2004

Host–Symbiont Stability and Fast Evolutionary Rates in an Ant–Bacterium Association: Cospeciation of Camponotus Species and Their Endosymbionts, Candidatus Blochmannia

Patrick H. Degnan
Josephine Bay Paul Center for Comparative Molecular Biology and Evolution

Adam B. Lazarus
Josephine Bay Paul Center for Comparative Molecular Biology and Evolution

Chad D. Brock
University of Nebraska-Lincoln

Jennifer J. Wernegreen
Josephine Bay Paul Center for Comparative Molecular Biology and Evolution

Syst. Biol. 53(1):95–110, 2004

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150490264842

Host–Symbiont Stability and Fast Evolutionary Rates in an Ant–Bacterium Association: Cospeciation of *Camponotus* Species and Their Endosymbionts, *Candidatus* Blochmannia

PATRICK H. DEGNAN,[1] ADAM B. LAZARUS,[1] CHAD D. BROCK,[2] AND JENNIFER J. WERNEGREEN[1]

[1]*Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, Massachusetts 02543, USA; E-mail: jwernegreen@mbl.edu (J.J.W.)*
[2]*School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588, USA*

Abstract.—Bacterial endosymbionts are widespread across several insect orders and are involved in interactions ranging from obligate mutualism to reproductive parasitism. *Candidatus* Blochmannia gen. nov. (*Blochmannia*) is an obligate bacterial associate of *Camponotus* and related ant genera (Hymenoptera: Formicidae). The occurrence of *Blochmannia* in all *Camponotus* species sampled from field populations and its maternal transmission to host offspring suggest that this bacterium is engaged in a long-term, stable association with its ant hosts. However, evidence for cospeciation in this system is equivocal because previous phylogenetic studies were based on limited gene sampling, lacked statistical analysis of congruence, and have even suggested host switching. We compared phylogenies of host genes (the nuclear *EF-1αF2* and mitochondrial *COI/II*) and *Blochmannia* genes (16S ribosomal DNA [rDNA], *groEL*, *gidA*, and *rpsB*), totaling more than 7 kilobases for each of 16 *Camponotus* species. Each data set was analyzed using maximum likelihood and Bayesian phylogenetic reconstruction methods. We found minimal conflict among host and symbiont phylogenies, and the few areas of discordance occurred at deep nodes that were poorly supported by individual data sets. Concatenated protein-coding genes produced a very well-resolved tree that, based on the Shimodaira–Hasegawa test, did not conflict with any host or symbiont data set. Correlated rates of synonymous substitution (d_S) along corresponding branches of host and symbiont phylogenies further supported the hypothesis of cospeciation. These findings indicate that *Blochmannia–Camponotus* symbiosis has been evolutionarily stable throughout tens of millions of years. Based on inferred divergence times among the ant hosts, we estimated rates of sequence evolution of *Blochmannia* to be ~0.0024 substitutions per site per million years (s/s/MY) for the 16S rDNA gene and ~0.1094 s/s/MY at synonymous positions of the genes sampled. These rates are several-fold higher than those for related bacteria *Buchnera aphidicola* and *Escherichia coli*. Phylogenetic congruence among *Blochmannia* genes indicates genome stability that typifies primary endosymbionts of insects. [*Blochmannia*; *Camponotus*; cospeciation; endosymbiosis; molecular clock; phylogenetics.]

Bacteria have forged intracellular symbioses with insects several times independently, with significant effects on the fitness and diversification of their hosts. Primary (P) endosymbionts occur in several insect orders, including Coleoptera, Dictyoptera, Diptera, Hemiptera, and Hymenoptera (Buchner, 1965; Douglas, 1989). Common features of P endosymbionts include their need for host survival and reproduction and their restriction to specialized host cells (bacteriocytes) and female ovaries (or milk glands; Cheng and Aksoy, 1999) through which they are transmitted to the offspring (Buchner, 1965; Douglas, 1989; Baumann et al., 2000; Moran and Baumann, 2000; Thao et al., 2000b). In contrast to facultative symbionts that have transitory associations with hosts, P endosymbionts have been vertically inherited over long evolutionary timescales. Phylogenetic congruence suggests host–symbiont cospeciation over tens or hundreds of millions of years that can be traced back to an ancient infection event within each host group (Moran et al., 1993; Baumann et al., 1995; Chen et al., 1999; Clark et al., 2000; Thao et al., 2000b; Spaulding and von Dohlen, 2001; Lo et al., 2003). Genomes of P endosymbionts have signatures of extreme specialization resulting from this long-term intracellular lifestyle, including small genome size (~450–809 kilobases [kb]; Gil et al., 2002; Wernegreen et al., 2002) compared with the 4.5–5.5-megabase genomes of their close relative *Escherichia coli* (Bergthorsson and Ochman, 1998). Full genome sequence analysis illustrates that this reduction involves the loss of metabolic functions and DNA repair capabilities (e.g., Shigenobu et al., 2000; Tamas et al., 2001; Akman et al., 2002; van Ham et al., 2003), the latter of which may account for the extreme AT-richness of endosymbiont DNA sequences. Known functions of P endosymbionts include providing essential nutrients that are absent or in low quantities from the hosts' typically unbalanced diets (Nogge, 1981; Douglas and Prosser, 1992; Akman et al., 2002). Although these insect associates have originated independently in several bacterial groups, they are particularly common in the γ-Proteobacteria, which also includes the well-studied *E. coli* and related enterobacteria.

In contrast to P endosymbionts, certain bacteria form labile associations with insects and in addition to maternal transmission can spread horizontally to new hosts (Fukatsu et al., 2000; Fukatsu, 2001; Sandström et al., 2001; Russell et al., 2003). These secondary (S) endosymbionts are thought to have more subtle effects on host fitness, such as increased defense against parasitoids (Oliver et al., 2003). Phylogenetic analyses have revealed frequent horizontal transfer of these endosymbionts among insect species, often to entirely new host groups. Reconstructing the evolutionary history of insect–bacterium associations can help to distinguish whether a given bacterial associate better fits the model of P endosymbiont (cospeciation) or S endosymbiont (host switching). The aim of this study was to test for cospeciation between *Camponotus*, a diverse ant genus that represents >10% of all ant species (Bolton, 1995),

and its bacterial associate *Blochmannia*, a γ-3 proteobacterium that exhibits certain characteristics of both P and S endosymbionts.

Blochmannia *of* Camponotus

Blochmannia shows many hallmarks of a P endosymbiont that suggest long-term stability of its association with ants. Phylogenetic analysis of the 16S ribosomal DNA (rDNA) gene indicates a close relationship with *Wigglesworthia* and other endosymbionts in the γ-3 subdivision of the Proteobacteria (Schroder et al., 1996; Charles et al., 2001). *Blochmannia* is maintained with high fidelity and has been found in all field specimens of *Camponotus*, *Polyrhachis*, and *Colobopsis* screened thus far (Dasch, 1975; Dasch et al., 1984; Sameshima et al., 1999). This symbiosis is maintained in a variety of ecological contexts because the ant hosts have a cosmopolitan distribution across Old and New World environments ranging from deserts to tropical rainforests (Bolton, 1995). The ubiquity of *Blochmannia* in its host genera contrasts with the sporadic distribution of a related bacterium within the ant genus *Formica*, whose presence is variable among ant colonies of the same species (Dasch, 1975; Dasch et al., 1984). Like other P endosymbionts, *Blochmannia* is located within bacteriocytes, where the bacterial cells float freely in the host cytoplasm (although certain other P endosymbionts are contained within host-derived membranes) and within oocytes of queens, consistent with maternal transmission and integration into larvae during development (Sauer et al., 2002). The genome of *Blochmannia*, like that of many other P endosymbionts, has an extremely low GC content ($\sim$27–30% GC; Dasch, 1975; Gil et al., 2003;) and is severely reduced ($\sim$706–809 kb; Wernegreen et al., 2002; Gil et al., 2003).

Other aspects of *Blochmannia* biology suggest it might be horizontally transmitted among hosts. Although bacteriocytes of other insects are often clustered into an organlike bacteriome, those containing *Blochmannia* are intercalated among the ant midgut epithelial cells (Buchner, 1965; Schröder et al., 1996; Sauer et al., 2000, 2002). By virtue of this close proximity to the midgut, the bacteria might have greater opportunities for horizontal transfer among species (e.g., via body secretions). It is unclear whether *Blochmannia* provides an essential function for the host; these bacteria have been eliminated from lab-reared ants with no apparent detrimental effects on the host (Sauer et al., 2002). Certain arboreal *Camponotus* and *Polyrhachis* species have restricted diets composed mostly of extrafloral nectar and insect exudates (Davidson, 1997, 1998; Davidson et al., 2003), but most *Camponotus* species have diverse and complex diets that include plant material, insect secretions, and other arthropods (Dasch et al., 1984; Hölldobler and Wilson, 1990). Despite the complex diet of its host, the full genome sequence of *Blochmannia* from *C. floridanus* strongly suggests a nutritional basis for the symbiosis (Gil et al., 2003). The small *Blochmannia* genome retains genes for the biosynthesis of several amino acids essential to the host. It also retains genes for urease, which hydrolyzes urea to CO_2 and ammonia, and genes for glutamine synthetase, which can recycle ammonia before it reaches toxic concentrations. The mutualism is truly reciprocal because *Blochmannia* relies on the host for most nonessential amino acids, vitamins, and cofactors. It has also been suggested that the energetic functions of *Blochmannia* might be most important at the colony founding stage (D. Hann, pers. comm.; Wernegreen et al., 2003), and that the bacterium might also be involved in the production of pheromones used for food recruitment (Sauer, 2000). If *Blochmannia* provides a more subtle or sporadic selective advantage than is typical of nutritional P endosymbionts, it may exhibit a more labile host association.

Reconstructing Host–Symbiont Associations

Phylogenetic approaches provide powerful tools for inferring the stability of host–symbiont associations. Even very rare symbiont transfer, if successful, is expected to disrupt phylogenetic congruence. An overall agreement of *Camponotus* and *Blochmannia* phylogenies suggests stability of symbiont transmission (Schröder et al., 1996; Sameshima et al., 1999; Sauer et al., 2000). Moreover, the monophyly of *Camponotus*, *Polyrhachis*, and *Colobopsis* and the co-occurrence of *Blochmannia* in all three genera (Sameshima et al., 1999) further support the hypothesis of a single infection of a common ancestor and subsequent codivergence. However, previous studies were based on a single gene of the host and symbiont, had limited power to resolve phylogenetic relationships, and suggested horizontal transfer of *Blochmannia* between two host species (from *C. pennsylvanicus* to *C. castaneus*; Sauer et al., 2000). Our aim in this study was to evaluate the stability of the *Blochmannia*–*Camponotus* association by analyzing several loci from the symbiont, host mitochondria, and host nucleus totaling >7 kb for each of 16 host species. In addition, we used parameter-rich maximum likelihood (ML) and Bayesian analyses to improve the reliability of the estimated relationships (Felsenstein, 1981; Huelsenbeck and Ronquist, 2001). A variety of approaches can be employed to compare alternative evolutionary hypotheses, including ML (Goldman et al., 2000), Bayesian analysis (Huelsenbeck et al., 2000), and maximum parsimony (Brooks, 1981; Farris et al., 1994). We have selected a likelihood-based approach, the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa, 1999) based on several advantages reviewed previously (Goldman et al., 2000).

Demonstrating host–symbiont cospeciation offers a unique opportunity to calibrate rates of bacterial molecular evolution. Such rates are very difficult to estimate for free-living prokaryotes that lack a fossil record (Ochman et al., 1999) and experience lateral gene transfer (Lawrence and Ochman, 2002). The most reliable estimates of evolutionary rates in bacteria have been calculated for *Buchnera*, based on divergence times among the hosts as inferred from the fossil record of aphids or their preferred host plants (Moran et al., 1993; Clark et al., 1999). Here, we built upon evidence for cospeciation

between *Blochmannia* and *Camponotus* by estimating absolute rates of sequence evolution in this endosymbiont for comparison with other bacterial groups.

Materials and Methods

Specimens Examined

DNA sequence data were collected from 16 *Camponotus* species and two outgroup species, *Formica fusca* and *Polyrhachis foreli*. Most ant specimens were collected near the Southwestern Research Station (SWRS) in Portal, Arizona during the summer of 2001 and from Falmouth, Massachusetts during the summers of 2001 and 2002 (Table 1). Ants were identified in the field by J.J.W. or A.B.L., and vouchers were confirmed by Stefan Cover of the Museum of Comparative Zoology (MCZ, Cambridge, MA). Vouchers are available for inspection in the Wernegreen lab (MBL) or at the MCZ. All specimens used for DNA analysis were frozen and stored at −80°C. Specimens of *C. floridanus*, *C. pennsylvanicus*, and *Polyrhachis foreli* were obtained from colleagues (see Acknowledgments).

DNA Extraction, Amplification, and Sequencing

Genomic DNA was extracted from single ants using the DNeasy Tissue Kit (Qiagen) and diluted for polymerase chain reactions (PCRs). Six gene fragments were amplified: the *Blochmannia* genes for 16S rDNA, a heat shock protein (*groEL*), glucose inhibited cell division protein A (*gidA*), and ribosomal protein subunit B (*rpsB*), the *Camponotus* mitochondrial genes encoding cytochrome oxidase subunits I and II (*COI*/*II*), and the *Camponotus* nuclear gene encoding elongation factor 1 alpha F2 copy (*EF-1αF2*). Combined, these gene fragments totaled ~7,342 base pairs (bp) of high-quality sequence data per *Camponotus* species sampled (Table 1). Specific PCR primers are listed in Table 2. PCRs contained 0.4 pmol/μl of primers, 1.0 mM dNTPs (Invitrogen), PCR buffer (Fisher and Promega), 0.02 U of *Taq* DNA polymerase (Fisher and Promega), 5 μl of diluted genomic DNA, and $MgCl_2$ (concentrations optimized for each primer pair) and were brought to a final volume of 50 μl with double-distilled water. Reactions were performed on the PTC-0200 gradient cycler (MJ Research) under the following conditions: 2 min at 94°C followed by 35 cycles of 20 sec at 95°C, 50 sec at an annealing temperature specific to the primer pair, 1–2 min at 72°C. PCR products were cleaned using the Qiaquick PCR purification kit (Qiagen) and quantified by spectrophotometry and agarose gel analysis. Some products were cloned into plasmid vectors to obtain sufficient template for DNA sequencing. Appendix 1 lists detailed methods, including PCR conditions for each gene (Appendix 2). PCR products and clones were sequenced on an ABI 3700 automated sequencer using Big Dye 3.0 chemistry (Applied Biosystems). Internal sequencing primers were used to sequence products >1 kb. Sequences were assembled and edited using PHRED, PHRAP, and CONSED. All DNA assemblies were checked by eye, and any ambiguous base assignments were changed to N.

Phylogenetic Analysis

Alignment and data sets considered.—Protein-coding genes were aligned using translated amino acid sequences in ClustalX (Thompson et al., 1997) and manually edited using MacClade 4.05 (Maddison and Maddison, 2002). Non–protein-coding regions, including the tRNA and intergenic spacer between *COI* and *COII* and the introns of *EF-1αF2*, were aligned using ClustalX. *Blochmannia* 16S rDNA sequences were aligned using the Ribosomal Database Project II Sequence Aligner (Maidak et al., 2001), which considers the secondary structure of the rRNA molecule, and the resulting alignment was edited manually in MacClade. Regions with ambiguous alignment were few because the sequences were closely related, but any such regions were excluded from the analysis. After confirming that individual gene regions gave similar phylogenies, we concatenated the five protein-coding data sets (referred to hereinafter as the complete data set). We also analyzed combined data sets of the *Blochmannia* protein-coding genes *gidA, groEL, and rpsB* (the ALLBloch data set) and of the *Camponotus* protein-coding genes *COI*/*II* and *EF-1αF2* (the ALLCamp data set). *Blochmannia* 16S rDNA was not included in these concatenated protein-coding data sets because of the requirement for distinct models of sequence evolution for 16S rRNA (rate variation among sites following a gamma distribution) and protein-coding genes (varying rates among codon positions) in Bayesian analyses.

Outgroup selection.—Phylogenetic analysis of 16S rDNA suggests that *Wigglesworthia*, the bacterial endosymbiont of tsetse flies, is closely related to *Blochmannia* (Charles et al., 2001). Thus, we used genes from the fully sequenced genome of *Wigglesworthia brevipalpis* (NC004344; Akman et al., 2002) as the outgroup for analyses of two of the endosymbiont protein-coding genes (*gidA* and *groEL*). As an outgroup for the 16S rDNA and *rpsB* analyses, we obtained these genes from *Blochmannia* of *Polyrhachis foreli* after confirming its position outside our *Camponotus–Blochmannia* sample at the host *COI*/*II* (data not shown). The related formicine ant *Formica fusca* was used as the outgroup for *Camponotus COI*/*II* and *EF-1αF2*.

ML and Bayesian analyses.—For ML analyses, substitution models for each data set were selected using Modeltest 3.06 (Posada and Crandall, 1998), and ML heuristic searches were performed using 100 random taxon addition replicates with tree bisection and reconnection (TBR) branch swapping. ML bootstrap support was determined by 100 bootstrap replicates, each using 10 random taxon addition replicates with TBR branch swapping. Searches were performed in parallel on a Beowulf cluster using the clusterpaup program (A.G. McArthur, jbpc.mbl.edu/mcarthur) and PAUP 4.0b10 (Swofford, 2002).

Bayesian analyses were performed using both informative and noninformative prior probabilities and a general time reversible (GTR) substitution model with unequal base frequencies and the portion of invariant

TABLE 1. Ant specimens, collection location, GenBank accession numbers, and average sequence lengths for gene regions sampled.

			Blochmannia				*Camponotus*	
Specimen	Subgenus	Collection location	16S rDNA (1,339 bp)	*rpsB* (423 bp)	*gidA* (933 bp)	*groES/EL* (1,522 bp)	*COI/II* (1,677 bp)	*EF-1αF2* (1,448 bp)
Camponotus Mayr								
C. americanus	*Camponotus* Mayr	Falmouth, MA, USA	AY334379	AY334461	AY334429	AY334445	AY334395	AY334412
C. chromaiodes	*Camponotus*	Falmouth, MA, USA	AY334376	AY334458	AY334426	AY334442	AY334392	AY334409
C. laevigatus	*Camponotus*	Portal, AZ, USA	AY334370	AY334450	AY334418	AY334434	AY334384	AY334401
C. pennsylvanicus	*Camponotus*	Concord, MA, USA	AY196850	AY334457	AY334425	AY334441	AY334391	AY334408
C. novaboracensis	*Camponotus*	Falmouth, MA, USA	AY334378	AY334460	AY334428	AY334444	AY334394	AY334411
C. schaefferi	*Camponotus*	Portal, AZ, USA	AY334373	AY334454	AY334422	AY334438	AY334388	AY334405
C. castaneus	*Tanaemyrmex*	Falmouth, MA, USA	AY334377	AY334459	AY334427	AY334443	AY334393	AY334410
C. festinatus	*Tanaemyrmex*	Portal, AZ, USA	AY196851	AY334452	AY334420	AY334436	AY334386	AY334403
C. ocreatus	*Tanaemyrmex*	Portal, AZ, USA	AY334372	AY334453	AY334421	AY334437	AY334387	AY334404
C. sansabeanus	*Tanaemyrmex*	Portal, AZ, USA	AY334368	AY334448	AY334416	AY334432	AY334382	AY334399
C. vafer	*Tanaemyrmex*	Portal, AZ, USA	AY334369	AY334449	AY334417	AY334433	AY334383	AY334400
C. vicinus	*Tanaemyrmex*	Portal, AZ, USA	AY334374	AY334455	AY334423	AY334439	AY334389	AY334406
C. ulcerosus	*Myrmaphaenus* Emery	Portal, AZ, USA	AY334375	AY334456	AY334424	AY334440	AY334390	AY334407
C. sayi	*Myrmentoma* Forel	Portal, AZ, USA	AY334371	AY334451	AY334419	AY334435	AY334385	AY334402
C. nearcticus	*Myrmentoma*	Falmouth, MA, USA	AY334380	AY334462	AY334430	AY334446	AY334396	AY334413
C. floridanus	*Myrmothrix* Forel	Gainesville, FL, USA	AY334381	AY334463	AY334431	AY334447	AY334397	AY334414
Formica L.								
F. fusca		Falmouth, MA, USA					AY334398	AY334415
Polyrhachis F. Smith								
P. foreli	*Myrma*	Townsville, Queensland, Australia	AY336986	AY429041				

TABLE 2. Sequences of PCR primers used to amplify host and symbiont gene regions.

Gene	Primer	Restriction site	Sequence (5′-3′)
16S rDNA	SL[a]	*Bam*HI	TTGGGATCCAGAGTTTGATCATGGCTCAGAT
	SR[a]	*Eco*RI	CACGAATTCTACCTTGTTACGACTTCACCCC
	Bloch16S_462F		AAACCCTGATGCAGCTATACCGTGTGTG
	Bloch16S-1299R		CCATTGTAGCACGTTTGTAGCCCTACTCA
groEL	groES367F	*Eco*RI	CACGAATTCGATATCCTTGCTATCGTTGAG
	groEL1581R	*Bam*HI	TTGGGATCCTCTTTAGGTAATTCAGTAACC
	groEL1645R		ACATCATGCCGCCCATACC
gidA	gidA40F	*Bam*HI	CGCGGATCCATTGGTGGAGGDCATGCWGGCACTGAAGCTGC
	gidA1450R	*Pst*1	AAACTGCAGCTTCTTCATAACCWGTAGTWCCATTAATTTGACC
rpsB	rpsB85F	*Bam*HI	CGCGGATCCCACCARACHCGTTAYTGGAAYCCRMAAATG
	rpsB571R	*Pst*1	AAACTGCAGGTATCAACWATAGMAAATACWGGAATACC
COI/II	C1-J-1754F[b]		CCACGTTTAAATAATATAAGATTTTGAC
	C2-N-3661R[b]		CCACAAATTTCTGAACATTGACCA
	C1F356F		GGATCAGGAACAGGTTGAAC
EF-1αF2	EF1αF2-559F		AGGTGAGCGTGGTATTACGA
	EF1αF2-2111R		CATGGGCTTGCTGGGAAC

[a]Schröder et al., 1996.
[b]Simon et al., 1994.

sites estimated from the data. Rates were allowed to vary among sites according to a gamma distribution for 16S rDNA and introns. All data sets containing protein-coding genes were analyzed using a partitioned-likelihood Bayesian search where each codon position was assigned an independent substitution model similar to those listed previously. First and second coding positions were given a prior uniform shape parameter of the gamma distribution (α range of 0.1–1), and third coding positions and introns were given a range of 5–30. Concatenated data sets were further partitioned by gene origin; nuclear, mitochondrial, or endosymbiont. The parameters for all partitions were estimated independently: unlink shape = (all) revmat = (all) pinvar = (all). Four replicates of each independent and concatenated data set were run, each with four simultaneous and incrementally heated Markov chain Monte Carlo (MCMC) chains, starting from a random tree, for 10,000,000 generations (MrBayes 3.0b4; Ronquist and Huelsenbeck, 2003). The likelihood parameters and trees were sampled every 100 generations. The 50% majority rule consensus and posterior probabilities were calculated from the 50,001 trees sampled after the initial burn-in period and were averaged across the four replicates. To confirm that Bayesian runs had adequate convergence and mixing for a given data set, we tested for similarity of the means and variances of model likelihoods after burn-in across the four independent analyses.

Statistical tests of phylogenetic congruence.—The above analyses generated the "best" ML and Bayesian phylogenies for each of nine data sets (six gene regions and three combined data sets). The ML tree topologies of the various data sets showed slight differences. We tested the significance of these topological differences using the SH test (1999), a nonparametric ML-based test of phylogenetic congruence. This test involved optimizing the ML parameters of each data set across its best ML tree and across alternative topologies (the best ML trees of other data sets). The significance of differences in negative log likelihoods (−lnL) was evaluated using the bootstrap (1,000 replicates) based on RELL sampling and full optimization (PAUP 4.0b10). For each tree topology, the summed likelihood scores of all data sets were calculated to identify the tree with the best cumulative support.

Calculating Rates of Sequence Evolution

Testing for correlated rates of synonymous divergence at host and symbiont genes.—Correlated branch lengths on host and symbiont phylogenies provide further evidence for cospeciation (Hafner and Nadler, 1988). If hosts and symbionts each evolved at constant (but not necessarily identical) rates, then cospeciation would result in a correlation in the lengths of corresponding branches. We calculated rates of synonymous substitution (d_S) along branches for each data set using codeML (PAML 3.13; Yang, 1997) across the complete topology. The codeML parameters allowed omega (ratio of the rates of nonsynonymous to synonymous substitution, or d_N/d_S) to vary across the topology (model = 1) and Ti/Tv was estimated from the data (fix_kappa = 0). We excluded any d_S values along the root branches, values of zero, or values <2, suggesting saturation of synonymous sites. The d_S values at *COI/II* were plotted against the d_S of corresponding branches of each symbiont protein-coding gene. Significance of the correlation was assessed using the Pearson product-moment correlation (r) (JMP 5.0, SAS Institute).

Evolutionary rates on an absolute timescale.—Because *Camponotus* lacks a detailed fossil record, we necessarily took a less direct approach than previous studies of absolute evolutionary rates (Moran et al., 1993; Clark et al., 1999) (see Appendix 1). We used the program r8s (Sanderson, 2003) to estimate divergence dates among *Camponotus* species and then constrained the date of a basal *Camponotus* node to estimate evolutionary rates in *Blochmannia*. First, using codeML, we calculated d_N along branches along a phylogeny of 828 bp of *COI*

available for the relevant taxa. This phylogeny included *Camponotus* and other ants in the subfamily Formicinae, a representative of the subfamily Myrmicinae, and wasp outgroups (Appendix 1, Fig. A1). We constrained the ancestral node between the Formicinae and Myrmicinae to 90–110 million years ago (MYA) based on fossil evidence (Grimaldi and Agosti, 2000; Nascimbene and Silverstein, 2000). From this divergence time and d_N at *COI*, we estimated more recent divergence times, including the basal node of the *Camponotus* species sampled (16.2–19.9 MYA). Next, using r8s we estimated absolute rates of *Blochmannia* sequence evolution based on this inferred date and branch lengths of *Blochmannia* gene trees. These branch lengths were estimated as d_S of protein-coding genes (codeML) or as overall divergences at 16S rDNA using the model of Tamura and Nei (1993) and accounting for rate variation among sites using a gamma distribution with shape 0.5, as implemented by PAUP* (Swofford, 2002).

All estimates of dates and evolutionary rates above were calculated with the semiparametric penalized likelihood (PL) and nonparametric rate smoothing (NPRS) methods implemented by r8s. These smoothing methods are intended to correct for deviations from a molecular clock. PL outperforms NPRS when data are limited and some branches are very short (Sanderson, 2002), which is true of the *COI* data set. Thus, we placed greater emphasis on our results from the PL method.

RESULTS

Sequence Characteristics

New sequence data for host nuclear, mitochondrial, and endosymbiont genomes totaled >7 kb for each of 16 *Camponotus* species representing five subgenera. Protein-coding genes of *Blochmannia* showed a strong AT bias that is consistent with previous estimates of 30% genomic GC content (Dasch, 1975) (Table 3). The AT bias of the mitochondrial genes *COI/II* was consistent with previous analyses of *Camponotus* (30.4% GC at *COI*; Brady et al., 2000), and mitochondrial genes showed a pattern of $GC_1 = GC_2 > GC_3$ observed previously in *Camponotus* and other ant genera (Chenuil and McKey, 1996; Wetterer et al., 1998; Brady et al., 2000). The exceptionally low GC content at silent sites of endosymbiont and mitochondrial genes would be expected if AT mutational pressure shaped base composition but was attenuated at replacement sites by selection on amino acid content (Muto and Osawa, 1987). In contrast to the AT-rich mitochondrial and endosymbiont genomes, the nuclear *EF-1αF2* gene had a more moderate base composition (50.2%).

The *COI/II* sequence derived from *C. vafer* contained three unusual insertions and deletions that may shift the coding frame of both cytochrome oxidases I and II. These changes occurred at *C. vafer* alignment positions 511 and 512 (2-bp insertion), at positions 726–728 (3-bp insertion), and between positions 922 and 923 (4-bp deletion). These sequence results were confirmed by the independent PCR amplification and direct sequencing of an identical product from a separate *C. vafer* isolate. We introduced gaps in the *COI* alignment to account for these indels. The downstream region of *COII*, positions 1929–2117, contained several deletions, and this region was not included in subsequent analyses. Despite apparent frameshifts in *C. vafer COI/II*, the pairwise Ka/Ks between *C. vafer* and *C. festinatus* for the *COI/II* region included was 0.038, a value lower than most pairwise Ka/Ks between other *Camponotus* species at the same gene region (unpubl. data). Thus, if *C. vafer COI* is a pseudogene, it is quite young and has not accumulated mutations at nonsynonymous sites. This finding justifies our inclusion of *C. vafer* in the *COI/II* analysis and assignment of codon positions based on the full alignment.

Blochmannia 16S rDNA genes had moderate base composition (average of 48.7%) and contained putative intervening sequences (IVSs) observed previously (Sauer et al., 2000). Because of alignment ambiguity, we excluded IVS-like regions from phylogenetic analyses. In other bacterial species, IVSs are excised from the rRNA molecule during posttranscriptional modification (Rainey et al., 1996). We identified a region that aligns with an insert designated II (Sauer et al., 2000), which is inserted in *E. coli* 16S rDNA between bp 207–212 and 208–211 and varied from 6 to 48 bp among 15 of the taxa we sampled (data are incomplete for *C. nearcticus*). It was previously reported that both *Wigglesworthia* and *Buchnera* lack a similar insertion (Sauer et al., 2000). However, although *Buchnera* sp. APS lacks the insert (Shigenobu et al., 2000), both copies of 16S rDNA in *W. brevipalpis* contain a 14-bp insert.

Phylogenetic Reconstruction

ML tree topologies were generally congruent across datasets and with phylogenies generated from Bayesian approaches (Fig. 1). For each data set, the four

TABLE 3. Base composition of host and symbiont protein-coding genes. %GC content was calculated at all coding positions (GC_{total}), each codon position (GC_{1-3}), fourfold degenerate sites ($GC_{4\text{-fold}}$), and twofold degenerate sites ($GC_{2\text{-fold}}$). The sequence length from which total %GC contents were calculated is listed.

Gene	GC_{total}	bp	GC_1	GC_2	GC_3	$GC_{4\text{-fold}}$	$GC_{2\text{-fold}}$
16S rDNA	48.75	1339					
rpsB	30.83	423	41.53	32.86	18.01	19.65	12.66
groEL	39.35	1522	58.72	38.91	20.44	16.76	20.62
gidA	34.35	933	45.03	39.61	18.39	18.90	16.71
EF-1αF2 (coding region)	50.16	886	57.29	42.46	50.72	40.16	60.74
COI/II	29.09	1677	34.56	33.99	18.72	19.19	18.47

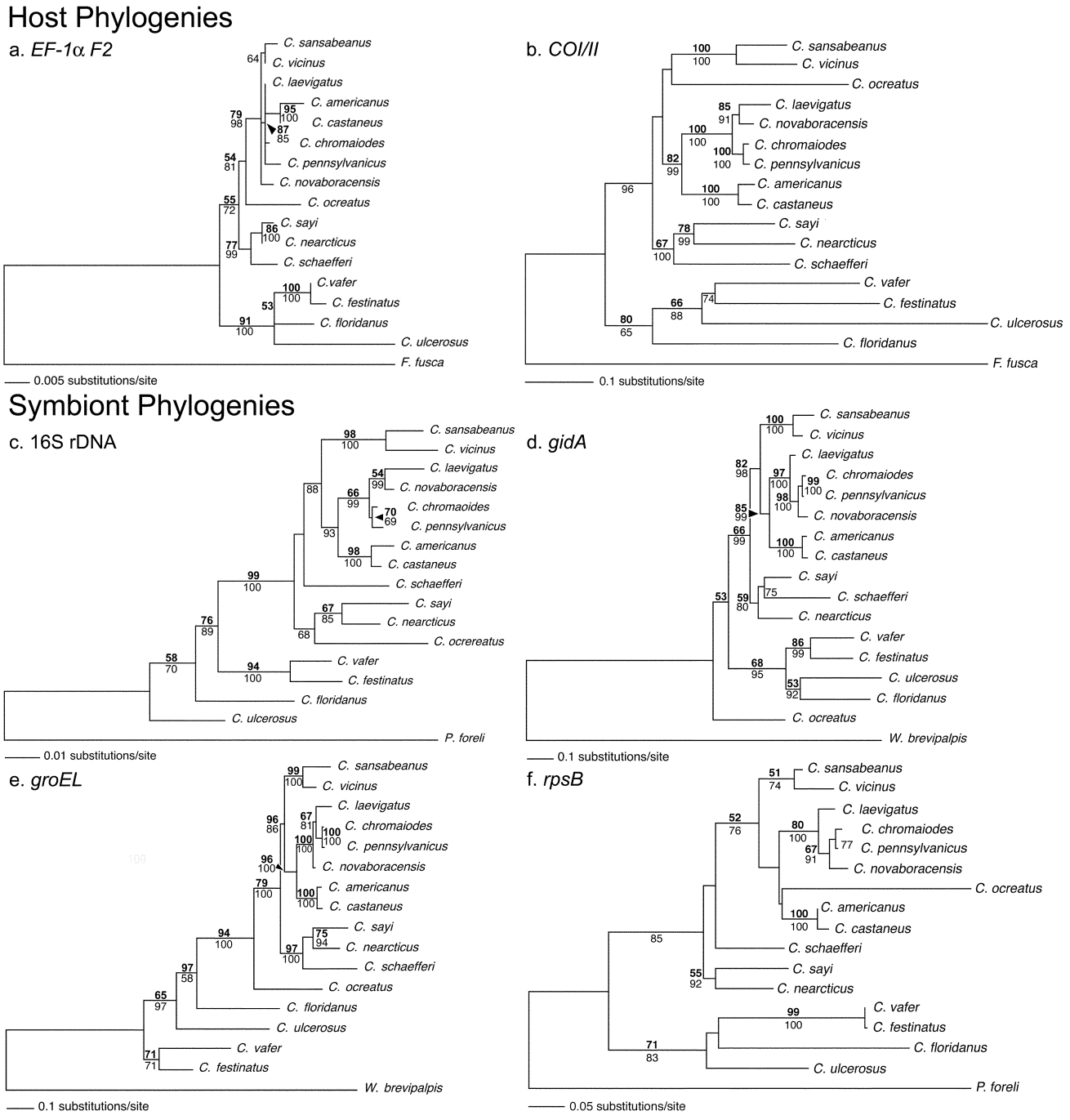

FIGURE 1. Maximum likelihood phylogenies estimated from each host and symbiont gene region. Phylogenetic analysis was performed using substitution models identified by Modeltest: *COI/II*, *gidA* and *groEL*: GTR+I+Γ *EF-1αF2*: GTR+Γ *rpsB* and 16S rDNA: TVM+I+Γ. These models allow for three to six classes of substitution, base frequencies estimated from the data, a proportion of sites to remain invariant (I), and the evolutionary rate of the remaining portion of sites varying according to a gamma distribution (Γ). Bayesian analyses were run using the GTR+I+Γ evolutionary model with all variables estimated from the data. Top numbers (bold) indicate ML bootstrap values, numbers below indicate Bayesian posterior probabilities of shared nodes (>50%), and arrows are used to distinguish support values for closely occurring nodes. Host and symbiont tree topologies consistently support four clades: (1) *C. sansabeanus* and *C. vicinus*, (2) *C. sayi*, *C. nearcticus*, and *C. schaefferi*, (3) *C. vafer, C. festinatus, C. ulcerosus*, and *C. floridanus*, and (4) *C. pennsylvanicus, C. novaboracensis, C. laevigatus, C. chromaiodes, C. americanus*, and *C. castaneus*. *Camponotus ocreatus* was not assigned to a clade because of its topological instability across trees.

TABLE 4. Estimated ML nucleotide substitution parameters for each data set. Initial parameters were identified using Modeltest (Posada and Crandall, 1998) and used in estimations of ML phylogenies of each data set (Fig. 1). Model parameters were then reoptimized across the ML phylogeny of that data set and are presented here. The GTR+I+Γ model (General Time Reversible model accounting for invariant sites and estimating rate variation according to the gamma distribution) was the most appropriate model for each data set.

		Substitution frequencies							
Data set	−lnL	AC	AG	AT	CG	CT	I	α	length (bp)
16S rDNA	4046.60	1.58	17.97	4.49	0.00	21.13	0.675	0.512	1,339
rpsB	2578.19	2.38	4.39	0.28	0.54	11.04	0.427	0.859	423
groEL	8998.35	1.41	6.94	1.45	0.13	5.95	0.557	1.041	1,522
gidA	7539.12	0.89	4.59	0.52	0.33	2.22	0.289	0.982	933
EF-1αF2	3206.17	0.92	6.76	2.38	0.81	14.85	0.394	0.741	1,448
COI/II	10397.15	5.33	23.01	9.55	1.54	107.03	0.454	0.792	1,677
Complete	33529.37	1.55	6.98	1.65	0.35	10.17	0.406	0.489	5,903
ALLBloch	19102.96	1.33	5.39	0.86	0.27	4.51	0.447	0.922	2,841

independent Bayesian runs produced the same topology and very similar posterior probabilities, with ≤3% variation among values of any such node. In addition, the four runs for a given data set had very similar means and variances of total likelihood. For all data sets, the mean of each run fell within the 95% credibility interval of the other three (MrBayes 3.0b4). For a given node, ML bootstrap values were typically lower than Bayesian posterior probability scores, as predicted from modeling approaches (Cummings et al., 2003). Most clades were highly congruent and well supported across trees, and each phylogeny clearly distinguished four major clades (Fig. 1). Any topological differences tended to occur at deep nodes, some of which had relatively low support in these single-gene analyses. ML nucleotide substitution parameters of each data set were reoptimized across the ML phylogeny of that data set. Variation in substitution patterns among data sets illustrated distinct modes of DNA sequence among the loci sampled (Table 4). Subgenera defined by morphological characters were not always monophyletic (Fig. 2), as shown previously (Brady et al., 2000).

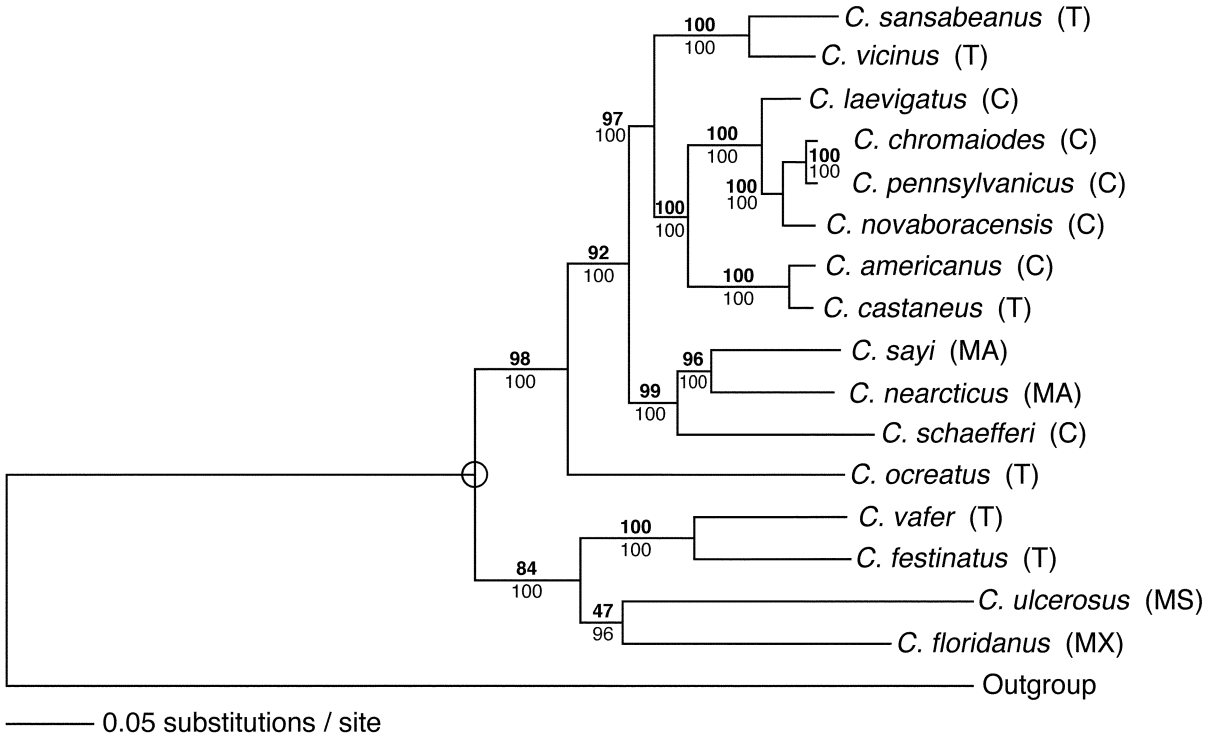

FIGURE 2. Phylogeny of concatenated protein-coding genes of *Camponotus* and *Blochmannia*. ML and Bayesian analyses generated identical trees for the combined host and symbiont protein-coding genes (the complete data set). *Camponotus* subgenera (Bolton, 1995) are indicated: C = *Camponotus*; T = *Tanaemyrmex*; MA = *Myrmentoma*; MS = *Myrmaphaenus*; and MX = *Myrmothrix*. Top numbers (bold) indicate ML bootstrap values and numbers below indicate Bayesian posterior probabilities. Using r8s (Sanderson, 2003), we performed a PL analysis of a *COI* data set specifying that the subfamilies Formicinae and Myrmicinae diverged 90–110 MYA. This analysis dated the node labeled with an open circle at 16.2–19.9 MYA. This date was used for subsequent analyses of rates of evolution at *Blochmannia* genes.

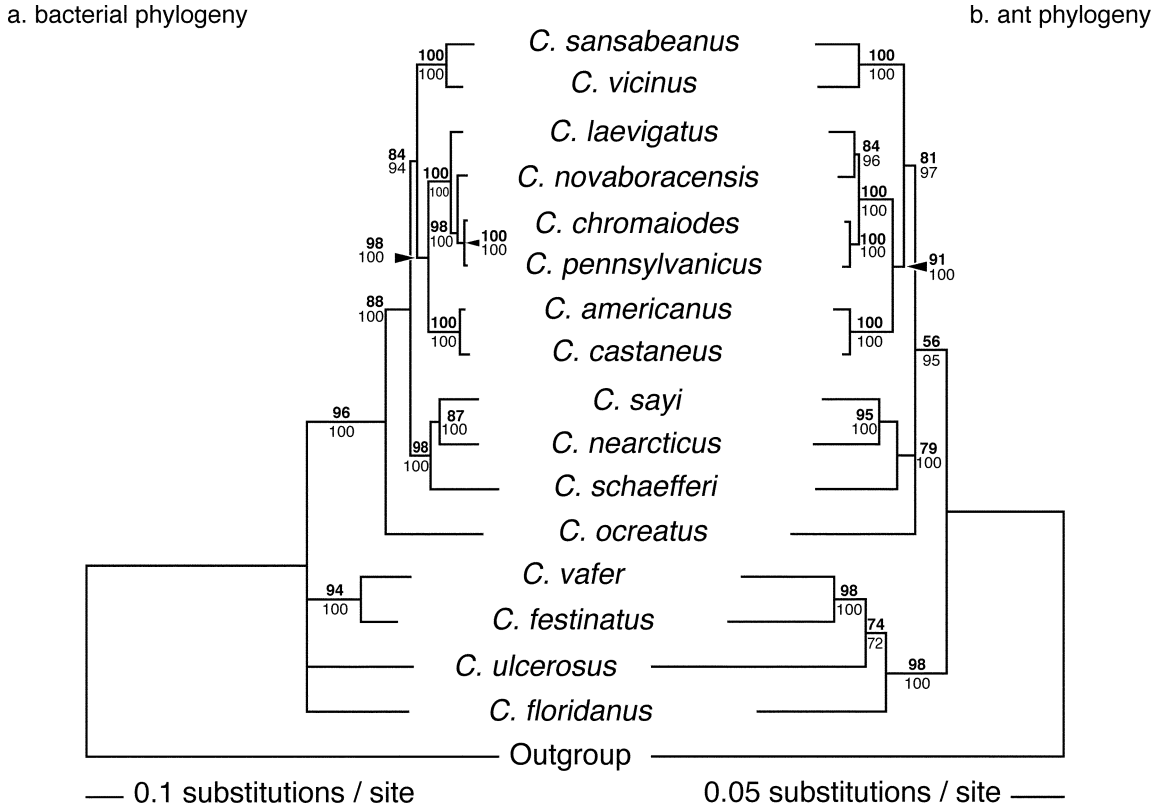

FIGURE 3. Comparison of host and symbiont phylogenies. Trees represent majority rule consensus topologies estimated from the ML analyses of bacterial protein-coding genes (a) and *Camponotus* genes (b). ML and Bayesian analyses used GTR+I+Γ parameters estimated from the data. Top numbers (bold) indicate ML bootstrap values and numbers below indicate Bayesian posterior probabilities.

The combined data set of all protein-coding genes (complete data set) produced the same phylogeny using ML and Bayesian approaches, with bootstrap scores and posterior probabilities generally >84% (Fig. 2). The one exception was the ML bootstrap support of 47% for the grouping of *C. ulcerosus* and *C. floridanus*, although this node had 96% posterior probability. Phylogenies estimated from the concatenated ALLCamp and ALLBloch data sets showed strong congruence at well-supported nodes (Fig. 3).

Statistical Comparison of Alternative Topologies

We compared the ML phylogenies of each data set using the SH test, which unlike the commonly implemented Kishino–Hasegawa test (Kishino and Hasegawa, 1989) accounts for the fact that alternative topologies were estimated from the data (Goldman et al., 2000). We first optimized each data set across its "best" ML tree and across the alternative topologies (i.e., the ML trees estimated from other data sets). The phylogeny estimated from all protein-coding genes (complete data set) had the best cumulative likelihood score, summed for all data sets across that tree ($L_{sum} = 70341.29$; Table 5). We next applied the SH test to evaluate the significance of difference between the –lnL of a given data set across its ML tree and the alternative topologies. We found few cases of significant conflict with alternative topologies, reflected in the scarcity of significant values in Table 6. The ML tree of *EF-1αF2* showed the most conflict with other data sets, but this topology was poorly supported even by the *EF-1α F2* data set because of low variation in this nuclear gene (Table 6; Fig. 1a). The topology of the tree generated by combined protein-coding genes (complete data set, Fig. 2) is identical to that of the tree estimated from all bacterial protein-coding genes (Fig. 3a) and was not rejected by any host or symbiont data set. We consider this tree our best estimate of the shared host and symbiont phylogeny. Although 16S rDNA was not included in the complete data set, the 16S rDNA phylogeny does not significantly disagree with the complete phylogeny.

Rates of Host and Symbiont Evolution

Correlated rates of synonymous substitution.—Branch lengths of host and symbiont trees were compared by plotting d_S values calculated along corresponding branches. The d_S values of each bacterial gene showed a significant correlation with d_S of *COI/II* (*groEL*, $r = 0.685$, $P < 0.0002$; *gidA*, $r = 0.817$, $P < 0.0001$; *rpsB*, $r = 0.435$, $P < 0.038$). *EF-1αF2* was excluded because of extremely low d_S values within *Camponotus*.

Absolute rates of sequence evolution.—Absolute rates of evolution within the subfamily Formicinae were estimated based on fossil evidence that this subfamily

TABLE 5. Negative log likelihoods of trees under GTR+I+Γ model for *Blochmannia* (16S rDNA, *rpsB*, *groEL*, and *gidA*) and *Camponotus* (*EF-1α F2* and *COI/II*) gene regions, and the combined data set of host and symbiont protein-coding genes (the complete data set). For each data set, the ML nucleotide substitution parameters were optimized separately for each tree considered. Specific model parameters for a given data set across its ML tree (underlined) are listed in Table 4. L_{sum} = summed −lnL values across all seven data sets constrained to a given topology. The lowest (best) likelihood scores for each data set and the best L_{sum} are underlined. The Complete topology has the best score for two of the seven data sets and the best L_{sum} (70341.29). The significance of differences in −lnL of a given data set topology was evaluated using the SH test (see Table 6).

	Data set							
Topology	16S rDNA	rpsB	*groEL*	*gidA*	*EF-1αF2*	*COI/II*	Complete	L_{sum}
16S rDNA	4046.60	2590.89	9035.36	7568.29	3232.18	10419.31	33633.75	70526.37
rpsB	4065.78	2578.19	9043.89	7559.10	3233.37	10426.93	33612.31	70519.57
groEL	4062.09	2586.73	8998.35	7560.16	3220.18	10416.21	33569.60	70413.32
gidA	4073.00	2588.25	9020.41	7540.66	3225.29	10413.87	33562.22	70423.71
EF-1αF2	4085.52	2620.28	9152.72	7680.07	3206.17	10554.78	34032.40	71331.94
COI/II	4059.39	2581.55	9020.95	7569.32	3221.61	10397.15	33577.83	70427.81
Complete	4059.99	2583.05	9008.76	7539.12	3213.27	10407.72	33529.37	70341.29

diverged from the ant subfamily Myrmicinae approximately 90–110 MYA (Grimaldi and Agosti, 2000). Using the PL method in r8s, we estimated rates of nonsynonymous substitution at an 828-bp region of *COI* to be 0.000425 ± 6.5e-05 substitutions per site per million years (s/s/MY) (assuming the subfamilies diverged 110 MYA) or 0.000521 ± 8.3e-05 s/s/MY (assuming they diverged 90 MYA) (Appendices 1, 3). This rate is faster than rates previously calculated across several insect orders (0.00022 s/s/MY at second codon positions of *COI*; Gaunt and Miles, 2002) but is consistent with elevated rates of mitochondrial evolution in Hymenoptera (e.g., Crozier and Crozier, 1992). The same r8s analysis estimated divergence times within the Formicinae. The ancestral node of all ants known to have *Blochmannia* was dated at ∼29.3–35.9 MYA (Appendix 1 Fig. A1), suggesting the *Blochmannia*–ant association is at least 30 MY old and perhaps much older if the symbiont once infected other genera but was subsequently lost from them. The divergence time of *Camponotus* included here was estimated at ∼16.2–19.9 MYA (Fig. 2).

From this estimated divergence time, we used r8s to calculate absolute rates of synonymous substitution at *Blochmannia gidA* and *groEL* and total substitution rates at 16S rDNA (Table 7). Rates of synonymous substitution at *Blochmannia gidA* and *groEL* ranged from 0.0854 to 0.1372 s/s/MY, depending on the gene and the date at which the basal *Camponotus* node was constrained (16.2 or 19.9 MYA). Compared with *Buchnera*, rates in *Blochmannia* were about 14 times faster at 16S rDNA and 5–10 times faster at synonymous sites of *gidA* and *groEL*. Compared with *E. coli, Blochmannia* apparently evolves about 29 times faster at the 16S rDNA gene (Table 7, Appendix 4). Direct comparisons of d_S values are difficult to interpret because of codon bias in *E. coli* at the high-expression genes sampled. Although the NPRS method gave slightly slower rate estimates than did PL (Table 7, Appendix 3), both methods suggested faster absolute rates in *Blochmannia* than in *Buchnera* and *E. coli*.

DISCUSSION

Obligate endosymbionts of insects, like most of the microbial world, cannot be cultured in the lab and have thus proved difficult to study with traditional genetic or physiological techniques. Full genome sequencing has illustrated the mechanisms and consequences of severe genome size reduction among obligate bacterial associates, including their extreme specialization for the host cellular environment and the retention of host-beneficial genes (e.g., Shigenobu et al., 2000; Tamas et al., 2001; Akman et al., 2002; Moran, 2002; Gil et al., 2003). Furthermore, molecular phylogenetic analysis has revealed that symbiont dynamics range from strict cospeciation of P endosymbionts and their hosts (Chen et al., 1999; Thao et al., 2000b), to the frequent host switching of S endosymbionts (Fukatsu et al., 2000; Thao et al., 2000a;

TABLE 6. Results of SH tests of alternative hypotheses for host and endosymbiont phylogenies. Differences in −lnL scores were calculated for each data set across the ML topology for that data set and the ML topologies of other data sets. The significance of these −lnL differences was measured using full optimization. An asterisk indicates significant conflict between a given data set and the topology tested ($P < 0.05$). Topologies that were not rejected by a given data set (i.e., nonsignificant differences in −lnL) are underlined.

	Data set								
Topology	16S rDNA	*rpsB*	*groEL*	*gidA*	*EF-1αF2*	*COI/II*	Complete	ALLBloch	ALLCamp
16S rDNA	(best)	0.2480	0.0760	0.1070	0.0520	0.2500	0.0150*	0.0390*	0.1420
rpsB	0.1230	(best)	0.0190*	0.2560	0.0360*	0.1430	0.0370*	0.1170	0.1140
groEL	0.1600	0.4120	(best)	0.2230	0.1670	0.2960	0.2470	0.3890	0.2620
gidA	0.0390*	0.3350	0.2540	0.8150	0.1400	0.3770	0.3770	0.5270	0.3240
EF-1αF2	0.0150*	0.0110*	0.0000*	0.0000*	(best)	0.0000*	0.0000*	0.0000*	0.0000*
COI/II	0.2910	0.7740	0.2030	0.0810	0.1720	(best)	0.1710	0.1140	0.8360
Complete/ALLBloch	0.2450	0.6600	0.5900	(best)	0.5980	0.5530	(best)	(best)	0.7280
ALLCamp	0.5500	0.7730	0.2640	0.0980	0.8080	0.8970	0.2090	0.1360	(best)

TABLE 7. Comparison of evolutionary rates in *Blochmannia*, *Buchnera*, and the enterobacteria at 16S rDNA and at synonymous sites of protein-coding genes. Rates were calculated at homologous gene regions and expressed as substitutions/site/million years (s/s/MY). For *Blochmannia*, rates of sequence evolution (±SE) were calculated across the *Camponotus* phylogeny using penalized likelihood (PL) and nonparametric rate smoothing (NPRS) methods implemented by the program r8s (Sanderson, 2003), assuming a divergence time of 19.9 MYA or 16.2 MYA for the basal node of *Camponotus* spp. considered (circled node in Fig. 2). Rates at homologous gene regions of *Buchnera* were calculated from pairwise sequence divergences between *Buchnera* from *Acyrthosiphon pisum* (Ap) and *Buchnera* from *Schizaphis graminum* (Sg) (50–70 MYA), and rates in the enterics were calculated from pairwise divergence between *E. coli* (Ec) and *Salmonella typhimurium* (St) (100–150 MYA). For *Buchnera* and the enterics, the range of rates reflects uncertainty in the divergence times of bacterial lineages and/or their insect hosts.

	Blochmannia			
	19.9 MYA	16.2 MYA	*Buchnera* (Ap–Sg)	Enterics (Ec–St)
16S rDNA	$0.00216^{PL} \pm 0.000362$	$0.002674^{PL} \pm 0.000422$	0.000148–0.000208	0.000066–0.000100
	$0.001062^{NPRS} \pm 0.000285$	$0.001298^{NPRS} \pm 0.000348$		
gidA	$0.08535^{PL} \pm 0.01069$	$0.1027^{PL} \pm 0.01025$	0.014531–0.020344	0.002663–0.003994
groEL	$0.1123^{PL} \pm 0.004025$	$0.1372^{PL} \pm 0.00359$	0.011290–0.015806	0.000794–0.001192

Fukatsu, 2001; Sandström et al., 2001; Russell et al., 2003) and reproductive parasites such as *Wolbachia* (Werren et al., 1995). Distinguishing among these transmission patterns has important implications for understanding the evolutionary stability, approximate age, and potential for coevolution within a particular association.

The aim of this study was to test for cospeciation between the endosymbiont *Blochmannia* and its *Camponotus* hosts by expanding the available sequence data sets and by using statistically reliable approaches to estimate and compare phylogenies. The restriction of this bacterium to bacteriocytes and the female reproductive tissues suggests that speciation of the host and symbiont are linked (Schröder et al., 1996; Sauer et al., 2002). Although lineage sorting can disrupt phylogenetic congruence (e.g., Hardig et al., 2000), it is unlikely to operate in this system. Our PCR analysis of multiple host species has not detected more than a single *Blochmannia* strain per host. However, the ability to eliminate this symbiont from lab-reared ants without detectable host effects (Sauer et al., 2002) and previous suggestions of host switching (Sauer et al., 2000) suggest that *Blochmannia* may form a more labile association than is typical of other P endosymbionts.

Prior to this study, phylogenetic analyses had not accounted for rate variation across loci and were based on relatively small data sets. A molecular phylogenetic analysis of *Camponotus* based on 385 bp of *COI* showed that certain subgenera might not be monophyletic, suggesting homoplasy of the morphological characters on which subgenera are based (pilosity and head shape; Brady et al., 2000). We sampled a nuclear gene of *Camponotus* and an extended region of mitochondrial genes (*COI*/*II*) often used in insect systematics (e.g., Rokas et al., 2002). Our analyses also indicate that certain *Camponotus* subgenera are not monophyletic. Inconsistencies we observed include the placement of *C. schaefferi* (subgenus *Camponotus*) with species of the subgenus *Myrmentoma* and the close relationship of *C. castaneus* (subgenus *Tanaemyrmex*) to *C. americanus* (subgenus *Camponotus*). Most notably, the subgenus *Tanaemyrmex* apparently does not represent a monophyletic group, as observed previously (Brady et al., 2000).

Congruence between the phylogenies of host data sets and those of endosymbiont genes indicate that stable vertical transmission of symbionts to host offspring leads to bacterium–ant cospeciation. The phylogeny based on all protein-coding genes has the best cumulative likelihood score, is identical to the that estimated from bacterial protein-coding genes, and is consistent with the host data sets. In addition, correlations in d_S values in host mitochondrial and endosymbiont phylogenies indicate similar levels of synonymous changes along branches for these two linked, maternally inherited genomes. These branch length correlations support the hypothesis of cospeciation but are not required by it. Even if cospeciation were true, variable evolutionary rates in either the host or symbiont lineages could disrupt the correlation in rates (Hafner and Nadler, 1988). Phylogenetic congruence among *Blochmannia* genes sampled here illustrates a lack of horizontal gene transfer and contrasts with frequent recombination among free-living bacterial strains (Dykhuizen and Green, 1991; Ochman et al., 2000). This apparent genome stability of *Blochmannia* is similar to that observed in *Buchnera* and *Wigglesworthia*, which show no evidence of lateral gene transfer (Funk et al., 2000; Wernegreen and Moran, 2001; Akman et al., 2002; Tamas et al., 2002).

Hypotheses other than cospeciation can explain topological congruence of host and symbiont phylogenies. For example, geographic effects may generate this pattern when hosts within a given geographic region tend to associate with symbionts in the same location. Our sample of *Camponotus* included both closely related species from distinct geographic regions and, conversely, phylogenetically diverse species from the same field site (often adjacent to the other species' nests). Thus, observed topological congruence in this study cannot be explained by a tendency for co-occurring host and symbionts to form associations. Moreover, geographic effects would not necessarily predict the observed correlation of host and symbiont branch lengths.

Establishing host–symbiont cospeciation opens the possibility of calibrating a symbiont clock based on host divergence times. Molecular clock estimates for *Buchnera* have been calibrated against a host fossil record and provide the most reliable clock available for bacteria (Moran et al., 1993; Clark et al., 1999). At synonymous sites of low codon bias genes, *Buchnera* evolves about

twice as fast as *E. coli* and *Salmonella typhimurium* on an absolute timescale (Clark et al., 1999). This rate elevation might result from higher mutation rates in small endosymbiont genomes that lack many repair genes (Eisen and Hanawalt, 1999; Moran and Wernegreen, 2000), and/or accelerated fixation of slightly deleterious mutations due to increased genetic drift in small populations (Moran, 1996; Funk et al., 2001; Woolfit and Bromham, 2003). Here, we calculated rates of sequence evolution in *Blochmannia* from inferred divergence times of its *Camponotus* hosts. At homologous gene regions, *Blochmannia* evolves surprisingly faster than either *Buchnera* or the enterobacteria. Our results suggest that *Blochmannia* 16S rDNA evolves about 14 times faster than that of *Buchnera* and about 29 times faster than that of the enterics (represented here by *E. coli* and *S. typhimurium*). At synonymous sites of *gidA* and *groEL*, rates in *Blochmannia* are ~5–10 times higher than those in *Buchnera*. Previous studies have interpreted high divergence among *Blochmannia* strains to mean that this ant–bacterium association is very old and perhaps an original attribute of ants that was subsequently lost in most ant lineages (Schröder et al., 1996; Boursaux-Eude and Gross, 2000). However, host–symbiont cospeciation means this high divergence occurred in the context of its association with *Camponotus*. Although our rate estimates may be error prone because of uncertain divergence times among *Camponotus* hosts, *Blochmannia* apparently evolves many times faster than *Buchnera*. For example, for the 16S rDNA, applying evolutionary rates for *Buchnera* (inferred here or by other investigators; Clark et al., 2000) to *Blochmannia* sequence divergences would predict that the *Blochmannia–Camponotus* association is >100 MY old, which far predates the inferred origin of the subfamily Formicinae (Grimaldi and Agosti, 2000). This evidence for fast evolutionary rates in *Blochmannia* would be strengthened by comparing rates across other genes and by better fossil data to assign dates to nodes within the *Camponotus* phylogeny.

ACKNOWLEDGMENTS

We thank Stefan Cover and Bill MacKay for their assistance with ant identification and for sharing their insights, along with those of Gary Alpert, Brian Fisher, and Daniel Hahn, into *Camponotus* ecology. We thank Stefan Cover, Mark Seid, Chris Starkweather, and Simon Robson for contributing insect specimens. We are grateful to Josh Herbeck, Andrew McArthur, and Seth Bordenstein for critical insights and comments on the manuscript and to Seán Brady for helpful discussion of *Camponotus* systematics. This article benefited greatly from the constructive suggestions of Carol von Dohlen, Kevin Johnson, Chris Simon, and an anonymous reviewer. We also thank Kevin Johnson and Dale Clayton for organizing the 2002 SSB Symposium "Untangling Coevolutionary History" at the SSB/SSE annual meeting in Champaign, Illinois, where some of these results were presented. Funding for this research was provided by the NIH (R01 GM62626-01), the NSF (DEB 0089455 and associated REU supplement award), the NASA Astrobiology Institute (NCC2-1054), and the Josephine Bay Paul and C. Michael Paul Foundation.

REFERENCES

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

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, P., N. Moran, and L. Baumann. 2000. Bacteriocyte-associated endosymbionts of insects. *In* The prokaryotes, a handbook on the biology of bacteria; ecophysiology, isolation, identification, applications (M. Dworkin, ed.), 3rd edition, release 3.1. Springer-Verlag, New York (http://link.springer-ny.com/link/service/books/10125).

Bergthorsson, U., and H. Ochman. 1998. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol. Biol. Evol. 15:6–16.

Bolton, B. 1995. Identification guide to the ant genera of the world. Harvard Univ. Press, Cambridge, Massachusetts.

Boursaux-Eude, C., and R. Gross. 2000. New insights into symbiotic associations between ants and bacteria. Res. Microbiol. 151:513–519.

Brady, S. G., J. Gadau, and P. S. Ward. 2000. Systematics of the ant genus *Camponotus* (Hymenoptera: Formicidae): A preliminary analysis using data from the mitochondrial gene cytochrome oxidase I. Pages 131–139 *in* Hymenoptera. Evolution, biodiversity and biological control (A. D. Austin and M. Dowton, eds.). CSIRO, Collingwood, Victoria, Australia.

Brooks, D. 1981. Hennig's parasitological method: A proposed solution. Syst. Zool. 30:229–249.

Buchner, P. 1965. Endosymbiosis of animals with plant microorganisms. Interscience, New York.

Charles, H., A. Heddi, and Y. Rahbe. 2001. A putative insect intracellular endosymbiont stem clade, within the Enterobacteriaceae, inferred from phylogenetic analysis based on a heterogeneous model of DNA evolution. C.R. Acad. Sci. III 324:489–494.

Chen, X., 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.

Cheng, Q., and S. Aksoy. 1999. Tissue tropism, transmission and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies. Insect Mol. Biol. 8:125–132.

Chenuil, A., and D. B. McKey. 1996. Molecular phylogenetic study of a myrmecophyte symbiosis: Did *Leonardoxa*/ ant associations diversify via cospeciation? Mol. Phylogenet. Evol. 6:270–286.

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 (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution 54:517–525.

Crozier, R. H., and Y. C. Crozier. 1992. The cytochrome *b* and AT-Pase genes of honeybee mitochondrial DNA. Mol. Biol. Evol. 9:474–482.

Cummings, M. P., S. A. Handley, D. S. Myers, D. L. Reed, A. Rokas, and K. Winka. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. Syst. Biol. 52:477–487.

Danforth, B. N., and S. Ji. 1998. Elongation factor-1 alpha occurs as two copies in bees: Implications for phylogenetic analysis of *EF-1* alpha sequences in insects. Mol. Biol. Evol. 15:225–235.

Dasch, G. A. 1975. Morphological and molecular studies on intracellular bacterial symbiotes of insects. Yale Univ. Press, New Haven, Connecticut.

Dasch, G., E. Weiss, and K. Chang. 1984. Endosymbionts of insects. Pages 811–833 *in* Bergy's manual of systematic bacteriology (J. Holt and N. Krieg, eds.). Williams & Wilkins, Baltimore.

Davidson, D. 1997. The role of resource imbalances in the evolutionary ecology of tropical arboreal ants. Biol. J. Linn. Soc. Lond. 61:153–181.

Davidson, D. 1998. Resource discovery versus resource domination in ants: A functional mechanism for breaking the trade-off. Ecol. Entomol. 23:484–490.

Davidson, D. W., S. C. Cook, R. R. Snelling, and T. H. Chua. 2003. Explaining the abundance of ants in lowland tropical rainforest canopies. Science 300:969–972.

Douglas, A. E. 1989. Mycetocyte symbiosis in insects. Biol. Rev. Camb. Philos. Soc. 64:409–434.
Douglas, A., and W. Prosser. 1992. Synthesis of the essential amino acid tryptophan in the pea aphid (*Acyrthosiphon pisum*) symbiosis. J. Insect Physiol. 38:565–568.
Dykhuizen, D. E., and L. Green. 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. 173:7257–7268.
Eisen, J. A., and P. C. Hanawalt. 1999. A phylogenomic study of DNA repair genes, proteins, and processes. Mutat. Res. 435:171–213.
Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1994. Testing significance of congruence. Cladistics 10:315–320.
Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.
Fukatsu, T. 2001. Secondary intracellular symbiotic bacteria in aphids of the genus *Yamatocallis* (Homoptera: Aphididae: Drepanosiphinae). Appl. Environ. Microbiol. 67:5315–5320.
Fukatsu, T., N. Nikoh, R. Kawai, and R. Koga. 2000. The secondary endosymbiotic bacterium of the pea aphid *Acyrthosiphon pisum* (Insecta: Homoptera). Appl. Environ. Microbiol. 66:2748–2758.
Funk, D. J., L. Helbling, J. J. Wernegreen, and N. A. Moran. 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. Proc. R. Soc. Lond. B Biol. Sci. 267:2517–2521.
Funk, D. J., J. J. Wernegreen, and N. A. Moran. 2001. Intraspecific variation in symbiont genomes: Bottlenecks and the aphid–*Buchnera* association. Genetics 157:477–489.
Gaunt, M. W., and M. A. Miles. 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol. Biol. Evol. 19:748–761.
Gil, R., B. Sabater-Munoz, A. Latorre, F. J. Silva, and A. Moya. 2002. Extreme genome reduction in *Buchnera* spp.: Toward the minimal genome needed for symbiotic life. Proc. Natl. Acad. Sci. USA 99:4454–4458.
Gil, R., F. J. Silva, E. Zientz, F. Delmotte, F. Gonzalez-Candelas, A. Latorre, C. Rausell, J. Kamerbeek, J. Gadau, B. Holldobler, R. C. van Ham, R. Gross, and A. Moya. 2003. The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. Proc. Natl. Acad. Sci. USA 100:9388–9393.
Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49:652–670.
Grimaldi, D., and D. Agosti. 2000. A formicine in New Jersey Cretaceous amber (Hymenoptera: Formicidae) and early evolution of the ants. Proc. Natl. Acad. Sci. USA 97:13678–13683.
Hafner, M. S., and S. A. Nadler. 1988. Phylogenetic trees support the coevolution of parasites and their hosts. Nature 332:258–259.
Hardig, T. M., P. S. Soltis, and D. E. Soltis. 2000. Diversification of the North American shrub genus *Ceanothus* (Rhamnaceae): Conflicting phylogenies from nuclear ribosomal DNA and chloroplast DNA. Am. J. Bot. 87:108–123.
Hölldobler, B., and E. O. Wilson. 1990. The ants. Belknap Press, Cambridge, Massachusetts.
Huelsenbeck, J. P., B. Rannala, and B. Larget. 2000. A Bayesian framework for the analysis of cospeciation. Evolution 54:353–364.
Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.
Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.
Lawrence, J. G., and H. Ochman. 2002. Reconciling the many faces of lateral gene transfer. Trends Microbiol. 10:1–4.
Lo, N., C. Bandi, H. Watanabe, C. Nalepa, and T. Beninati. 2003. Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. Mol. Biol. Evol. 20:907–913.
Maddison, D. R., and W. P. Maddison. 2002. MacClade 4: Analysis of phylogeny and character evolution, version 4.05. Sinauer, Sunderland, Massachusetts.
Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt, and J. M. Tiedje. 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. 29:173–174.
Moran, N. A. 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA 93:2873–2878.
Moran, N. A. 2002. Microbial minimalism: Genome reduction in bacterial pathogens. Cell 108:583–586.
Moran, N. A., and P. Baumann. 2000. Bacterial endosymbionts in animals. Curr. Opin. Microbiol. 3:270–275.
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 J. J. Wernegreen. 2000. Lifestyle evolution in symbiotic bacteria: Insights from genomics. Trends Ecol. Evol. 15:321–326.
Muse, S. V., and B. S. Weir. 1992. Testing for equality of evolutionary rates. Genetics 132:269–276.
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.
Nascimbene, P., and H. Silverstein. 2000. The preparation of fragile Cretaceous ambers for conservation and study of organismal inclusions. Pages 93–102 *in* Studies on fossils in amber, with particular reference to the Cretaceous of New Jersey (D. Grimaldi, ed.). Backhuys, Leiden.
Nogge, G. 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology 82:299–304.
Ochman, H., S. Elwyn, and N. A. Moran. 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. USA 96:12638–12643.
Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304.
Ochman, H., and A. C. Wilson. 1987. Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. 26:74–86.
Oliver, K. M., J. A. Russell, N. A. Moran, and M. S. Hunter. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc. Natl. Acad. Sci. USA 100:1803–1807.
Posada, D., and K. A. Crandall. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817–818.
Rainey, F. A., N. L. Ward-Rainey, P. H. Janssen, H. Hippe, and E. Stackebrandt. 1996. *Clostridium paradoxum* DSM 7308T contains multiple 16S rRNA genes with heterogeneous intervening sequences. Microbiology 142:2087–2095.
Rokas, A., J. A. Nylander, F. Ronquist, and G. N. Stone. 2002. A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): Implications for insect phylogenetic studies. Mol. Phylogenet. Evol. 22:206–219.
Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
Russell, J. A., A. Latorre, B. Sabater-Munoz, A. Moya, and N. A. Moran. 2003. Side-stepping secondary symbionts: Widespread horizontal transfer across and beyond the Aphidoidea. Mol. Ecol. 12:1061–1075.
Sameshima, S., E. Hasegawa, O. Kitade, N. Minaka, and T. Matsumoto. 1999. Phylogenetic comparison of endosymbionts with their host ants based on molecular evidence. Zool. Sci. 16:993–1000.
Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.
Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Mol. Biol. Evol. 19:101–109.
Sanderson, M. J. 2003. r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.
Sandström, J. P., J. A. Russell, J. P. White, and N. A. Moran. 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. Mol. Ecol. 10:217–228.
Sauer, C. 2000. Charakterisierung intrazellulärer, bakterieller Endosymbionten im Mitteldarm von Ameisen der Gattung *Camponotus*. Page 112 *in* Biozentrum Universität Würzburg. Würzburg, Germany.
Sauer, C., D. Dudaczek, B. Hölldobler, and R. Gross. 2002. Tissue localization of the endosymbiotic bacterium "*Candidatus* Blochmannia floridanus" in adults and larvae of the carpenter ant *Camponotus floridanus*. Appl. Environ. Microbiol. 68:4187–4193.
Sauer, C., E. Stackebrandt, J. Gadau, B. Hölldobler, and R. Gross. 2000. Systematic relationships and cospeciation of

bacterial endosymbionts and their carpenter ant host species: Proposal of the new taxon *Candidatus* Blochmannia gen. nov. Int. J. Syst. Evol. Microbiol. 50:1877–1886.
Schröder, D., H. Deppisch, M. Obermayer, 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. 21:479–489.
Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81–86.
Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.
Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 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.
Spaulding, A. W., and C. D. von Dohlen. 2001. Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing substitutions in ribosomal RNA. Insect Mol. Biol. 10:57–67.
Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, Massachusetts.
Tamas, I., L. Klasson, B. Canback, A. K. Naslund, A. S. Eriksson, J. J. Wernegreen, J. P. Sandström, N. A. Moran, and S. G. Andersson. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science 296:2376–2379.
Tamas, I., L. M. Klasson, J. P. Sandström, and S. G. Andersson. 2001. Mutualists and parasites: How to paint yourself into a (metabolic) corner. FEBS Lett. 498:135–139.
Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512–526.
Thao, M. L., M. A. Clark, L. Baumann, E. B. Brennan, N. A. Moran, and P. Baumann. 2000a. Secondary endosymbionts of psyllids have been acquired multiple times. Curr. Microbiol. 41:300–304.
Thao, M. L., N. A. Moran, P. Abbot, E. B. Brennan, D. H. Burckhardt, and P. Baumann. 2000b. Cospeciation of psyllids and their primary prokaryotic endosymbionts. Appl. Environ. Microbiol. 66:2898–2905.
Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.
van Ham, R. C., J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J. M. Fernandez, L. Jimenez, M. Postigo, F. J. Silva, J. Tamames, E. Viguera, A. Latorre, A. Valencia, F. Moran, and A. Moya. 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. USA 100:581–586.
Wernegreen, J. J., P. H. Degnan, A. B. Lazarus, C. Palacios, and S. R. Bordenstein. 2003. Genome evolution in an insect cell: Distinct features of an ant–bacterial partnership. Biol. Bull. 204:221–232.
Wernegreen, J. J., A. B. Lazarus, and P. H. Degnan. 2002. Small genome of *Candidatus* Blochmannia, the bacterial endosymbiont of *Camponotus*, implies irreversible specialization to an intracellular lifestyle. Microbiology 148:2551–2556.
Wernegreen, J. J., and N. A. Moran. 2001. Vertical transmission of biosynthetic plasmids in aphid endosymbionts (*Buchnera*). J. Bacteriol. 183:785–790.
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. 261:55–63.
Wetterer, J. K., T. R. Schultz, and R. Meier. 1998. Phylogeny of fungus-growing ants (tribe Attini) based on mtDNA sequence and morphology. Mol. Phylogenet. Evol. 9:42–47.
Woolfit, M., and L. Bromham. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20:1545–1555.
Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.

Final acceptance 12 September 2003
Associate Editor: Kevin Johnson

APPENDIX 1: ONLINE RESOURCES

PCR Amplification Strategies

A fragment of the 16S rDNA gene was amplified using the SL and SR universal eubacterial primers (Schroder et al., 1996) often in combination with a *Blochmannia*-specific 16S rDNA primer to generate two overlapping PCR fragments. A gene fragment spanning *COI/II* was generated using the C1-J-1754F and C2-N-3661R primers (Simon et al., 1994) for most isolates. The only exception was *C. sansabeanus*, for which a new forward *COI* primer was designed to create a slightly smaller fragment. The nuclear *EF-1α* genes occur in two copies (F1 and F2) in Hymenoptera (Danforth and Ji, 1998). We first amplified both copies using primers designed from the *EF-1α* genes of *Apis mellifera* (AF015267). Both *EF-1α* copies were TA cloned and sequenced from two divergent *Camponotus* species (*C. pennsylvanicus* and *C. festinatus)* to design PCR primers specific to the *Camponotus* F1 and F2 copies. *EF-1αF2* was selected for analysis because of its slightly larger size (reflecting a greater number of introns). Primers for *Blochmannia* genes *rpsB*, *gidA*, and *groEL* were designed from alignments of enterobacteria and *Buchnera* sequences available in GenBank. Combined, these gene fragments totaled ~7,342 bp of high quality sequence data per isolate sampled (Appendix 2).

When PCR failed to yield sufficient template for direct sequencing, fragments were cloned into plasmid vectors using the TOPO TA cloning kit (Invitrogen) and electrocompetent cells according to the manufacturer's instructions. Purified plasmid clones were prepared using the Qiaquick spin miniprep kit (Qiagen) or the RevOrbit alkaline lysis 96-well plate purification system (Genomic Solutions). Clones containing inserts were screened prior to sequencing, using both restriction digests with *Eco*RI and *Not*I (New England Biolabs) and PCR amplification with M13F and M13R primers.

Estimating Absolute Rates of Sequence Evolution: Details of the Approach and Calculations

The goal of this analysis was to estimate rates of nucleotide substitution in *Blochmannia* on an absolute timescale (as substitutions per site per million years: s/s/MY) for comparison with absolute rate estimates for *Buchnera* and the enterics (Moran et al., 1993; Clark et al., 1999; Ochman et al., 1999). To estimate the rates of substitution in *Blochmannia*, we would ideally assign the dates of divergence for their *Camponotus* host lineages. However, because *Camponotus* itself lacks a detailed fossil record, we necessarily took a less direct approach that involved the following steps.

Estimating rates for the subfamily Formicinae on an absolute timescale.—We first estimated sequence divergence between the subfamilies Myrmicinae and Formicinae, which diverged 90–110 MYA according to fossil evidence (Grimaldi and Agosti, 2000). Divergences were calculated at nonsynonymous sites of the mitochondrial gene *COI*, the locus that provided the most robust molecular clock in a previous study of several insect orders (Gaunt and Miles, 2002). Two regions of *COI* were selected based on the availability of myrmicine sequences in GenBank: bp 1816–2214 (region A) and bp 2254–2652 (region B) (positions correspond to genome sequence of the *Drosophila melanogaster* mitochondrion, NC001709). Among myrmicine species, region A was available for the *Crematogaster borneensis* and region B was available for *Leptothorax recedens*. Relative rates tests (calculated by the method of Muse and Weir, 1992) showed no significant difference in rates of nonsynonymous substitution in formicines and myrmicines at either region of *COI* (A or B) when either Vespidae or Scoliidae were used as outgroups. We then calculated an evolutionary rate on an absolute timescale for each of the regions A and B by averaging the pairwise estimates of d_N (PAML; Yang, 1997) between the single myrmicine representative and formicine *COI* sequences and then dividing by the inferred divergence time (90–110 MYA). The rates based on regions A and B both performed similarly; they both estimated comparable divergence times of *Colobopsis* and *Camponotus*. Given the lack of conflict between the dates estimated from regions A (*Crematogaster*) and B (*Leptothorax*), we combined these regions for the rates analysis below.

The r8s program (Sanderson, 2003) implements several methods and algorithms for estimation of absolute rates of molecular evolution

and divergence times across a phylogeny. The penalized likelihood (PL) method was most appropriate for the *COI* data set, but the nonparametric rate smoothing (NPRS) method was also used for comparison of results (Appendix 3). All analyses assumed a divergence time of 90–110 MYA between Formicinae and Myrmicinae. These values were independently fixed (using the "fixage" command in r8s) to the shared node representing the most recent common ancestor of Formicinae and Mymicinae (Fig. A1). Cross-validation analyses of the PL method were used to optimize the smoothing parameters for each age, 90 and 110 MYA, using both the Powell and truncated Newton (TN) algorithms. Calculations were rerun using the optimal smoothing parameter (110 MYA: smoothing = 150; 90 MYA: smoothing = 100), which estimated rates at nonsynonymous sites of this *COI* region to be 0.000425 ± 6.5e-05 s/s/MY and 0.000521 ± 8.3e-05 s/s/MY, respectively. NPRS was run using the Powell algorithm and estimated a slower rates of 0.000251 ± 2.3e-04 and 0.000306 ± 2.8e-04 s/s/MY under the assumption of 110 and 90 MYA divergence times, respectively. The lower rate is generally consistent with a molecular clock previously calculated at the same gene across several insect orders (0.00022 s/s/MY at second codon positions of *COI*, which was calibrated using the Blattaria/Orthoptera divergence at ~380–354 MYA; Gaunt and Miles, 2002).

Evolutionary rates generated by the PL and NPRS methods are quite different (Appendix 3). This result is not surprising, because simulation studies have shown that NPRS performs best when the data set is large and the number of substitutions per branch is high (Sanderson, 1997, 2002). The short length of this *COI* data set (828 bp) and zero values for d_N along certain branches may compromise the performance of NPRS (Sanderson, 2002). For these reasons, we consider the PL results to be more reliable for this data set.

Based on the 90–110 MYA divergence and the estimated d_N values at *COI*, we then used the PL method of r8s to estimate the divergence of *Camponotus* and *Colobopsis* at 29.3–35.9 MYA, a minimum age for the *Blochmannia*–ant association. The same r8s analysis calculated the basal *Camponotus* node to be 16.2–19.9 MYA. This dated node was used to estimate absolute rates of *Blochmannia* sequence evolution.

Calculating absolute rates of Blochmannia *sequence evolution.*—We calculated d_S values for *Blochmannia gidA* and *groEL* along branches of the complete topology, using codeML. Branch lengths for the 16S rDNA phylogeny were calculated in PAUP* using the Tamura and Nei (1993) model, accounting for rate variation among sites under a gamma distribution ($\alpha = 0.5$). By constraining the basal *Camponotus* node to 16.2 or 19.9 MYA, *Blochmannia* branch lengths were then used to estimate absolute rates (r8s, PL method) (Table 7).

Estimating rates in Buchnera *and the enterobacteria.*—For comparison, we calculated rates for *Buchnera* and the enterics at

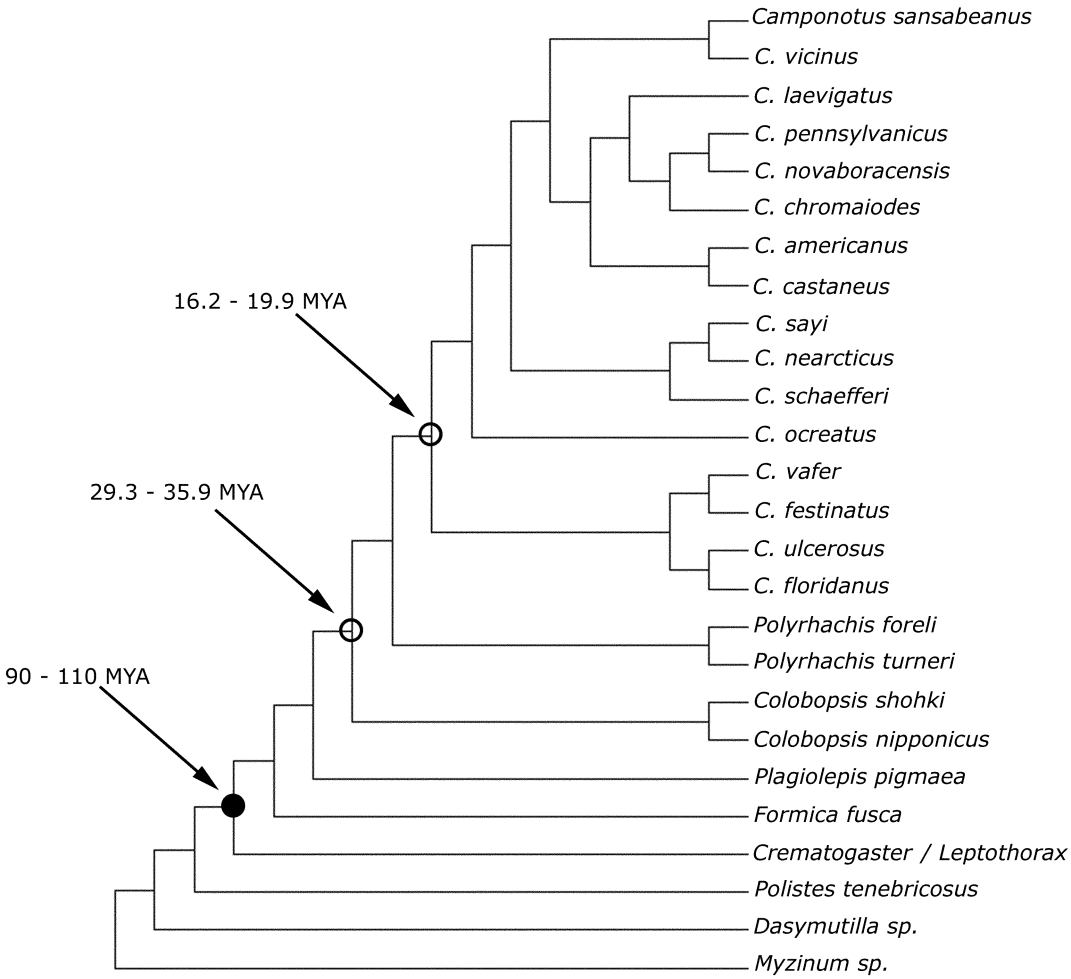

FIGURE A1. *COI* phylogeny used to estimate absolute rates of sequence evolution and divergence dates within Formicinae. Calculations were performed using r8s (Sanderson, 2003), constraining the divergence of myrmicines and formicines to 90–110 MYA (solid circle). Based on this constraint, the dates of nodes corresponding to the common ancestor of all *Blochmannia*-associated ants and the basal node of sampled *Camponotus* species were calculated (open circles). Non-formicine taxa were included in the analyses to root the phylogeny: *Leptothorax recedens* (AF096117), *Crematogaster borneensis* (AB030558), *Myzinum* sp. (AF142546/AF142535), *Dasymutilla* sp. (AF142541/AF142531), and *Polistes tenebricosus* (AF142550/AF142539).

the same gene regions we analyzed for *Blochmannia*. Rates for *Buchnera* were based on pairwise divergences between *Buchnera*-Ap and *Buchnera*-Sg with an estimated divergence time of 50–70 MYA (Tamas et al., 2002). Rates for the enterics were calculated for *E. coli* and *S. typhimurium*, which diverged ~100–150 MYA (Ochman and Wilson, 1987). See Appendix 4 for details of these calculations. Rates for *Buchnera* and *E. coli* were consistent with previous estimates based on many more genes (Clark et al., 1999). In the previous study, average rates of K_S for *Buchnera* ranged from 0.0051 to 0.0095 s/s/MY depending on *Buchnera* lineages considered and assumptions about their divergence time (Clark et al.,1999). Our estimates of d_S between *Buchnera*-Ap and *Buchnera*-Sg were similar but had a slightly higher maximum rate (0.0058–0.020344 s/s/MY at *gidA* and *groEL*). Rates at *Buchnera* 16S rDNA ranged from 0.00019 to 0.00045 s/s/MY (depending on the pair of *Buchnera* taxa and inferred divergence time) (Clark et al., 2000). Our slightly lower estimates (0.000148 to 0.000208 s/s/MY) may reflect different *Buchnera* taxa used or the restriction of our estimates to 1.36 kb of the of 16S rDNA gene (a region that is also available for *Blochmannia*). Previous estimates of 16S rDNA rates for the enterics were 0.00012–0.00018 s/s/MY (Clark et al., 1999); ours were about half that (0.000066–0.000100 s/s/MY) at the particular 16S rDNA region sampled.

APPENDIX 2

Specific PCR conditions used to amplify various *Blochmannia* and *Camponotus* gene regions.

Primers	$MgCl_2$ (mM)	Annealing temp (°C)	Elongation time
C1-J-1754F, C2-N-3661R	3.5	48	1 min, 45 sec
C1F365F, C2-N-3661R	2.5	50	1 min, 45 sec
groES367F, groEL1581R	2.5	58	1 min, 45 sec
groES367F, groEL1645R	2.0	56	1 min, 45 sec
SL, Bloch16S-1299R	3.5	56	1 min, 15 sec
Bloch16S-462F, SR	3.5	53	1 min, 15 sec
rpsB85F, rpsB571R	3.5	51	1 min, 15 sec
gidA40F, gidA1450R	3.0	55	1 min, 15 sec
EF1aF2-559F, EF1aF2-2111]	3.5	51	1 min, 30 sec

APPENDIX 3

Calculations of rates at nonsynonymous sites of *COI* and the ages nodes, using r8s (Sanderson, 2003) with penalized likelihood (PL) and nonparametric rate smoothing (NPRS) methods. Estimations are substitutions per site per million years (s/s/MY). The divergence times of *Camponotus* and *Camponotus*/*Colobopsis* were inferred based on a divergence of 90–110 MYA between the formicines and myrmicines.

		Cross-valdation smoothing parameter		Age of basal *Camponotus* node		Age of *Camponotus*/*Colobopsis* node		Estimated rate at nonsynonymous sites (s/s/MY)	
Model	Algorithm	90 MYA	110 MYA	90 MYA	110 MYA	90 MYA	110 MYA	110 MYA	90 MYA
PL	Powell	100	150	16.15	19.87	29.24	35.87	0.0004253 ± 6.53E-05	0.0005228 ± 8.34E-05
PL	Truncated Newton	100	150	16.19	19.89	29.30	35.90	0.0004245 ± 6.48E-05	0.0005212 ± 8.30E-05
NPRS	Powell	—	—	31.53	38.54	47.50	58.00	0.0002596 ± 2.30E-04	0.0003051 ± 2.80E-04

APPENDIX 4

Calculation of evolutionary rates in *Buchnera* and the enterics on an absolute time scale. To allow direct comparisons, rates were estimated for the same gene regions analyzed in *Blochmannia* and using the same evolutionary models (except as pairwise calculations here, as opposed to calculations across branches for *Blochmannia*). Rates were calculated at synonymous sites for protein-coding genes and at all sites of a ~1.36-kb gene region of 16S rDNA (homologous to the 16S rDNA region analyzed for *Blochmannia*). Pairwise d_S values at protein-coding genes were calculated using codeML (runmode = −2), and pairwise divergences at 16S rDNA were estimated using the Tamura and Nei (1993) model with a gamma distribution ($\alpha = 0.5$), as implemented in the program MEA (Molecular Evolutionary Analysis program, courtesy of E. Moriyama, Univ. Nebraska). The range of rates reflects uncertainty in divergence times. Absolute rates were estimated by dividing sequence divergences by the minimum and maximum estimated divergence times. Gene sequences were obtained from the full genome sequences of *Buchnera*-Sg (AE013218), *Buchnera*-Ap (NC002528), *E. coli* K12 (NC000916), and *S. typhimurium* LT2 (AE006468). Table 7 provides a comparison of these rates with those calculated for *Blochmannia*.

			Rates (s/s/MY)	
	d_S	Divergence time (MYA)	Minimum	Maximum
Buchnera (Ap vs. Sg)				
gidA	2.0344 ± 0.3664	70–50	0.014531	0.020344
groEL	1.5806 ± 0.2512	70–50	0.011290	0.015806
E. coli vs. *S. typhimurium*				
gidA	0.7988 ± 0.1013	150–100	0.002663	0.003994
groEL	0.2383 ± 0.0279	150–100	0.000794	0.001191
16S rDNA	K_S (pairwise TN93 + Γ)			
Buchnera (Ap vs. Sg)	0.0208 ± 0.0042	70–50	0.000148	0.000208
E. coli vs. *S. typhimurium*	0.0199 ± 0.0040	150–100	0.000066	0.000100