflecting the recent origin of mutations following a bottleneck and the low frequencies of slightly deleterious alleles (*e.g.*, Tajima 1989).

## MATERIALS AND METHODS

**Samples and loci:** *U. ambrosiae* (Thomas) (redescribed in Olive 1963) (Aphididae) is a cyclically parthenogenetic aphid with a broad North American distribution. It uses native food plants from the Asteraceae and especially the giant ragweed, *Ambrosia trifida*, on which it specializes over much of its range (Moran 1984). Our *U. ambrosiae* samples consisted of single individuals collected from food plants in the field at different localities across the United States, with two individuals from a Minnesota site in different years (Table 1, Figure 1). Most collections for the principal analyses of this study were made from *A. trifida*, but New York and Utah samples were collected from unidentified Ambrosia species and the Sapelo Island, Georgia, sample was taken on *Iva frutescens*, a close relative of Ambrosia and an alternative food plant of *U. ambrosiae* along the U.S. East Coast (Olive 1963).

All aphid samples were frozen alive and stored at −80° with vouchers from the same aphid colony stored in 70% ethanol. From each aphid, DNA sequences were collected from a segment spanning portions of the cytochrome oxidase I (COI) and cytochrome oxidase II (COII) genes of aphid mitochondrial DNA [1354 nucleotides (nt)], the entire *dnaN* gene on the Buchnera chromosome (1107 nt), and segments spanning parts of *leuBC* (1674 nt) and *trpEG* (1200 nt) from the leucine and tryptophan plasmids, respectively. Results reported below refer to coding sequences at these four loci, totaling 5325 nucleotides per haplotype.

**DNA extraction, amplification, and sequencing:** For each sample, total genomic DNA was extracted from a single aphid and suspended in TE following a protocol modified from Bender *et al.* (1983). Sequencing templates were amplified in a single polymerase chain reaction (PCR) experiment by submitting 50-μl reactions (*cf.* Moran *et al.* 1998) to 30–35 cycles of 1 min at 94°, 1 min at 54–60°, and 1.5–2.5 min at 72°. PCR products were filtered using the Concert Rapid PCR purification system (Life Technologies), resuspended in water, and submitted to the Laboratory of Molecular Systematics and Evolution Automated Sequencing Facility at the University of Arizona. For COI/II, *dnaN*, and *leuBC*, sequencing was performed in both directions using PCR primers plus an inter-

**TABLE 1**

**Collection information for aphid specimens sequenced for the principal analyses**

| Code | Locality | Date | Collector |
|---|---|---|---|
| AL | AL, Calhoun Co., Anniston | 10-2-97 | D. Funk |
| AZ | AZ, Santa Cruz Co., Tubac | 10-?-98 | D. Funk |
| GA.1 | GA, McIntosh Co., Sapelo Island | 4-8-99 | S. Pennings |
| GA.2 | GA, Dade Co., New England | 9-27-98 | D. Funk |
| IL.1 | IL, Wayne Co., Barnhill | 10-5-97 | D. Funk |
| IL.2 | IL, Dewitt Co. | 10-12-97 | D. Voegtlin |
| IN | IN, Monroe Co., Bloomington | 10-13-97 | N. Takebayashi |
| KY | KY, Grant Co., Crittenden | 9-27-98 | D. Funk |
| LA | LA, Madison Co., Tallulah | 10-3-97 | D. Funk |
| MD | MD, Montgomery Co., Silver Spring | 10-6-97 | L. Shapiro |
| MI | MI, Wayne Co., Plymouth | 9-26-98 | D. Funk |
| MN.a | MN, Ramsey Co., St. Paul | 9-?-97 | S. Scheffer |
| MN.b | MN, Ramsey Co., St. Paul | 9-24-98 | D. Funk |
| MS | MS, Madison Co., Jacksonville | 9-28-98 | D. Funk |
| NY | NY, Tompkins Co., Ithaca | 10-13-97 | M. Caillaud |
| OH | OH, Allen Co., Lima | 9-26-98 | D. Funk |
| TN.1 | TN, Davidson Co., Nashville | 9-16-97 | Ö. Pellmyr |
| TN.2 | TN, Campbell Co., Jellico | 9-27-98 | D. Funk |
| UT | UT, Millard Co., Cove | 9-29-97 | C. von Dohlen |
| VA | VA, Albermarle Co., Charlottesville | 10-6-97 | D. Funk |
| WV | WV, Raleigh Co., Beckley | 10-6-97 | D. Funk |

nal sequencing primer, yielding a continuous sequence fragment. For *trpEG*, analyzed sequences represent concatenated 600 nt 5′ and 3′ segments of the amplified fragment, with several hundred nucleotides missing between them.

Primers 8 and 18 from the Insect Mitochondrial DNA Primer Oligonucleotide Set (obtained from the University of British Columbia Nucleic Acid-Protein Service Unit; described in Simon *et al.* 1994) were used to amplify COI/II sequences and an internal sequencing primer in COI was designed based on initial *U. ambrosiae* sequences. Primers designed and used to amplify the other study loci are the following: *dnaN* (dnaN.430seqF 5′-TCT ATG GGM AAA CAA G-3′), *leuBC* (leuB.689F 5′-TGT ATG CAA ATT ATT AAA RAT CC-3′; leuC.1337R 5′-GCW GCC ATA ATA GGA CTA ACT A-3′; leuC.180F 5′-AAT ACA TTT GCH ACT ATG GAT C-3′), and *trpEG* (trpEUrol.151F 5′-AAG YAT YAT GAT YAT YGA YAG TGC-3′; trpGUrol.507R 5′-ATA GAT TCR GGR TGA AAT TG-3′).

Figure 1.—Collection localities for *U. ambrosiae* aphids. Dark sites yielded haplotypes belonging to the "common" lineage and pale sites yielded the "divergent" alleles (see results).

**Sequence alignment and data analysis:** Sequences from two Minnesota aphids were analyzed, and *trpEG* sequences were not collected for the IL.1 and TN.2 samples (Table 1) due to technical difficulties. With these exceptions, the principal analyses of this study treated mitochondrial, bacterial, leucine plasmid, and tryptophan plasmid DNA sequences obtained from a single aphid from each of the 20 sampling localities. These sequences can be found under GenBank accession nos. AF196354–AF196467.

Sequences were inspected and nucleotides at all variable sites confirmed using Sequence Navigator (Applied Biosystems, Foster City, CA). Alignment of sequences using Clustal, as implemented in Sequence Navigator, was straightforward as no insertions or deletions were inferred within *U. ambrosiae* and overall variation was low. Using MacClade (Maddison and Maddison 1992), coding regions were identified and translated by comparison with previously published sequences and then aligned with sequences from *U. rudbeckiae* and *U. aenum*, two congeners sequenced for another study (J. J. Wernegreen and N. A. Moran, personal communication). *U. rudbeckiae* is the most closely related Uroleucon species for which sequences were available (Moran *et al.* 1998).

Using the four-gamete test (Hudson and Kaplan 1985) in the computer program DnaSP (Rozas and Rozas 1999), we obtained the minimum number of historical recombination events implied by a combined data set including all four gene regions. The degree of association between nucleotides at different polymorphic sites was also tested in DnaSP using chi-square tests and applying a Bonferroni adjustment to maintain a constant experiment-wide critical $\alpha$ despite multiple comparisons (Sokal and Rohlf 1991).

We used DnaSP and another program, MEA (Moriyama 1997), to calculate synonymous and nonsynonymous sites and two common measures of nucleotide variation, $\pi$ (Nei and Li 1979) and $\theta_w$ (Watterson 1975). Nucleotide diversity, $\pi$, was calculated as the average number of nucleotide differences

between all pairs of sequences, and $\theta_w$ is based on the proportion of segregating nucleotide sites in our sample. Both $\pi$ and $\theta_w$ are estimators of the neutral parameter $\theta = 2N_e\mu$ for maternally inherited haploid genomes, where $N_e$ is female effective population size. We used $\theta_w$ and a published rate of synonymous mutation in these bacteria (Clark *et al.* 1999) to calculate effective population size for Buchnera (Ua) using the Buchnera chromosomal gene *dnaN*. Our mutation rate estimate was an average derived from 14 Buchnera loci with similar synonymous rates and calibrated using the host fossil record and estimates of the number of Buchnera cell divisions per generation (Clark *et al.* 1999).

Phylogenetic relationships among haplotypes were determined by branch-and-bound parsimony analyses of the equally weighted data set as implemented in PAUP* (Swofford 1999) using different combinations of *U. rudbeckiae* and *U. aenum* as outgroups. Statistical support for each node was obtained by bootstrap resampling based on 1000 branch-and-bound replicate trees. Evolutionary reconstructions of polymorphic nucleotide positions and amino acid residues over the topology were inferred with MacClade (Maddison and Maddison 1992).

We compared observed sequence variation with neutral expectations for each of the four gene regions of our study. In an equilibrium population undergoing neutral mutation, the estimate of $\theta$ based on the number of segregating sites will be the same as the estimate derived from $\pi$. We used Tajima's *D* (Tajima 1989) to evaluate whether our estimates of $\pi$ and $\theta_w$ were significantly different. Tajima's *D* depends on the frequency distribution of segregating sites, with negative values ($\pi < \theta_w$) indicating an excess of rare alleles and positive values ($\pi > \theta_w$) indicating an excess of intermediate frequencies. Fu and Li's *D* (Fu and Li 1993) evaluates whether the numbers of mutations on external branches (singletons) and internal branches in a genealogy are equal, as expected under neutrality. Fu's *F*'s is a statistic that compares the observed to the expected number of alleles from a sample for a given $\pi$ and may offer a more powerful means of rejecting neutrality than Tajima's *D* and Fu and Li's *D* (Fu 1997). Several other statistical comparisons of $\theta$ estimators (Tajima's *T*, Fu and Li's *F*, Fu and Li's *D**, Fu and Li's *F**; Fu and Li 1993; Fu 1997) were also applied.

The previous tests evaluate the hypothesis that all mutations in a DNA region are neutral in a population at stationarity. We also employed the method of McDonald and Kreitman (1991), which asks whether the ratio of replacement to synonymous polymorphisms within species is equal to the ratio of replacement to synonymous substitutions between species, as per the neutral prediction. For the interspecific comparisons in these tests, we used sequences from *U. rudbeckiae.*

**Local analyses of *U. ambrosiae* from southeastern Arizona:** To complement the principal analyses of this study, we also sequenced aphids collected from several localities and several food plants across southeastern Arizona, where *U. ambrosiae* is rather generalized in its food plant preferences (Funk and Bernays 2001). Along with the *A. trifida* Arizona specimen from the principal analyses, all 5325 nucleotides were also sequenced for three aphids from three different non-Ambrosia food plants [*Viguiera dentata* (Asteraceae), *Heterotheca subaxillaris* (Asteraceae), and *Sphaeralcea* sp. (Malvaceae)] (Table 2). For 18 additional specimens, the first 534 nucleotides of *dnaN* were sequenced. This data set was compiled to evaluate the possibility that different food plants support genetically differentiated aphid populations and to compare genetic diversity at different geographic scales.

## RESULTS

**Sequence and haplotype variation:** Polymorphic nucleotide sites and amino acid residues are illustrated in Table 3 and summarized by gene region in Table 4. No variable positions were detected in noncoding regions (66 nt of tRNA between COI and COII, a 1-nt spacer between *leuB* and *leuC*, and a 2-nt spacer between *trpE* and *trpG*). A total of 59 nt sites varied, of which 25 (42%) were nonsynonymous polymorphisms. No insertions or deletions were detected among *U. ambrosiae* sequences, but *U. ambrosiae* exhibited a five-codon insertion in COI and a one-codon insertion in *trpEG* compared to *U. rudbeckiae* and *U. aenum* sequences.

The 21 individuals sampled for the principal analyses yielded eight different haplotypes (Table 3). Sequences collected from 13 individuals were identical and represent the "common" haplotype. Together with the common haplotypes, distinct haplotypes from IL.1 and IN and a third shared by KY and VA samples constitute the highly homogeneous "common lineage" and differ from each other at a maximum of two of 5325 sites (0.04%) (Tables 5 and 6). The four remaining individuals showed distinct haplotypes differing from the common haplotype and one another at 17–30 sites (0.32–0.56% total sequence divergence). They are collectively referred to here as the "divergent" haplotypes (Tables 3 and 6). This combination of extremely closely and comparatively distantly related alleles yielded a distinctly bimodal frequency distribution of pairwise sequence divergences (Figure 2a). For each gene region, ~75% of polymorphic sites were singletons, and the minority nucleotide was observed in no more than four alleles at any site (Tables 3 and 4), yielding a skewed frequency distribution (Figure 2b).

**Estimates of $\theta$ and $N_e$:** Overall $\pi$ and $\theta_w$ values were low for each gene region compared to estimates from other bacteria and insect mitochondria (see discussion). Values are miniscule when calculated for the common lineage alone (Table 5). On the basis of our estimate of $\theta_w$ from synonymous sites in the Buchnera chromosomal gene *dnaN* (Table 5, 0.36%) and the per generation mutation rate from Clark *et al.* (1999; $1.4–1.9 \times 10^{-10}$ mutations/site/generation), effective population sizes for Buchnera (Ua) are about $1 \times 10^7$.

**Phylogenetic reconstruction:** For each gene region, ~25% of polymorphic sites, a total of 15 sites, were parsimony informative. The strong bootstrap support provided by these 15 characters (Figure 3) reflects a lack of homoplasy. Among the 59 nt sites that are polymorphic among *U. ambrosiae* haplotypes, only 1 (*leuC* site 1539) appears to have experienced multiple (two) substitutions. Using other Uroleucon species to root the tree introduced some homoplasy, with the result that the basal MN.b and GA alleles (Figure 3) switch positions when *U. aenum* is used as an outgroup. The rooted topology clearly demonstrates the common lineage to be cladistically derived relative to the basal divergent alleles. Evidence for recombination or the horizontal transfer of bacterial or plasmid symbionts among hosts is absent, and significant linkage disequilibrium was detected among gene regions. These phenomena are fur-

**TABLE 2**

**Collection information for southeastern Arizona aphid specimens sequenced for the local analyses**

| Food plant species | Food plant tribe | Locality | Date |
|---|---|---|---|
| *Ambrosia ambrosioides* | Heliantheae | Pima Co., Catalina St. Pk. | 11-24-96 |
| *A. ambrosioides* | Heliantheae | Pima Co., Tucson | 11-17-96 |
| *A. confertaflora* | Heliantheae | Pima Co., Arivaca | 10-15-96 |
| *A. confertaflora* | Heliantheae | Cochise Co., Fairbank | 10-18-96 |
| *A. confertaflora* | Heliantheae | Cochise Co., rt 90, San Pedro R. | 11-1-96 |
| *A. confertaflora* | Heliantheae | Santa Cruz Co., Tubac | 11-6-96 |
| *A. confertaflora* | Heliantheae | Santa Cruz Co., Tubac | 11-6-96 |
| *A. trifida* | Heliantheae | Pima Co., Arivaca | 10-15-96 |
| *A. trifida* | Heliantheae | Santa Cruz Co., Patagonia | 11-1-96 |
| *A. trifida* | Heliantheae | Cochise Co., St. David | 10-18-96 |
| *A. trifida* | Heliantheae | Santa Cruz Co., Tubac | 11-6-96 |
| *A. trifida* | Heliantheae | Santa Cruz Co., Tubac | 10-?-98 |
| *Heterotheca subaxillaris* | Astereae | Santa Cruz Co., Tubac | 11-6-96 |
| *H. subaxillaris* | Astereae | Santa Cruz Co., Patagonia | 11-15-97 |
| *Lactuca* sp. | Lactuceae | Pima Co., Tucson | 3-26-99 |
| *Senecio* sp. | Heliantheae | Santa Cruz Co., Madera Cyn | 10-17-96 |
| *Solidago wrightii* | Astereae | Pima Co., Mt. Lemon | 9-10-96 |
| *S. wrightii* | Astereae | Pima Co., Mt. Lemon | 9-10-96 |
| *Sphaeralcea* sp. | (Malvaceae) | Pima Co., Tucson | 3-17-99 |
| *Viguiera dentata* | Heliantheae | Cochise Co., Fairbank | 10-18-96 |
| *V. dentata* | Heliantheae | Cochise Co., Fairbank | 11-7-97 |
| *V. dentata* | Heliantheae | Pima Co., Tucson | 11-17-96 |
| *V.* sp. | Heliantheae | Santa Cruz Co., Madera Cyn | 10-17-96 |
| *V.* sp. | Heliantheae | Pima Co., Mt. Lemon | 9-10-96 |

The first 534 nucleotides of *dnaN* were sequenced from these specimens; for underlined samples only, the entire four-locus 5325-nucleotide haplotype was collected. All listed food plants belong to the Asteraceae except for *Sphaeralcea* sp., which belongs to the Malvaceae.

ther illustrated by the complete congruence among trees separately reconstructed for each of the assayed loci, using the data described here (Funk *et al.* 2000). The data exhibit no obvious phylogeographic structure as alleles from the common lineage are intermingled geographically with the divergent alleles, and both are widely scattered among U.S. collecting localities (Figure 1).

**Tests of neutrality:** The absence of multiple hits, recombination (or horizontal transfer), and population structure evident from our data is consistent with three assumptions of many population genetic models, those of infinite sites, no recombination, and panmixia. Multiple changes may separate the *U. rudbeckiae* sequences used in our McDonald-Kreitman tests from *U. ambrosiae* at some sites. However, per site rates of synonymous divergence between these species are $<0.3$ for each gene region (Table 5), below the onset of saturation.

Every gene region showed departures from a neutral frequency distribution for a majority of the tests of neutrality applied, as summarized in Table 7. In most cases, significantly negative skews were observed, implying an excess of rare polymorphisms or alleles. This is reflected in the prevalence of singletons (Table 4, Figure 2b) and the concentration of changes on external branches (31) relative to internal branches (21; Figure 3). Deviation from neutrality was also supported by a MacDonald-Kreitman test for *dnaN*, which showed an excess of nonsynonymous intraspecific polymorphism (Table 8). In contrast, the other gene regions showed similar $K_a/K_s$ values for polymorphism and divergence (Table 4).

**Local analyses:** Sequencing 22 Arizona aphid specimens for the first 534 nucleotides of *dnaN* yielded levels and patterns of polymorphism very similar to those observed at the larger geographic scale of the principal analyses. Of these sequences, 18 proved identical to the homologous segment of the common allele from the principal analyses, 3 specimens from non-Ambrosia food plants shared a second haplotype that differed from the common allele at two nucleotide sites (Table 9), and 1 specimen from *A. confertaflora* possessed a third haplotype. One of the three polymorphic sites in the local data set (site 111) was also polymorphic at the larger geographic scale. For the local data, $\pi = 0.00110$ and $\theta_w = 0.00155$. The common allele was observed for each of the three individuals from non-Ambrosia hosts for which the entire 5325-nucleotide data set was collected (Table 9).

In any study of intraspecific variation, the presence of morphologically cryptic species may compromise interpretations if unrecognized. In herbivorous insects, such cryptic species can sometimes be distinguished by their use of different food plants (see references in Funk 1998). Our data provide no evidence that food

**TABLE 3**

**Polymorphic nucleotide sites in coding regions from four symbiotic genomes**

| | | Aphid mitochondrion | | | | | | | | | Bacterial chromosome: | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | COI (801 bp) | | | | | | | COII (543 bp) | | *dnaN* (1107 bp) | | | | | | | | | | |
| | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | 0 | 1 | 2 | 3 | 3 | 6 | 6 | 9 | 0 | 0 | 1 | 3 | 4 | 5 | 5 | 7 | 7 | 7 | 8 | 8 |
| | | 0 | 0 | 9 | 8 | 9 | 5 | 6 | 2 | 8 | 8 | 1 | 9 | 0 | 4 | 4 | 1 | 3 | 5 | 1 | 7 |
| Allele | *N* | 9 | 3 | 5 | 4 | 3 | 8 | 3 | 5 | 1 | 4 | 1 | 1 | 9 | 4 | 7 | 9 | 9 | 3 | 7 | 4 |
| Common[a] | 13 | T | T | T | T | T | C | C | C | A | T | T | G | G | A | C | A | A | A | T | T |
| IL.1 | 1 | . | . | | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| IN | 1 | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . |
| KY, VA | 2 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C |
| OH | 1 | . | . | . | . | . | T | T | T | . | . | . | . | . | . | A | . | G | G | . | . |
| UT | 1 | . | . | . | . | C | . | T | . | . | . | C | . | . | . | . | C | G | . | . | . |
| MN.b | 1 | C | C | C | . | . | . | T | . | . | C | C | . | . | . | . | . | G | . | G | . |
| GA.1 | 1 | . | . | C | C | . | . | T | . | T | . | C | C | . | G | . | . | G | . | G | . |
| Ti/tv: | | i | i | i | i | i | i | i | i | v | i | i | v | i | i | v | v | i | i | v | i |
| Codon position: | | 3 | 1 | 1 | 3 | 3 | 1 | 3 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 2 | 1 | 3 | 1 | 1 |
| A.a. residue: | | F | L | L | H | Y | L | I | L | M | P | I | D | E | T | L | E | N | L | S | F |
| A.a. polymorphism: | | | | | | | | | | L | | | H | K | A | I | A | D | | A | L |

(*continued*)

plant association likewise structures genetic variation in *U. ambrosiae*. The common allele was observed in aphids from a variety of food plants in the southwestern United States (Table 2), and three other specimens collected on three different plant genera shared a single uncommon allele (Table 9). These findings corroborate the supposition that southwestern *U. ambrosiae* are food plant generalists rather than genetically differentiated populations that specialize on different food plants.

## DISCUSSION

**Bottlenecks and Buchnera:** Our findings support the hypothesis that reduced $N_e$ in Buchnera promotes the fixation of nonsynonymous substitutions and contributes to the accelerated molecular evolution of Buchnera relative to free-living bacteria. This support is most clearly manifest in the very low levels of synonymous polymorphism exhibited by Buchnera (Ua) relative to those of nonsymbiotic bacteria, as would be expected if increased exposure to genetic drift had promoted the loss of neutral alleles in these symbionts. In another recent study, Sreevatsan *et al.* (1997) similarly reported strikingly low levels of synonymous polymorphism when comparing their global samples of the human pathogen *Mycobacterium tuberculosis* with previously reported data from other bacterial species (*Vibrio cholerae*, *Streptococcus pyogenes*, *Neisseria meningtidis*, *M. avium-intracellulare*, *Escherichia coli*, *Borrelia burgdorferi*, and *Salmonella enterica*, in order of increasing polymorphism). The synonymous nucleotide diversity we observed for *dnaN* was also lower than for each of these seven bacterial taxa and was about three orders of magnitude lower than for *E. coli* or *S. enterica*, the closest relatives to Buchnera in this set, although it was still 20 times higher than in *M. tuberculosis*. Since Buchnera has a higher per site mutation rate (Ochman *et al.* 1999), the low observed $\pi$ must result from a substantial decrease in $N_e$.

Achtman *et al.* (1999) describe yet greater genetic homogeneity from global assays of the plague-causing *Yersinia pestis*, finding zero synonymous polymorphisms in sequences from several gene regions. These authors propose that *Y. pestis* represents a recent clone descending from within a more diversified species, much as the common lineage of Buchnera (Ua) is nested within a set of more divergent lineages (Figure 3). Similarly analyzing this geographically widespread common lineage as a distinct evolutionary entity yields estimates of synonymous polymorphism on par with those for *Y. pestis* and *M. tuberculosis* ($\pi$ averaged across *dnaN*, *leuBC*, and *trpEG* = 0.01%, due to a single synonymous change; Table 5, Figure 3). However, there is no evidence of such ecological differentiation from the samples yielding the common lineage of Buchnera (Ua), which were collected over a wide geographic range and principally on a single host plant.

Another point is raised by the *Y. pestis* example, however. If cryptic bacterial species or ecological types such as *Y. pestis* go unrecognized, estimates of polymorphism in the bacterial species that include them will be artificially inflated, with the result that other species [such as Buchnera (Ua)] might be inappropriately viewed as

**TABLE 3**

**(Continued)**

| | | Leucine plasmid | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *leuB* (366 bp) | | | | *leuC* (1308 bp) | | | | | | | | | | | | | | | | | |
| | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | 0 | 1 | 1 | 3 | 4 | 4 | 5 | 6 | 6 | 8 | 9 | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 5 | 5 | |
| | | 3 | 2 | 3 | 0 | 3 | 3 | 0 | 0 | 0 | 8 | 8 | 2 | 0 | 0 | 0 | 8 | 9 | 9 | 1 | 2 | 3 | |
| Allele | *N* | 1 | 6 | 5 | 6 | 0 | 4 | 6 | 0 | 9 | 5 | 4 | 8 | 3 | 2 | 9 | 1 | 0 | 5 | 6 | 1 | 9 | |
| Common[a] | 13 | G | T | A | C | G | A | G | T | A | A | A | A | T | T | T | T | A | G | T | G | T | |
| IL.1 | 1 | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | |
| IN | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | |
| KY, VA | 2 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | |
| OH | 1 | . | C | . | . | . | . | . | G | G | G | . | . | . | A | G | . | . | A | . | . | A | |
| UT | 1 | . | . | . | . | . | . | . | . | G | . | . | T | . | . | G | G | . | A | . | . | G | |
| MN.b | 1 | . | . | . | . | . | . | A | . | G | . | G | . | C | . | G | . | C | A | . | . | G | |
| GA.1 | 1 | A | . | G | . | A | G | . | . | G | . | G | . | . | . | G | . | . | A | C | A | G | |
| Ti/tv: | | i | i | i | i | i | i | i | v | i | i | i | v | i | v | v | v | v | i | i | i | v | |
| Codon position: | | 1 | 3 | 3 | 3 | 1 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 1 | 3 | 3 | 3 | 3 | |
| A.a. residue: | | V | G | A | N | G | E | S | S | K | G | G | L | T | N | L | S | S | V | N | R | G | |
| A.a. polymorphism: | | I | | | | S | G | N | | | | | F | | K | V | A | R | | | | | |

(*continued*)

depauperate in polymorphism by the kinds of comparisons made here. Nonetheless, because Buchnera (Ua) are consistently and appreciably less variable than those from an unbiased and fairly comprehensive set of published bacterial studies, it can safely be said that its overall levels of nucleotide polymorphism are exceedingly small.

In addition to reduced overall neutral variation, Buchnera (Ua) exhibited an excess of rare alleles relative to neutral expectations in both chromosomal (*dnaN*) and plasmid loci (*leuBC* and *trpEG*; Table 7). This excess may reflect the persistence of slightly deleterious mutations at low frequency under reduced $N_e$, as per the bottlenecks hypothesis. However, it is important to recognize that such patterns can also arise due to other deviations from the null model. Indeed, the observed concentration of rare alleles within a few unusual haplotypes (Figure 3) is not readily explained as the result

**TABLE 3**

**(Continued)**

| | | Tryptophan plasmid | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *trpE* (702 bp) | | | | | | | | | | | | *trpG* (498 bp) | | | | | | |
| | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | |
| | | 0 | 0 | 0 | 0 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 6 | 8 | 9 | 9 | 1 | 1 | 1 | |
| | | 2 | 5 | 6 | 7 | 6 | 7 | 9 | 1 | 5 | 6 | 9 | 9 | 3 | 2 | 4 | 7 | 9 | 9 | |
| Allele | *N* | 2 | 7 | 8 | 0 | 2 | 3 | 9 | 4 | 1 | 2 | 1 | 7 | 0 | 4 | 4 | 8 | 4 | 1 | 4 |
| Common[a] | 13 | C | C | T | C | C | T | C | A | C | C | T | G | G | T | C | G | G | C | |
| IL.1 | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | |
| IN | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | |
| KY, VA | 2 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | |
| OH | 1 | . | . | C | . | . | C | T | . | . | T | . | . | . | . | . | A | . | . | |
| UT | 1 | . | . | C | A | A | . | T | . | . | . | . | A | . | . | . | . | . | T | |
| MN.b | 1 | T | . | C | . | . | . | T | . | T | . | . | . | . | . | T | . | . | T | |
| GA.1 | 1 | . | T | C | . | . | . | T | G | T | . | C | . | A | C | . | . | A | T | |
| Ti/tv: | | i | i | i | v | v | i | i | i | i | i | i | i | i | i | i | i | i | i | |
| Codon position: | | 3 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 3 | 3 | |
| A.a. residue: | | L | H | H | N | Q | I | T | K | Y | I | D | T | L | V | I | V | R | R | |
| A.a. polymorphism: | | | Y | | K | K | T | I | E | | | | | | | | I | | | |

[a] The "common" four-locus allele was obtained from a majority of sequenced specimens and is thus equivalent to the consensus sequence.

**TABLE 4**

**Summary of haplotype and nucleotide variation across gene regions**

| | $N$ | Alleles | Solitary alleles | Bp | S | η(s) (%) | Rep. (%) | $K_a/K_s$ Poly | $K_a/K_s$ Div. | Div./poly. |
|---|---|---|---|---|---|---|---|---|---|---|
| COI/COII | 21 | 5 | 4 | 1344 | 9 | 78 | 11 | 0.02 | 0.06 | 9.1 |
| *dnaN* | 21 | 7 | 5 | 1107 | 11 | 73 | 73 | 0.64 | 0.15 | 3.6 |
| *leuB/leuC* | 21 | 6 | 5 | 1674 | 21 | 76 | 43 | 0.12 | 0.17 | 3.6 |
| *trpE/trpG* | 21 | 5 | 4 | 1200 | 18 | 78 | 39 | 0.15 | 0.18 | 5.2 |
| All loci | 21 | 8 | 6 | 5325 | 59 | 75 | 42 | | | |

*N*, number of samples providing sequences; alleles, observed unique haplotypes; solitary alleles, haplotypes recovered from a single sample; S, observed segregating sites; η(s) (%), percentage of segregating sites that are singletons; rep. (%), percentage of segregating sites involving an amino acid replacement; $K_a/K_s$, ratio of changes per site at nonsynonymous and synonymous sites for *U. ambrosiae* polymorphisms and for *U. ambrosiae/U. rudbeckiae* differences; div./poly., ratio of polymorphic sites between and within these species.

of slightly deleterious mutations alone. This pattern is suggestive of a history of population subdivision resulting in divergence among haplotypes or of a selective sweep involving positive selection (Atwood *et al.* 1951; Begun and Aquadro 1992) on haplotypes from the common lineage.

Applying the McDonald-Kreitman test to *dnaN* revealed a significant excess of nonsynonymous polymorphism for Buchnera sequences within *U. ambrosiae* compared to nonsynonymous divergence between *U. ambrosiae* and *U. rudbeckiae*. This pattern is as expected if some of the slightly deleterious mutations maintained by low $N_e$ and drift have been purged by selection prior to fixation (Table 8). Similar excesses of intraspecific amino acid polymorphism occur in the mitochondrial DNA of many animals, where they have likewise been interpreted as evidence for slightly deleterious mutations (Nachman 1998; Rand and Kann 1998).

Although the smaller number of polymorphic sites in our COI/II mitochondrial sequences provides more limited evidence, their patterns are also consistent with low $N_e$ of the aphid hosts. The nucleotide diversities of *U. ambrosiae* (Table 5) are, for example, somewhat smaller than those reported for the ND5 mitochondrial gene in *Drosophila melanogaster* and *D. simulans* (Rand and Kann 1996) and are generally low by insect standards.

**TABLE 5**

**Levels of polymorphism at synonymous and nonsynonymous sites**

| | Within sp. div. (%) | Between sp. div. (%) | All *U. ambrosiae* alleles π (%) | All *U. ambrosiae* alleles θ (%) | Common lineage[a] only π (%) | Common lineage[a] only θ (%) |
|---|---|---|---|---|---|---|
| COI/COII | | | | | | |
| All | 0.37 | 6.56 | 0.087 | 0.187 | 0.000 | 0.000 |
| Synonymous | 1.56 | 26.82 | 0.359 | 0.742 | 0.000 | 0.000 |
| Nonsynonymous | 0.09 | 1.56 | 0.009 | 0.027 | 0.000 | 0.000 |
| *dnaN* | | | | | | |
| All | 0.54 | 3.98 | 0.145 | 0.276 | 0.031 | 0.053 |
| Synonymous | 0.93 | 11.78 | 0.195 | 0.364 | 0.000 | 0.000 |
| Nonsynonymous | 0.45 | 1.94 | 0.132 | 0.253 | 0.039 | 0.067 |
| *leuB/leuC* | | | | | | |
| All | 0.72 | 5.17 | 0.180 | 0.365 | 0.007 | 0.018 |
| Synonymous | 2.34 | 14.60 | 0.555 | 1.067 | 0.032 | 0.081 |
| Nonsynonymous | 0.31 | 2.60 | 0.076 | 0.170 | 0.000 | 0.000 |
| *trpE/trpG* | | | | | | |
| All | 0.92 | 7.54 | 0.202 | 0.417 | 0.000 | 0.000 |
| Synonymous | 2.98 | 21.57 | 0.600 | 1.205 | 0.000 | 0.000 |
| Nonsynonymous | 0.42 | 3.71 | 0.094 | 0.206 | 0.000 | 0.000 |

Within sp. div. (%): percentage sequence divergence (uncorrected) between GA.1 and OH haplotypes, the most divergent in the multi-locus data set; between sp. div. (%): percentage sequence divergence (uncorrected) between the common haplotype and *U. rudbeckiae*.

[a] The common lineage includes 17 of the 21 multi-locus haplotypes, including those with the common allele plus the IL.1, IN, and KY/VA alleles (Table 3).

**TABLE 6**

**Percentage uncorrected nucleotide (above diagonal) and amino acid (below) divergences among multi-locus haplotypes and the outgroup**

| | Common | IL.1 | IN | KY/VA | OH | UT | MN.b | GA.1 | *U. rud.* |
|---|---|---|---|---|---|---|---|---|---|
| Common | | 0.02 | 0.02 | 0.02 | 0.36 | 0.32 | 0.41 | 0.56 | 5.82 |
| IL.1 | 0.00 | | 0.04 | 0.04 | 0.38 | 0.34 | 0.43 | 0.58 | 5.40 |
| IN | 0.06 | 0.06 | | 0.04 | 0.38 | 0.34 | 0.43 | 0.58 | 5.84 |
| KY/VA | 0.06 | 0.06 | 0.11 | | 0.38 | 0.34 | 0.43 | 0.58 | 5.84 |
| OH | 0.39 | 0.39 | 0.45 | 0.45 | | 0.36 | 0.43 | 0.56 | 5.77 |
| UT | 0.45 | 0.45 | 0.51 | 0.51 | 0.51 | | 0.36 | 0.51 | 5.73 |
| MN.b | 0.34 | 0.34 | 0.39 | 0.39 | 0.39 | 0.45 | | 0.45 | 5.61 |
| GA.1 | 0.62 | 0.62 | 0.68 | 0.68 | 0.73 | 0.79 | 0.56 | | 5.67 |
| *U. rud.* | 5.24 | 5.24 | 5.30 | 5.30 | 5.30 | 5.35 | 5.13 | 5.13 | |

COI/II allele frequency distributions are also negatively skewed and sometimes diverge significantly from neutral expectations (Table 7). However, no excess of within-species nonsynonymous polymorphism was detected in COI/II, providing no direct evidence for slightly deleterious variation.

**Selection on symbionts:** Increased rates of nonsynonymous substitution are a consistent pattern across Buchnera loci in previous studies, suggesting that positive selection is not a general cause of increased rates of molecular evolution in these bacteria (Moran 1996). As the first assay of DNA sequence variation within Buchnera of a single aphid species, the present study is able to address an alternative hypothesis: that purifying selection is relaxed for endosymbiotic bacteria, perhaps due to the relative stability of their intracellular environment. As a sole explanation of the accelerated sequence evolution observed in Buchnera, the relaxed selection hypothesis does not predict, nor readily explain, our findings of lower levels of genetic diversity within species or the observed non-neutral distributions of allele frequencies. By contrast, these results are expected under the hypothesis of low $N_e$, as described above.

Our ability to infer the specific mechanisms responsible for low polymorphism in Buchnera (Ua) is, however, limited by the fact that multiple processes can contribute to small $N_e$. These processes include fluctuations in census population size (or bottlenecks), selective sweeps involving linkage and strong positive selection (Atwood *et al.* 1951; Begun and Aquadro 1992), and background selection involving linkage and negative selection (Charlesworth *et al.* 1993). Our analyses of the combined multi-locus data set indicate significant linkage and no evidence for recombination. These findings complement documented patterns of phylogenetic congruence (historical linkage) among aphid, bacterial, and plasmid genomes in this system. They also suggest ample opportunity for genetic hitchhiking since selection at a site in any of these regions will have consequences for sequence variation at the others. However,

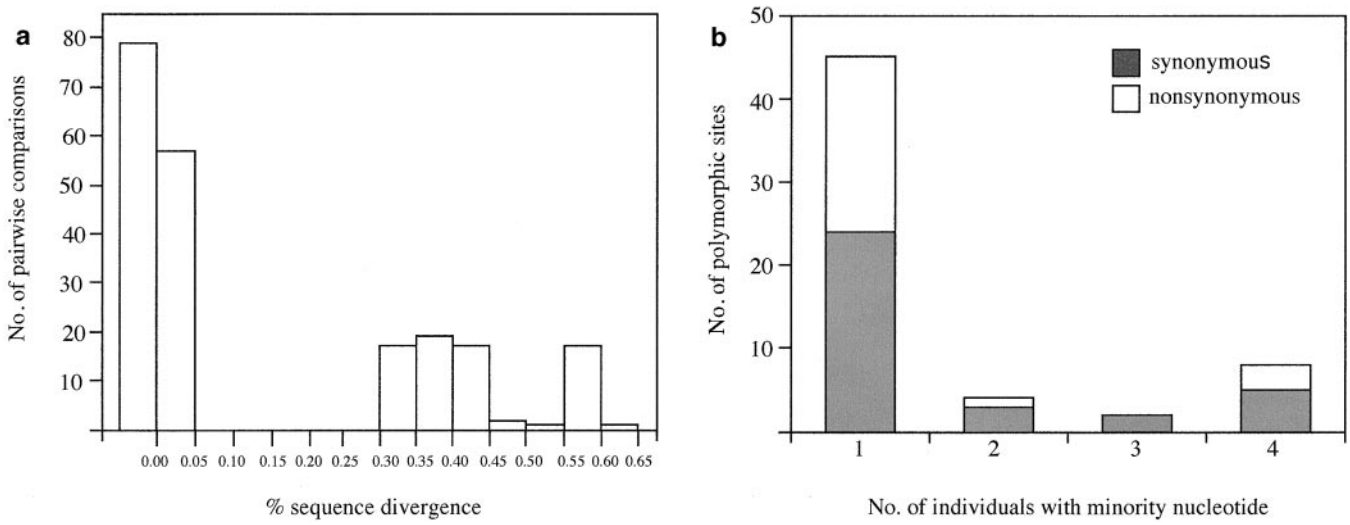


Figure 2.—(a) Frequency distribution of uncorrected sequence divergences derived from all pairwise comparisons among the 21 *U. ambrosiae* multi-locus haplotypes. (b) Frequency distribution of the number of synonymous and nonsynonymous polymorphic sites at which the minority nucleotide was observed in a given number of the 21 multi-locus haplotypes.

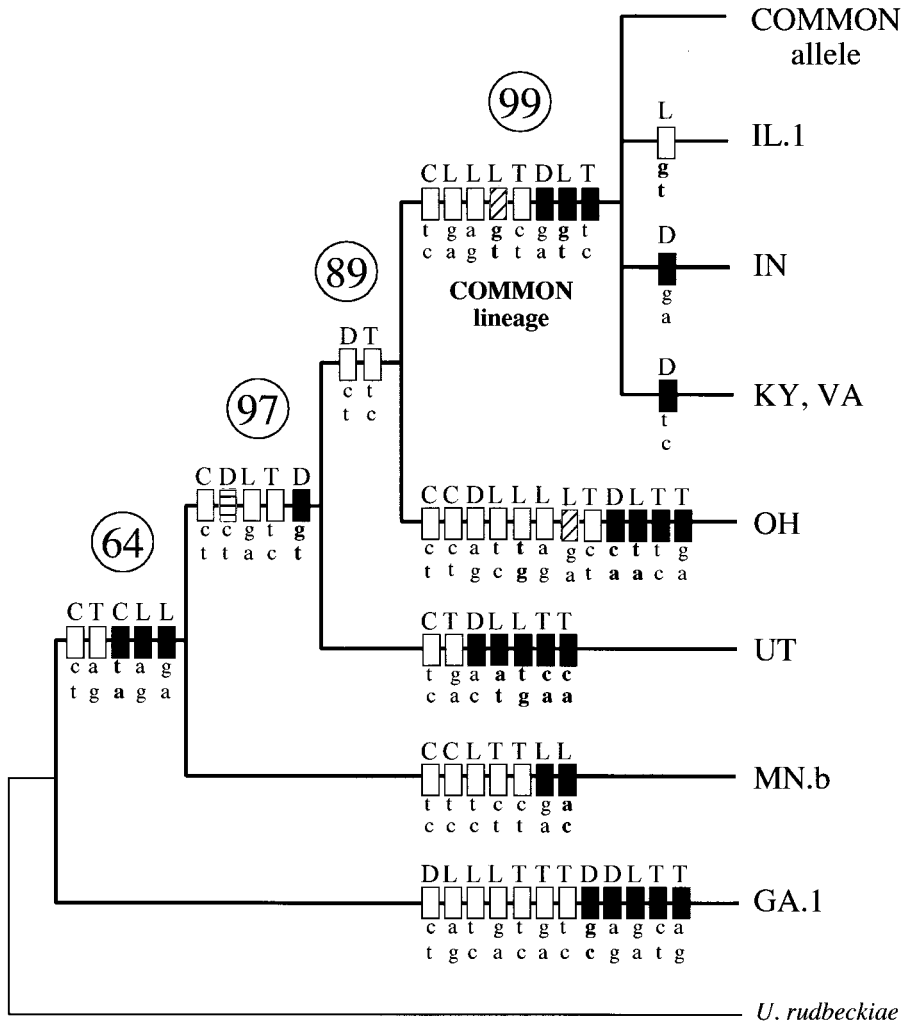


FIGURE 3.—Phylogenetic relationships among multi-locus haplotypes with bootstrap support. Inferred synonymous (open bars) and nonsynonymous (solid bars) changes are mapped onto the tree. Changes are labeled by nucleotides (below: transversions in boldface) and by gene region (above: C, COI/II; D, *dnaN*; L, *leuBC*; T, *trpEG*). Diagonally striped bars reflect the observation of two substitutions at one site; a horizontally striped bar indicates the instance of homoplasy introduced by outgroup comparison. Consistency index, 0.98.

hitchhiking is not predicted to give an excess of nonsynonymous polymorphism as we observed at *dnaN*. Rather, if selective sweeps result in fixation of certain genotypes that experience favorable mutations, an excess of nonsynonymous differences between species might be expected, with a relative lack of nonsynonymous polymorphism. In contrast, our results suggest that small census population sizes have played an important role in the evolution of this endosymbiosis.

Our results do not, however, completely exclude relaxed selection, perhaps in combination with small $N_e$, as a partial basis for accelerated substitution of nonsynonymous alleles in Buchnera. It is plausible, for example, that the evolutionary transition to endosymbiosis may have simultaneously introduced a less stressful selective environment and bottleneck-induced population structure in the bacterial ancestor of Buchnera. Thus, relaxed selection may coincide with small $N_e$, making their relative contributions difficult to disentangle. Similar limitations apply to other studies that have argued

**TABLE 7**

**Tests of the neutral model**

| | Tajima's *D* | Tajima's *T* | Fu's *F*'s | Fu and Li's *D* | Fu and Li's *F* | Fu and Li's *D** | Fu and Li's *F** |
|---|---|---|---|---|---|---|---|
| COI/COII | −1.806* | −1.678* | −0.646 | −1.414 | −1.612 | −2.158* | −2.152* |
| *dnaN* | −1.671 | −1.637* | −1.960 | −1.930* | −1.982* | −2.019* | −2.035* |
| *leuB/leuC* | −1.823* | −1.752* | 1.107 | −2.535* | −2.465** | −2.495* | −2.432* |
| *trpE/trpG* | −1.924* | −1.797* | 1.327 | −2.348* | −2.345* | −2.454* | −2.414* |
| Total | | −1.885* | 2.812 | −2.732*** | −2.647*** | −2.568* | −2.530* |

*$P < 0.05$; **$P < 0.01$; ***$P < 0.005$.

**TABLE 8**

**McDonald-Kreitman tests comparing nonsynonymous and synonymous nucleotide variation within and between aphid host species**

| | Polymorphisms within *U. ambrosiae* | Differences between *U. ambrosiae* and *U. rudbeckiae* | *G* | *P* |
|---|---|---|---|---|
| COI/COII | | | 0.039 | 0.8430 |
| Nonsynonymous | 1 | 11 | | |
| Synonymous | 8 | 71 | | |
| *dnaN* | | | 4.394 | 0.0361 |
| Nonsynonymous | 8 | 15 | | |
| Synonymous | 3 | 25 | | |
| *leuB/leuC* | | | 0.084 | 0.7714 |
| Nonsynonymous | 9 | 30 | | |
| Synonymous | 13 | 50 | | |
| *trpE/trpG* | | | 0.150 | 0.6983 |
| Nonsynonymous | 7 | 32 | | |
| Synonymous | 9 | 51 | | |

In each of 10 codons, two positions differed between *U. ambrosiae* and *U. rudbeckiae* sequences. Assuming only two changes, both must have been replacements in three instances, while in seven instances (two in COI/II, three in *leuBC*, two in *trpEG*) one change could be and was counted as silent, yielding a slightly less conservative test. In one *trpE* codon, all positions differed, and two were counted as silent. Within *U. ambrosiae*, a single (*leuC*) site with three nucleotides was treated as two separate changes. Three sites at which one *U. ambrosiae* sequence shared a nucleotide with *U. rudbeckiae* and parsimony unambiguously indicated convergence were counted as both a polymorphism and an interspecific difference. The two most 5′ polymorphic *trpG* sites (Table 3) are not included because no *U. rudbeckiae* data was available for comparison. The last two columns present test statistics and probability values for *G*-tests of independence (Sokal and Rohlf 1991).

for small $N_e$ as the basis for accelerated evolution in mammals (Ohta 1993, 1995) and in *D. melanogaster* (Akashi 1996) since changes in population structure may often be accompanied by ecological shifts that affect selection intensity. However, in these cases and in Buchnera the acceleration in sequence evolution affects loci throughout the genome, a pattern unlikely to be caused by many ecological scenarios, but consistent with reduced $N_e$.

**TABLE 9**

**Polymorphic nucleotide sites in *dnaN* for the local analyses**

| | | Bacterial chromosome *dnaN* (534 bp) | | |
|---|---|---|---|---|
| | | 1 | 2 | 4 |
| | | 1 | 9 | 4 |
| Allele | *N* | 1 | 1 | 1 |
| Common | 18 | T | G | G |
| Mt. Lemmon, *Solidago wrightii* | 3 | C | A | . |
| Mt. Lemmon, *Viguiera* sp. | | C | A | . |
| Madera Cyn., *Senecio* sp. | | C | A | . |
| rt. 90, San Pedro R., *A. confertaflora* | 1 | . | . | A |
| Ti/tv: | | i | i | i |
| Codon position: | | 3 | 3 | 3 |
| A.a. residue: | | I | M | G |
| A.a. polymorphism: | | | I | |

These treated specimens were collected on various food plants in southeastern Arizona (see Table 2).

**Insights from aphid biology:** Observations on natural *U. ambrosiae* populations suggest that they may experience repeated and severe bottlenecking. In southeastern Arizona, these aphids occur at very low densities for much of the year, but may be quite common during March and April and again in October and November. This fluctuation in census population sizes is due, in part, to changes in food plant abundance resulting from variable rainfall patterns between seasons and years. During the present study, for example, *U. ambrosiae* were relatively uncommon at Arizona collecting localities during the autumns of 1997 and 1999 but very abundant at the same localities in the autumn of 1998. This capacity of aphid populations to rapidly rebound from reduced densities owes to their rapid development and parthenogenetic life history, which allow a single female to produce thousands of descendants in a single season.

Our data also suggest that the genetic homogeneity of *U. ambrosiae* may be further promoted by considerable long-distance gene flow. The lack of geographic structure in our principal data set and the distribution of the common allele on opposite sides of North America indicate that dispersal of *U. ambrosiae* is far and frequent on an evolutionary time scale. These results, plus the observation of similar levels and patterns of nucleotide diversity at continental and regional scales, suggest that *U. ambrosiae* haplotypes are randomly mixed across the

United States. The long-distance dispersal of aphids as aeroplankton has been documented in other aphid taxa and may provide a mechanism for such pronounced gene flow (Loxdale *et al.* 1993).

Population genetic studies on other aphid species mostly indicate low levels of polymorphism compared to other insect species, although aphid taxa vary in the extent of both polymorphism and geographic structure (*e.g.*, De Barro *et al.* 1995; Simon *et al.* 1996, 1999; Loxdale *et al.* 1998; Wilson *et al.* 1999). Some studies suggest that selection linked to host plants or other ecological factors sometimes erodes polymorphism or produces geographic or host-plant-related subdivision within aphid populations (*e.g.*, Sunnucks *et al.* 1997). Such selection could decrease $N_e$ below expectations on the basis of census population sizes alone. In the case of *U. ambrosiae*, behavioral studies have indicated genetic differentiation between eastern and southwestern populations in their degree of food plant specificity (Funk and Bernays 2001). However, the lack of geographic or food plant-associated genetic structure observed in the present study indicates that this regional and possibly adaptive differentiation occurs in the face of gene flow (Ehrlich and Raven 1969).

Previous molecular genetic studies of aphids have focused on pest populations using introduced crops. By contrast, our study on *U. ambrosiae* represents the first extensive molecular population genetic assay of an aphid using wild host plants in its native range. A parallel study on *Pemphigus obesinymphae*, a distantly related, gall-dwelling aphid of cottonwoods with a very different life cycle, has revealed low levels of polymorphism comparable to those reported here (P. Abbot, personal communication). Such natural systems may provide unique insights into the population genetics underlying the aphid-Buchnera symbiosis.

We thank various colleagues (Table 1) for contributing aphid specimens and Michael Nachman and Howard Ochman for comments on the analyses and text. Louise Helbling and Michael Gelsey helped to collect sequence data. D.J.F. and J.J.W. were funded by a National Institutes of Health (NIH) training grant in Molecular Insect Science administered through the Center for Insect Science at the University of Arizona. This work was supported by NIH and by National Science Foundation grants DEB9815413 and DEB9978518 to N.A.M.

## LITERATURE CITED

Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule *et al.*, 1999 *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. USA **96:** 14043–14048.

Akashi, H., 1996 Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics **144:** 1297–1307.

Atwood, K. C., L. K. Schneider and F. J. Ryan, 1951 Selective mechanisms in bacteria. Cold Spring Harbor Symp. Quant. Biol. **16:** 345–355.

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

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.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally-occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

Bender, W., P. Spierer and D. S. Hogness, 1983 Chromosomal walking and jumping to isolate DNA from the *ace* and *rosy* loci and the bithorax complex in *Drosophila melanogaster*. J. Mol. Biol. **168:** 17–33.

Bergstrom, C. T., and J. Pritchard, 1998 Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics **149:** 2135–2146.

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 for *tuf* genes in endosymbionts of aphids. Mol. Biol. Evol. **15:** 574–582.

Buchner, P., 1965 *Endosymbiosis of Animals with Plant Microorganisms*. John Wiley & Sons, New York.

Charles, H., and I. 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.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

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

Clark, M. A., N. A. Moran, P. Baumann and J. J. Wernegreen, 2000 Cospeciation between bacterial endosymbionts and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution **54:** 517–525.

De Barro, P. J., T. N. Sherratt, G. R. Carvalho, D. Nicol, A. Iyengar *et al.*, 1995 Geographic and microgeographic genetic differentiation in 2 aphid species over southern England using the multilocus (GATA)(4) probe. Mol. Ecol. **4:** 375–382.

Douglas, A. E., 1998 Nutritional interactions in insect-microbe symbioses: aphids and their symbiotic bacteria *Buchnera*. Annu. Rev. Entomol. **43:** 17–37.

Ehrlich, P. A., and P. H. Raven, 1969 Differentiation of populations. Science **165:** 1228–1232.

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

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

Funk, D. J., 1998 Isolating a role for selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. Evolution **52:** 1744–1759.

Funk, D. J., and E. A. Bernays, 2001 Geographic variation in host specificity reveals host range evolution in *Uroleucon ambrosiae* aphids. Ecology (in press).

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

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

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. Genetics **111:** 147–164.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.

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.

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

Loxdale, H. D., J. Hardie, S. Halbert, R. Foottit, N. A. C. Kidd *et al.*, 1993 The relative importance of short- and long-range movement of flying aphids. Biol. Rev. **68:** 291–311.

Loxdale, H. D., C. P. Brookes, I. R. Wynne and S. J. Clark, 1998 Genetic variability within and between English populations of the damson-hop aphid, *Phorodon humuli* (Hemiptera: Aphididae), with special reference to esterases associated with insecticide resistance. Bull. Entomol. Res. **88:** 513–526.

Lynch, M., 1996 Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol. Biol. Evol. **13:** 209–220.

Lynch, M., 1997 Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. Mol. Biol. Evol. **14:** 914–925.

Maddison, W. P., and D. R. Maddison, 1992 MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer Associates, Sunderland, MA.

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

Moran, N. A., 1984 The genus *Uroleucon* (Homoptera: Aphididae) in Michigan: key, host records, biological notes, and descriptions of three new species. J. Kansas Entomol. Soc. **57:** 596–616.

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

Moran, N. A., and A. Telang, 1998 Bacteriocyte-associated symbionts of insects—a variety of insect groups harbor ancient prokaryotic endosymbionts. Bioscience **48:** 295–304.

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. Ser. B **253:** 167–171.

Moran, N. A., M. E. Kaplan, M. J. Gelsey, T. G. Murphy and E. A. Scholes, 1998 Phylogenetics and evolution of the aphid genus *Uroleucon* based on mitochondrial and nuclear DNA sequences. Syst. Entomol. **24:** 1–9.

Moriyama, E. N., 1997 Molecular Evolutionary Analysis Package. Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT.

Muller, H. J., 1964 The relation of recombination to mutational advance. Mutat. Res. **1:** 2–9.

Munson, M. A., P. Baumann, M. A. Clark, L. Baumann, N. A. Moran *et al.*, 1991 Aphid-eubacterial endosymbiosis: evidence for its establishment in an ancestor of four aphid families. J. Bacteriol. **173:** 6321–6324.

Nachman, M. W., 1998 Deleterious mutations in animal mitochondrial DNA. Genetica **102/103:** 61–69.

Nei, M., and W.-H. Li, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

Ochman, H., S. Elwyn and N. A. Moran, 1999 Calibrating bacterial evolution. Proc. Natl. Acad. Sci. USA **96:** 12638–12643.

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

Ohta, T., 1993 An examination of the generation-time effect on molecular evolution. Proc. Natl. Acad. Sci. USA **90:** 10676–10680.

Ohta, T., 1995 Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40:** 56–63.

Olive, A. T., 1963 The genus *Dactynotus* Rafinesque in North Carolina (Homoptera: Aphididae). Misc. Publ. Entomol. Soc. Am. **56:** 556–559.

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

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

Rand, D. M., and L. M. Kann, 1998 Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica **102/103:** 393–407.

Rispe, C., and N. A. Moran, 2000 Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. Am. Nat. **156:** 425–441.

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

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

Sandström, J., and N. Moran, 1999 How nutritionally imbalanced is phloem sap for aphids? Entomol. Exp. Appl. **91:** 203–210.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu *et al.*, 1994 Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. **87:** 651–701.

Simon, J. C., D. MartinezTorres, A. Latorre, A. Moya and P. D. N. Hebert, 1996 Molecular characterization of cyclic and obligate parthenogens in the aphid *Rhopalosiphum padi* (L). Proc. R. Soc. Lond. Ser. B **263:** 481–486.

Simon, J. C., S. Baumann, P. Sunnucks, P. D. N. Hebert, J. S. Pierre *et al.*, 1999 Reproductive mode and population genetic structure of the cereal aphid Sitobion avenae studied using phenotypic and microsatellite markers. Mol. Ecol. **8:** 531–545.

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

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.

Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth *et al.*, 1997 Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA **94:** 9869–9874.

Sunnucks, P., P. J. De Barro, G. Lushai, N. Maclean and D. Hales, 1997 Genetic structure of an aphid studied using microsatellites: cyclic parthenogenesis, differentiated lineages and host specialization. Mol. Ecol. **6:** 1059–1073.

Swofford, D. L., 1999 *PAUP* Phylogenetic Analysis Using Parsimony (* and other methods). Version 4.* Sinauer Associates, Sunderland, MA.

Tajima, F., 1989 Statistical method for testing the neutral mutation model in finite populations. Genetics **128:** 183–192.

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* (endosymbiont of aphids). J. Bacteriol. **179:** 4768–4777.

Watterson, G. A., 1975 On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

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

Wilson, A. C. C., P. Sunnucks and D. F. Hales, 1999 Microevolution, low clonal diversity and genetic affinities of parthenogenetic Sitobion aphids in New Zealand. Mol. Ecol. **8:** 1655–1666.

Communicating editor: W. F. Eanes