Evolutionary rates are correlated between cockroach symbiont and mitochondrial genomes

Daej A. Arab[1], Thomas Bourguignon[1,2,3], Zongqing Wang[4], Simon Y. W. Ho[1], & Nathan Lo[1]

[1]School of Life and Environmental Sciences, University of Sydney, Sydney, Australia
[2]Okinawa Institute of Science and Technology Graduate University, Tancha, Onna-son, Okinawa, Japan
[3]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic
[4]College of Plant Protection, Southwest University, Chongqing, China

Authors for correspondence:
Daej A. Arab
e-mail: daej.arab@sydney.edu.au
Nathan Lo
e-mail: nathan.lo@sydney.edu.au

Keywords: host-symbiont interaction, *Blattabacterium cuenoti*, phylogeny, molecular evolution, substitution rate, cockroach.

Abstract

Bacterial endosymbionts evolve under strong host-driven selection. Factors influencing host evolution might affect symbionts in similar ways, potentially leading to correlations between the molecular evolutionary rates of hosts and symbionts. Although there is evidence of rate correlations between mitochondrial and nuclear genes, similar investigations of hosts and symbionts are lacking. Here we demonstrate a correlation in molecular rates between the genomes of an endosymbiont (*Blattabacterium cuenoti*) and the mitochondrial genomes of their hosts (cockroaches). We used partial genome data for multiple strains of *B. cuenoti* to compare phylogenetic relationships and evolutionary rates for 55 cockroach/symbiont pairs. The phylogenies inferred for *B. cuenoti* and the mitochondrial genomes of their hosts were largely congruent, as expected from their identical maternal and cytoplasmic mode of inheritance. We found a correlation between evolutionary rates of the two genomes, based on comparisons of root-to-tip distances and on comparisons of the branch lengths of phylogenetically independent species pairs. Our results underscore the profound effects that long-term symbiosis can have on the biology of each symbiotic partner.

1. Introduction

Rates of molecular evolution are governed by a multitude of factors and vary significantly among species [1, 2]. In the case of symbiotic organisms, such rates may be influenced by the biology of their symbiotic partner, in addition to their own. This is particularly the case for strictly vertically transmitted, obligate intracellular symbionts (hereafter 'symbionts'), which have a highly intimate relationship with their hosts [3]. For example, a small host effective population size will potentially lead to increased fixation of slightly deleterious mutations within both host and symbiont genomes, owing to the reduced efficacy of selection.

When the phylogenies of host and symbiont taxa are compared, simultaneous changes in evolutionary rate between host-symbiont pairs might be evident in their branch lengths. Some studies have found a correlation in evolutionary rates between nuclear and mitochondrial genes in sharks [4], herons [5], and turtles [6], between plastid and mitochondrial genes in angiosperms [7], and between nuclear, plastid, and mitochondrial genes in green algae [8, 9]. These results suggest that host biology affects substitution rates in nuclear and cytoplasmic genomes in similar ways. In insects, nuclear genes that interact directly with mitochondrial proteins have shown rate correlations with mitochondrial genes [10].

Potential correlations in evolutionary rates between hosts and bacterial symbionts remain untested. Evidence for correlated levels of synonymous substitutions was found in a study of one nuclear gene and two mitochondrial genes from *Camponotus* ants and three genes from their *Blochmannia* symbionts [11]. However, the study did not determine whether this correlation was driven by rates of evolution, time since divergence, or both. Numbers of substitutions tend to be low for closely related pairs of hosts and their corresponding symbionts, and high for more divergent pairs, leading to a correlation with time that does not necessarily reflect correlation in evolutionary rates.

Blattabacterium cuenoti (hereafter *Blattabacterium*) is an intracellular bacterial symbiont that has been in an obligatory intracellular and mutualistic relationship with cockroaches for over 200 million years [12, 13]. Found in highly specialized cells in the fat bodies of cockroaches, *Blattabacterium* is required for host fitness and fertility, and is transovarially transmitted from the mother to the progeny [14, 15]. Genome-wide analyses of the symbiont have confirmed its role in host nitrogen metabolism and the synthesis of essential amino acids [16, 17]. The genomes of 21 *Blattabacterium* strains sequenced to date are highly reduced compared with those of their free-living relatives, ranging in size from

590 to 645 kb [16–20]. They contain genes encoding enzymes for DNA replication and repair, with some exceptions (*holA*, *holB*, and *mutH*) [16–20]. The extent to which host nuclear proteins are involved in the cell biology of *Blattabacterium*, and particularly DNA replication, is not well understood.

We recently performed a study of cockroach evolution and biogeography using mitochondrial genomes [12]. During this study, we obtained partial genomic information for several *Blattabacterium* strains. These data provide the opportunity to test for correlation of molecular evolutionary rates between *Blattabacterium* and host-cockroach mitochondrial DNA. Here we infer phylogenetic trees for 55 *Blattabacterium* strains on the basis of 104 genes and compare branch lengths and rates of evolution for host-symbiont pairs across the phylogeny. We find evidence of markedly increased rates of evolution in some *Blattabacterium* lineages, which appear to be matched by increased rates of evolution in mitochondrial DNA of host lineages.

2. Materials and methods

A list of samples and collection data for each cockroach examined is provided in table S1 (see electronic supplementary material, ESM). For the majority of taxa examined in this study, we obtained *Blattabacterium* sequence data from genomic libraries originally used in a previous study of cockroach mitochondrial genomes carried out by our laboratories [12]. In some cases, new genomic data were obtained from fat bodies of individual cockroaches (see ESM for further details). We obtained 104 genes of 55 *Blattabacterium* strains from these data and aligned them with orthologues from seven outgroup taxa from Flavobacteriales (details provided in ESM).

Genomic data were assembled and annotated, and then aligned and tested for saturation. After the exclusion of 3rd codon sites in each data set, total lengths for the mitochondrial and *Blattabacterium* nucleotide alignments were 11,051 bp and 71,458, respectively. The former was partitioned into four subsets (1st codon sites, 2nd codon sites, rRNAs, and tRNAs), and the latter into two subsets according to codon positions. Translated amino acid alignments were also prepared for both host and symbiont. Trees were inferred for both nucleotide and amino acid alignments using maximum likelihood in RAxML v8.2 [21], using 1000 bootstrap replicates to estimate node support. We examined congruence between host and symbiont phylogenies using the distance-based ParaFit [22] in R 3.5.1 [23].

Root-to-tip distances from the RAxML analyses for each host and symbiont pair were subjected to Pearson correlation analysis. Branch-length differences between hosts and symbionts were compared for 27 phylogenetically independent species pairs across the topology (see figure S1 in ESM). These were calculated using a fixed topology (derived from the *Blattabacterium* analysis described above) for each of the following three data sets: 1) 1st+2nd codon sites of protein-coding genes; 2) translated amino acid sequences; 3) 1st+2nd codon positions of protein-coding genes plus the inclusion of rRNAs and tRNAs in the case of the mitochondrial data set. Further details on phylogenetic methods are provided in the ESM.

3. Results

In all analyses, there was strong support for the monophyly of each cockroach family with the exception of Ectobiidae (figures 1, S2, S3). The topologies inferred from the host and symbiont data sets were congruent ($p = 0.001$). In only two cases was a disagreement found to be supported by >85% bootstrap support in both trees (the sister group of Lamproblattidae+Anaplectidae; the sister group to *Carbrunneria paraxami*+*Beybienkoa kurandensis*).

We found a correlation between root-to-tip distances for protein-coding genes from hosts and their symbionts ($R = 0.75$, figure 2*a*). Similar results were found when rRNAs and tRNAs were included in the host data set ($R = 0.73$, figure S4*a*). The highest rates of evolution in the host and symbiont data sets (on the basis of branch lengths; see figure 1) were in members of an ectobiid clade containing *Allacta* sp., *Amazonina* sp., *Balta* sp., *Chorisoserrata* sp., and *Euphyllodromia* sp., and a separate clade containing the Anaplectidae. After excluding these taxa, evolutionary rates remained correlated, although to a lesser degree ($R = 0.35$, figure 2*b*). The sharing of branches between taxa in the estimation of root-to-tip distances renders the data in these plots phylogenetically non-independent and precludes statistical analysis.

A comparison of branch lengths among phylogenetically independent pairs of host and symbiont taxa based on protein-coding genes (figure S1) revealed a significant correlation between their rates of evolution ($R = 0.40$, $p = 0.039$; figure 2*c*). Equivalent analyses of branch lengths inferred from amino acid data also revealed a significant rate correlation between host and symbiont ($R = 0.43$, $p = 0.023$; figure S5*a*). However, there was

no significant rate correlation between host and symbiont following inclusion of rRNAs and tRNAs in the host mitochondrial data set ($R = 0.27$, $p = 0.17$; figure 2*d*). Analyses involving standardization of branch-length differences yielded significant rate correlations for the protein-coding gene and amino acid data sets ($R = 0.43–0.48$, $p = 0.011–0.023$; see figures S5, S6), and mixed results in the case of the inclusion of rRNAs and tRNAs in the host mitochondrial data set ($R = 0.34–0.40$, $p = 0.041–0.085$; see figure S4).

4. Discussion

We have detected a correlation in molecular evolutionary rates between *Blattabacterium* and host mitochondrial genomes, using two different methods of analysis. To our knowledge, this is the first evidence of such a correlation in a host-symbiont relationship. Previous studies found a correlation in evolutionary rates between mitochondrial and nuclear genes in various animal groups [4–6, 10].

Similar forces acting on the underlying mutation rates of both host and symbiont genomes could translate into a relationship between their substitution rates. This could potentially occur if symbiont DNA replication depends on the host's DNA replication and repair machinery [24]. In leafhoppers, a number of nuclear-encoded proteins targeted to mitochondria, including those involved in DNA replication and repair, are thought to be retargeted to nutritional symbionts, potentially leading to similarities in their mutation rates [25]. Interactions between host mitochondrial and symbiont proteins could also lead to correlations in evolutionary rates, as has been found between insect mitochondrial genes and nuclear genes that encode proteins targeted to mitochondria [10]. The level of integration of host-encoded proteins in the metabolism of *Blattabacterium*, and interactions between *Blattabacterium* and mitochondria, are not well understood. Further exploration of these interactions will shed light on the causes of the correlation in rates that we have found here.

Short host generation times could potentially lead to elevated evolutionary rates in host and symbiont [26], assuming that increased rates of symbiont replication are associated with host reproduction, as is found in *Blochmannia* symbionts of ants [27]. Variations in metabolic rate and effective population size between host taxa, as well as increased transmission bottlenecks of both mitochondria and symbionts, could also explain the rate correlations that we have observed. Unfortunately, with the exception of a few pest and other species, generation time, metabolic rates, and host and symbiont effective population sizes are poorly

understood in cockroaches. This precludes an examination of their influence on evolutionary rates in host and symbiont.

The addition of mitochondrial rRNA and tRNA data weakened the correlations found in the branch-length comparisons of species pairs. The reasons for this are unclear but they might be associated with the conserved nature of tRNAs and the stem regions of rRNAs, or highly variable loop regions in the latter.

Blattabacterium is a vertically transmitted, obligate intracellular mutualistic symbiont, whose phylogeny is expected to mirror that of its hosts. This is especially the case for phylogenies inferred from mitochondrial DNA, since mitochondria are linked with *Blattabacterium* through vertical transfer to offspring through the egg cytoplasm. As has been found in previous studies [28–30], we observed a high level of agreement between the topologies inferred from cockroach mitochondrial and *Blattabacterium* genome data sets. In some cases, however, disagreements were observed between well-supported relationships. The variability in rates that we observed between some lineages, and/or the highly increased rate of mitochondrial DNA compared with *Blattabacterium* DNA, could be responsible for these disagreements. Owing to long periods of co-evolution and co-cladogenesis between cockroaches and *Blattabacterium* [12, 28], potential movement of strains between hosts (for example, via parasitoids) is not expected to result in the establishment of new symbioses, especially between hosts that diverged millions of years ago.

In conclusion, our results highlight the profound effects that long-term symbiosis can have on the biology of each symbiotic partner. The rate of evolution is a fundamental characteristic of any species; our study indicates that it can become closely linked between organisms as a result of symbiosis. Further studies are required to determine whether the correlation that we have found here also applies to the nuclear genome of the host. Future investigations of generation time, metabolic rate, and effective population sizes in cockroaches and *Blattabacterium* will allow testing of their potential influence on evolutionary rates.

Funding. D.A.A. was supported by an International Postgraduate Research Stipend from the Australian Government. S.Y.W.H. and N.L. were supported by Future Fellowships from the Australian Research Council. T.B. was supported by the Japan Society for the Promotion of Science KAKENHI 18K14767, and by an EVA4.0 grant (No.

CZ.02.1.01/0.0/0.0/16_019/0000803) from the OP RDE. Z.W. acknowledges funding from the National Natural Sciences Foundation of China (31872271).

Acknowledgements. We thank Qian Tang, Frantisek Juna, James Walker, and David Rentz for providing specimens, and Charles Foster for assistance with data curation.

References

1. Bromham, L., 2009. Why do species vary in their rate of molecular evolution? *Biol. Lett.*, rsbl. 2009.0136.
2. Ho SYW, Lo N. 2013 The insect molecular clock. *Aust. J. Entomol.* **52**, 101–105.
3. Douglas AE. 2010 *The Symbiotic Habit*. Princeton, NJ: Princeton University Press.
4. Martin AP. 1999 Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). *Mol. Biol. Evol.* **16**, 996–1002.
5. Sheldon FH, Jones CE, McCracken KG. 2000 Relative patterns and rates of evolution in heron nuclear and mitochondrial DNA. *Mol. Biol. Evol.* **17**, 437–450.
6. Lourenço JM, Glémin S, Chiari Y, Galtier N. 2013 The determinants of the molecular substitution process in turtles. *J. Evol. Biol.* **26**, 38–50.
7. Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012 Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol. Evol.* **4**, 294–306.
8. Smith DR, Lee RW. 2010 Low nucleotide diversity for the expanded organelle and nuclear genomes of *Volvox carteri* supports the mutational-hazard hypothesis. *Mol. Biol. Evol.* **27**, 2244–2256.
9. Hua J, Smith DR, Borza T, Lee RW. 2012 Similar relative mutation rates in the three genetic compartments of *Mesostigma* and *Chlamydomonas*. *Protist* **163**, 105–115.
10. Yan Z, Ye G, Werren JH. 2019 Evolutionary rate correlation between mitochondrial-encoded and mitochondria-associated nuclear-encoded proteins in insects. *Mol. Biol. Evol.* **36**, 1022–1036.
11. Degnan PH, Lazarus AB, Brock CD, Wernegreen JJ. 2004 Host–symbiont stability and fast evolutionary rates in an ant–bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus blochmannia*. *Syst. Biol.* **53**, 95–110.
12. Bourguignon T *et al.* 2018 Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Mol. Biol. Evol.* **35**, 970–983.
13. Evangelista DA *et al.* 2019 An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proc. R. Soc. B* **286**, 20182076.
14. Cornwell PB. 1968 *The Cockroach*. London: Hutchinson.
15. Sacchi L, Corona S, Grigolo A, Laudani U, Selmi MG, Bigliardi E. 1996 The fate of the endocytobionts of *Blattella germanica* (Blattaria: Blattellidae) and *Periplaneta americana* (Blattaria: Blattidae) during embryo development. *Ital. J. Zool.* **63**, 1–11.
16. López-Sánchez MJ, Neef A, Peretó J, Patiño-Navarrete R, Pignatelli M, Latorre A, Moya A. 2009 Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLOS Genet.* **5**, e1000721.

17. Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA. 2012 Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Appl. Environ. Microbiol.* **78**, 204–210.
18. Kinjo Y *et al.* 2018 Parallel and gradual genome erosion in the *Blattabacterium* endosymbionts of *Mastotermes darwiniensis* and *Cryptocercus* wood roaches. *Genome Biol. Evol.* **10**, 1622–1630.
19. Vicente CSL, Mondal SI, Akter A, Ozawa S, Kikuchi T, Hasegawa K. 2018 Genome analysis of new *Blattabacterium* spp., obligatory endosymbionts of *Periplaneta fuliginosa* and *P. japonica*. *PLOS ONE* **13**, e0200512.
20. Kambhampati S, Alleman A, Park Y. 2013 Complete genome sequence of the endosymbiont *Blattabacterium* from the cockroach *Nauphoeta cinerea* (Blattodea: Blaberidae). *Genomics* **102**, 479–483.
21. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
22. Legendre P, Desdevises Y, Bazin E. 2002 A statistical test for host–parasite coevolution. *Syst. Biol.* **51**, 217–234.
23. R Development Core Team. 2018 R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
24. Moran NA, Bennett GM. 2014 The tiniest tiny genomes. *Annu. Rev. Microbiol.* **68**, 195–215.
25. Mao M, Yang X, Bennett GM. 2018 Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proc. Natl Acad. Sci. U.S.A.* **115**, E11691–E11700.
26. Bromham L. 2009 Why do species vary in their rate of molecular evolution? *Biol. Lett.* **5**, 401–404.
27. Wolschin F, Hölldobler B, Gross R, Zientz E. 2004 Replication of the endosymbiotic bacterium *Blochmannia floridanus* is correlated with the developmental and reproductive stages of its ant host. *Appl. Environ. Microbiol.* **70**, 4096–4102.
28. Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T. 2003 Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Mol. Biol. Evol.* **20**, 907–913.
29. Clark JW, Hossain S, Burnside CA, Kambhampati S. 2001 Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective. *Proc. R. Soc. B* **268**, 393–398.
30. Garrick RC, Sabree ZL, Jahnes BC, Oliver JC. 2017 Strong spatial-genetic congruence between a wood-feeding cockroach and its bacterial endosymbiont, across a topographically complex landscape. *J. Biogeogr.* **44**, 1500–1511.
31. Arab D, Bourguignon T, Wang Z, Ho SYW, Lo N. 2019 Data from: Evolutionary rates are correlated between cockroach symbiont and mitochondrial genomes. Dryad Digital Repository. (https://doi.org/10.5061/dryad.v6wwpzgqw)

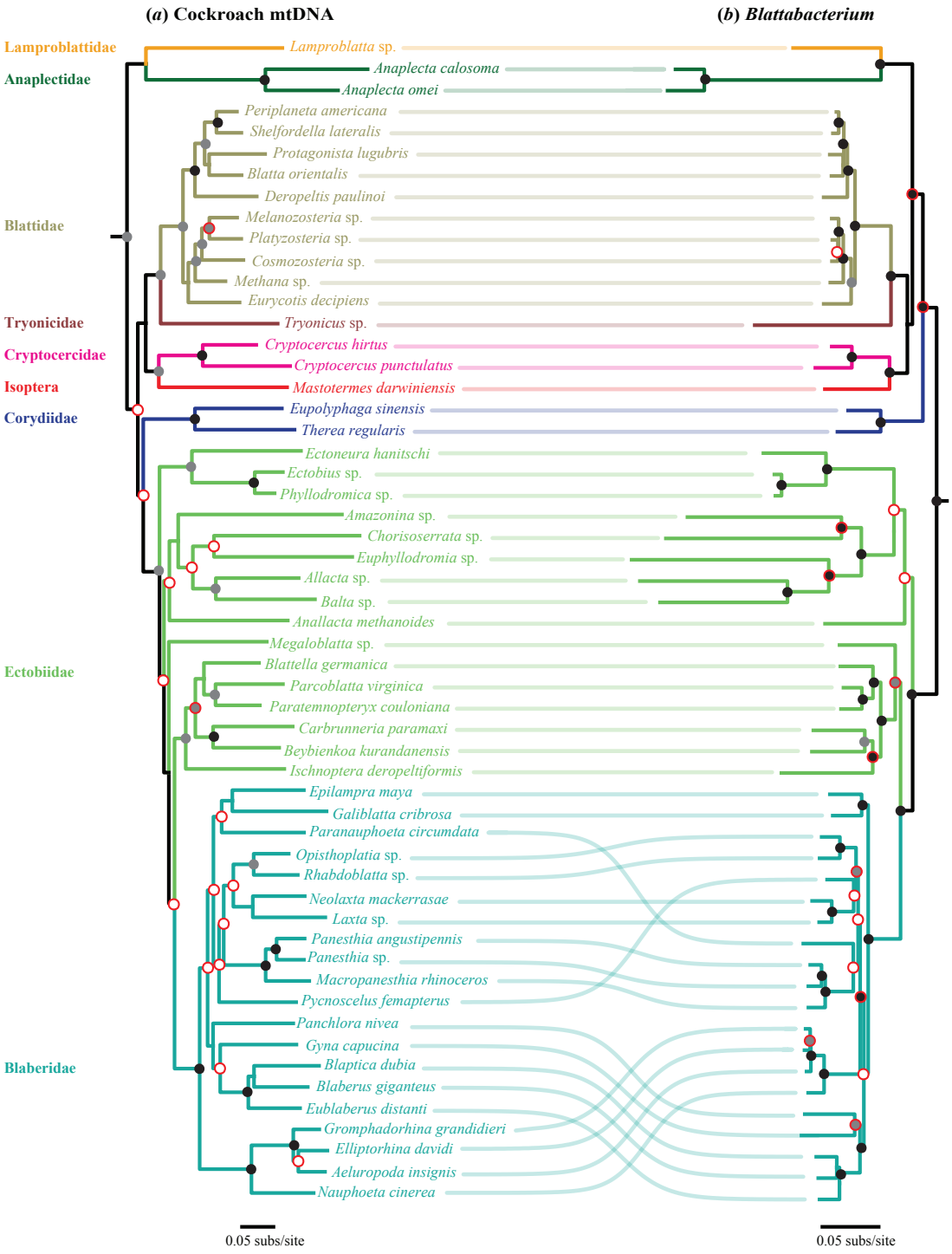


Figure 1. Congruence between (*a*) phylogenetic tree of host cockroaches inferred using maximum likelihood from whole mitochondrial genomes, and (*b*) phylogenetic tree of *Blattabacterium* inferred using maximum likelihood from 104 protein-coding genes (3rd codon sites excluded from both data sets). Shaded circles at nodes indicate bootstrap values (black = 100%, grey = 85–99%). Nodes without black or grey circles have bootstrap values <85%. Red outlines on circles indicate disagreement between the phylogenies. Branches are coloured according to their membership of different cockroach families.

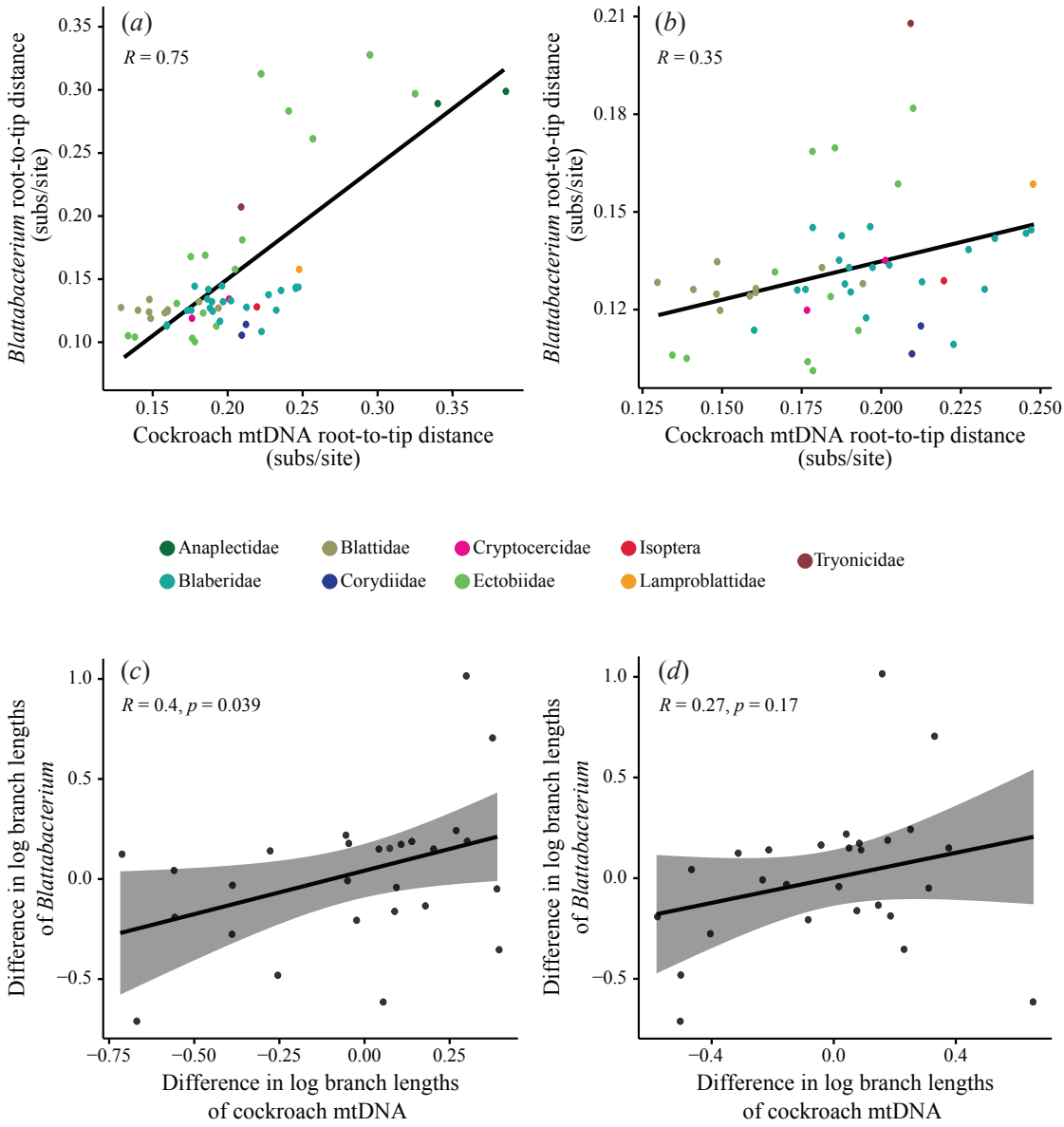


Figure 2. Comparison of evolutionary rates of *Blattabacterium* symbionts and their host cockroaches. (*a*) Correlation of root-to-tip distances in phylogenies of *Blattabacterium* and cockroaches, inferred using maximum-likelihood analysis of protein-coding genes from each data set, with 3rd codon sites excluded. (*b*) Correlation of root-to-tip differences following the removal of five rapidly evolving ectobiid taxa (*Amazonina* sp., *Chorisoserrata* sp., *Allacta* sp., *Balta* sp., and *Euphyllodromia* sp.) and two anaplectids. Colours represent data from representatives of different cockroach families, as shown in the colour key. Grey circles represent internal branches. (*c*) Correlation of log-transformed branch-length differences between phylogenetically independent pairs of host and symbiont taxa, based on protein-coding genes only, and (*d*) with the addition of rRNAs and tRNAs to the host mitochondrial data set.

Electronic supplementary material

Evolutionary rates are correlated between cockroach symbiont and mitochondrial genomes

Daej A. Arab[1], Thomas Bourguignon[1,2,3], Zongqing Wang[4], Simon Y. W. Ho[1], & Nathan Lo[1]

[1]School of Life and Environmental Sciences, University of Sydney, Sydney, Australia
[2]Okinawa Institute of Science and Technology Graduate University, Tancha, Onna-son, Okinawa, Japan
[3]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic
[4]College of Plant Protection, Southwest University, Chongqing, China

Sampling and Blattabacterium genomic data

A list of samples and collection data for each cockroach and outgroup examined is provided in table S1. All specimens examined in this study are stored at the Okinawa Institute of Science and Technology, Japan. For the majority of taxa examined in this study, we obtained *Blattabacterium* sequence data from genomic libraries originally used in a previous study of cockroach mitochondrial genomes carried out by our laboratories [1]. In some cases, new genomic data were obtained from fat bodies of individual cockroaches, as follows. DNA was extracted using a DNeasy Blood and Tissue kit (Qiagen), according to the manufacturer's protocol. Individual DNA samples were tagged with unique barcode combinations, mixed in equimolar concentration, and 150 bp paired-end-reads-sequenced with an Illumina HiSeq4000, following the methods described previously [1].

For each cockroach species, raw sequence data from the previous study [1] or the current study were assembled using CLC, and subject to blastn analysis using the published *Blattabacterium* genomes from *Blattella germanica* [2], *Periplaneta americana* [3], and *Cryptocercus punctulatus* [4] as subject sequences. Contigs identified as being derived from *Blattabacterium* during this step were then annotated using Prokka v1.12 [5]. Across the 55 strains examined in this study, a total of 104 orthologous genes were used for analysis. These

were found in ≥95% of all taxa. All taxa had over 90% of 104 genes, except for *Aeluropoda insignis* which only had 83 (79.8%) genes. Missing genes were presumed to be a result of uneven sequencing coverage of samples and the relatively low sequencing coverage used, rather than the actual absence of these genes from their genomes; further work is required to confirm their presence or absence. The genome sequences of outgroups were obtained from GenBank and included three strains of *Sulcia muelleri* (accession numbers CP002163, AP013293, and CP010828), one *Flavobacterium gilvum* (CP017479), one *Lutibacter* sp. (CP017478), one *Tenacibaculum dicentrarchi* (CP013671), and one *Polaribacter* sp. (LT629752).

The 104 orthologous *Blattabacterium* genes were each aligned at the amino acid level individually using TranslatorX [6] and concatenated into a 107,187 bp alignment. The mitochondrial genome data set included all protein-coding genes from each taxon plus 12S rRNA, 16S rRNA, and the 22 tRNA genes, and were obtained during a previous study [1]. All mtDNA protein coding genes were free of stop codons and indels, and could be translated into complete amino acid sequences, indicating that they were not nuclear insertions. Mitochondrial protein-coding genes were aligned using TranslatorX, while MAFFT [7] was used to align 12S rRNA, 16S rRNA, and the 22 tRNAs. All mitochondrial sequences were then concatenated into a 14,802 bp alignment. MEGA7 [8] was used to calculate the nucleotide composition of cockroach mtDNA and *Blattabacterium* data sets. The percentage of A+T of host and symbiont nucleotide datasets is shown in figure S7. We tested for substitution saturation using Xia's method implemented in DAMBE 6 [9, 10]. Third codon sites in the mitochondrial data set were saturated (NumOTU = 32, I_{SS} = 0.804, $I_{SS.C}$Asym = 0.809) and were excluded from our analyses. Although the *Blattabacterium* sequences were not significantly saturated at 3rd codon sites (NumOTU = 32, I_{SS} = 0.649, $I_{SS.C}$Asym = 0.819), we excluded these sites from our analyses because the test statistic was close to the critical value. After the exclusion of 3rd codon sites, the total lengths of the final data sets were 11,051 bp and 71,458 for the mitochondrial and *Blattabacterium* alignments, respectively.

Phylogenetic analysis

Maximum-likelihood phylogenetic analyses were carried out in RAxML v8.2 [11], using 1000 bootstrap replicates to estimate node support. The cockroach mtDNA data set was

partitioned into four subsets: 1st codon sites, 2nd codon sites, rRNAs, and tRNAs. The *Blattabacterium* data set was partitioned into two subsets: 1st codon sites and 2nd codon sites. Using jModelTest [12], we selected the GTR+G substitution model for each subset based on Bayesian information criterion scores. Using ProtTest v3.4 [13], the translated amino acid data set for *Blattabacterium* was assigned the CAT+CpREV model and the translated amino acid data set for cockroach mtDNA was assigned the CAT+MtART model based on Bayesian information criterion scores.

We used ParaFit in R 3.5.1 [14] to quantify congruence between host and symbiont topologies. We first created matrices of patristic distances calculated from maximum-likelihood host and symbiont phylogenies and a host-symbiont association matrix. We then performed a global test with 999 permutations, using the ParafitGlobal value and a *p*-value threshold of 0.05 to determine significance.

Root-to-tip distances and comparison of phylogenetically independent pairs of host and symbiont branch lengths

Root-to-tip distances from the RAxML analyses for each host and symbiont pair were calculated and subjected to Pearson correlation analysis using the R packages ape [15], phylobase [16], and adephylo [17]. The use of root-to-tip distances removes the confounding effects of time, because all lineages leading to the tips of the tree have experienced the same amount of time since evolving from their common ancestor. However, the sharing of internal branches by groups of taxa renders these data non-independent. Therefore, we compared branch-length differences between hosts and symbionts for 27 phylogenetically independent species pairs across the topology (see figure S1). These were calculated using a fixed topology (derived from the *Blattabacterium* analysis described above) for each of the following three data sets: 1) 1st+2nd codon sites of protein-coding genes; 2) translated amino acid sequences; and 3) 1st+2nd codon positions of protein-coding genes plus the inclusion of rRNAs+tRNAs in the case of the mitochondrial data set. Branch lengths were log transformed, and differences between pairs of hosts and pairs of symbionts were calculated and compared via Pearson correlation analysis.

To test for potential biases in the data that violate the assumptions of linear regressions, we compared the absolute mean value of log-transformed branch lengths with the log-transformed branch-length differences [18]. We found no significant correlation

between these values ($R = 0.07$, $p = 0.63$ for data from host cockroaches; $R = 0.06$, $p = 0.66$ for data from *Blattabacterium*), indicating that the data were suitable for use in our analyses. We also performed analyses in which branch-length differences were standardized following previous recommendations [19], to account for the potential confounding effects of the different amounts of time that sister pairs have had to diverge. Three standardizations were carried out, each based on dividing log-transformed branch-length differences by the square root of an estimate of time since divergence for the pair. In the first, time since divergence for host pairs was estimated as the average branch length of the host pair, divided by an assumed rate of 0.001 subs/site/million years, while for corresponding symbionts it was estimated as the average branch length of the symbiont pair, divided by the same assumed rate. In the second and third standardizations, times since divergence for both symbionts and hosts were based either on average branch lengths of host pairs only or symbiont pairs only.

Supplementary table S1. A list of samples and collection data for each cockroach examined.

Species	Family	Sample ID	Collecting locality	Collector	Date
Aeluropoda insignis	Blaberidae	B002	Breeding colony of Kyle Kandilian	N/A	N/A
Allacta australiensis	Ectobiidae	AUS Allacta	James Cook University, Rainforest site, Queensland, Australia	David Rentz	22-Jun-2015
Amazonina sp.	Ectobiidae	Z256E	Ecuador, Bosque Protector del Alto Nangaritza	Frantisek Juna	Apr-2016
Anallacta methanoides	Ectobiidae	B057	Breeding colony of Kyle Kandilian	N/A	N/A
Anaplecta calosoma	Anaplectidae	Cockroach contig 1688	Kuranda, Queensland, Australia	David Rentz	17-Nov-2015
Anaplecta omei	Anaplectidae	Anaplecta_omei	Mt Emei, Sichuan, China	Zongqing Wang	01-Jul-2013
Balta sp.	Ectobiidae	Balta_sp.	Cairns, Queensland, Australia	David Rentz	18-Dec-2015
Beybienkoa kurandanensis	Ectobiidae	Beybienkoa_kurandanensis	Cairns, Queensland, Australia	David Rentz	18-Dec-2015
Blaberus giganteus	Blaberidae	BGIGA	GenBank	N/A	N/A
Blaptica dubia	Blaberidae	B056	Breeding colony of Kyle Kandilian	N/A	N/A
Blatta orientalis	Blattidae	BOR	GenBank	N/A	N/A
Blattella germanica	Ectobiidae	BGE	GenBank	N/A	N/A
Carbrunneria paramaxi	Ectobiidae	Carbru	Cairns, Queensland, Australia	David Rentz	05-Oct-2015
Chorisoserrata sp.	Ectobiidae	CHORI	Yunnan, China	Zongqing Wang	01-Jul-2013
Cosmozosteria sp.	Blattidae	B117	Cape Upstart, Queensland, Australia	James Walker	13-Oct-2015
Cryptocercus hirtus	Cryptocercidae	HIR	Mt Taibai, Shaanxi, China	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPU	GenBank	N/A	N/A
Deropeltis paulinoi	Blattidae	B069	Breeding colony of Kyle Kandilian	N/A	N/A
Ectobius sp.	Ectobiidae	Z254C	Slovenia	Frantisek Juna	Apr-2016
Ectoneura hanitschi	Ectobiidae	Ectoneura_hanitschi	James Cook University, Rainforest site, Queensland, Australia	David Rentz	18-Dec-2015
Epilampra maya	Blaberidae	B095	Arcadia, Florida, USA	Kyle Kandilian	07-Jul-2009
Eublaberus distanti	Blaberidae	B025	Breeding colony of Kyle Kandilian	N/A	
Euphyllodromia sp.	Ectobiidae	Z257	Podocarpus National Park, Ecuador	Frantisek Juna	Apr-2016
Eupolyphaga sinensis	Corydiidae	B081	Breeding colony of Kyle Kandilian	N/A	N/A
Eurycotis decipiens	Blattidae	B071	Breeding colony of Kyle Kandilian	N/A	N/A
Galiblatta cribrosa	Blaberidae	Z98	Nouragues, French Guiana	N/A	14-Jun-2015
Gromphadorhina grandidieri	Blaberidae	B030	Breeding colony of Kyle Kandilian	N/A	N/A
Gyna capucina	Blaberidae	Z139GY	Ebogo, Cameroon	Frantisek Juna	08-Sep-2015
Ischnoptera deropeltiformis	Ectobiidae	B083	Torreya State Park, Bristol, Florida, USA	Kyle Kandilian	07-Jul-2009
Lamproblatta sp.	Lamproblattidae	LA male	Petit Saut, French Guiana	Frantisek Juna	08-Jul-2009
Laxta sp.	Blaberidae	AUS2	Olney State Forest, New South, Wales, Australia	Nathan Lo and Thomas Bourguignon	25-Aug-2015
Macropanesthia rhinoceros	Blaberidae	B092	Breeding colony of Kyle Kandilian	N/A	N/A
Mastotermes darwiniensis	Isoptera	MADAR	GenBank	N/A	N/A
Megaloblatta sp.	Ectobiidae	ECMD1	Podocarpus National Park, Ecuador	Frantisek Juna	Apr-2016
Melanozosteria sp.	Blattidae	Melanozosteria_sp.	Cairns, Queensland, Australia	David Rentz	18-Dec-2015
Methana sp.	Blattidae	AUS1	North Manly, New South Wales, Australia	Nathan Lo	01-Aug-2015

Nauphoeta cinerea	Blaberidae	BNCIN	GenBank	N/A	N/A
Neolaxta mackerrasae	Blaberidae	B107	Paluma Range, Queensland, Australia	David Rentz	15-Oct-2015
Opisthoplatia orientalis	Blaberidae	Z15100	Breeding colony of J. Hromádka	N/A	N/A
Panchlora nivea	Blaberidae	B044	Breeding colony of Kyle Kandilian	N/A	N/A
Panesthia angustipennis	Blaberidae	Z138	Breeding colony in Czech Republic, orig. Vietnam	N/A	N/A
Panesthia sp.	Blaberidae	Panesthia_sp	Bubeng, Yunnan, China	N/A	N/A
Paranauphoeta circumdata	Blaberidae	PARA	N/A	N/A	N/A
Paratemnopteryx couloniana	Ectobiidae	B061	Breeding colony of Kyle Kandilian	N/A	N/A
Parcoblatta virginica	Ectobiidae	B102	Breeding colony of Kyle Kandilian	N/A	N/A
Periplaneta americana	Blattidae	BPLAN	GenBank	N/A	N/A
Phyllodromica sp.	Ectobiidae	Phil	Czech Republic	Thomas Bourguignon	01-Aug-2015
Platyzosteria sp.	Blattidae	AUS3	Olney State Forest, New South, Wales, Australia	Nathan Lo and Thomas Bourguignon	25-Aug-2015
Protagonista lugubris	Blattidae	Cockroach contig 4907	Mt Diaoluo, Hainan, China	Zongqing Wang	25-May-2015
Pycnoscelus femapterus	Blaberidae	B048	Breeding colony of Kyle Kandilian	N/A	N/A
Rhabdoblatta sp.	Blaberidae	RHA	Kuranda, Queensland, Australia	David Rentz	16-Sep-2015
Shelfordella lateralis	Blattidae	B080	Breeding colony of Kyle Kandilian	N/A	N/A
Therea regularis	Corydiidae	B091	Palm plantation between Puducherry and Auroville, India	Kyle Kandilian	N/A
Tryonicus parvus	Tryonicidae	Tryonicus_parvus	Olney State Forest, New South, Wales, Australia	Nathan Lo and Thomas Bourguignon	10-Mar-2016
Sulcia muelleri	Flavobacteriaceae	CARI	GenBank	N/A	N/A
Sulcia muelleri	Flavobacteriaceae	PSPU	GenBank	N/A	N/A
Sulcia muelleri	Flavobacteriaceae	CARI	GenBank	N/A	N/A
Flavobacterium gilvum	Flavobacteriaceae	EM1308	GenBank	N/A	N/A
Lutibacter sp.	Flavobacteriaceae	LPB0138	GenBank	N/A	N/A
Tenacibaculum dicentrarchi	Flavobacteriaceae	AY7486TD	GenBank	N/A	N/A
Polaribacter sp.	Flavobacteriaceae	LT629752	GenBank	N/A	N/A

Supplementary table S2. A list of GenBank accession numbers and names of all 104 *Blattabacterium* genes used for this study.

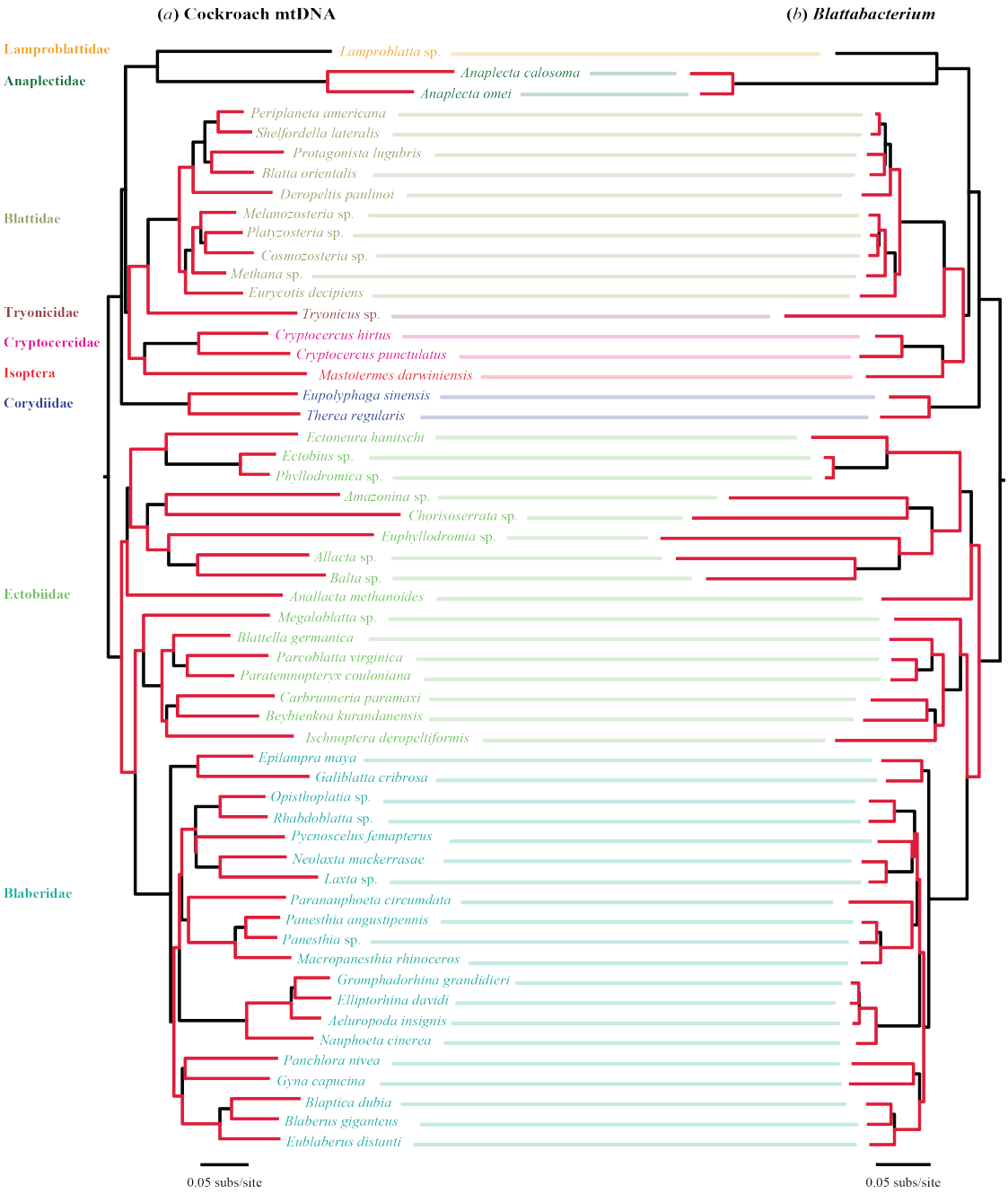
Accession No.	Gene name
MN038417 - MN038462	Dihydrolipoyllysine-residue succinyltransferase component of 2oxoglutarate dehydrogenase complex.
MN038463 - MN038510	Cysteine desulfuration protein
MN038511 - MN038558	hypothetical protein
MN038559 - MN038606	Putative 1,2-phenylacetyl-CoA epoxidase, subunit D
MN038607 - MN038654	UDP-N-acetylglucosamine--N-acetylmuramyl-pentapeptide pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
MN038655 - MN038703	50S ribosomal protein L11
MN038704 - MN038752	50S ribosomal protein L11
MN038753 - MN038800	Fumarate reductase flavoprotein subunit
MN038801 - MN038848	Glutamate dehydrogenase
MN038849 - MN038895	Asparagine tRNA ligase
MN038896 - MN038942	Polyribonucleotide nucleotidyltransferase
MN038943 - MN038989	RNA polymerase sigma factor SigA
MN038990 - MN039036	3-oxoacyl-acyl-carrier-protein synthase 2
MN039037 - MN039083	Acyl carrier protein
MN039084 - MN039130	ATP synthase epsilon chain
MN039131 - MN039177	30S ribosomal protein S2
MN039178 - MN039223	50S ribosomal protein L13
MN039224 - MN039271	10 kDa chaperonin
MN039272 - MN039318	60 kDa chaperonin
MN039319 - MN039365	Glutamine tRNA ligase
MN039366 - MN039412	DNA-directed RNA polymerase subunit beta'
MN039413 - MN039459	Glyceraldehyde-3-phosphate dehydrogenase A

MN039460 - MN039506	50S ribosomal protein L21
MN039507 - MN039553	50S ribosomal protein L22
MN039554 - MN039600	30S ribosomal protein S19
MN039601 - MN039647	50S ribosomal protein L2
MN039648 - MN039695	50S ribosomal protein L3
MN039696 - MN039742	50S ribosomal protein L1
MN039743 - MN039790	Cysteine desulfurase SufS
MN039791 - MN039837	FeS cluster assembly protein SufB
MN039838 - MN039884	Protein translocase subunit SecY
MN039885 - MN039931	50S ribosomal protein L15
MN039932 - MN039979	30S ribosomal protein S10
MN039980 - MN040027	Elongation factor G
MN040028 - MN040075	30S ribosomal protein S7
MN040076 - MN040123	30S ribosomal protein S12
MN040124 - MN040171	Methionine aminopeptidase 1
MN040172 - MN040219	30S ribosomal protein S5
MN040220 - MN040267	Alternate 30S ribosomal protein S14
MN040268 - MN040314	50S ribosomal protein L14
MN040315 - MN040361	30S ribosomal protein S17
MN040362 - MN040408	50S ribosomal protein L16
MN040409 - MN040456	30S ribosomal protein S3
MN040457 - MN040504	Two names: 1)] Acetylornithine deacetylase; 2)] Succinyl-diaminopimelate desuccinylase
MN040505 - MN040552	Aspartate semialdehyde dehydrogenase
MN040553 - MN040599	50S ribosomal protein L17
MN040600 - MN040647	DNA-directed RNA polymerase subunit alpha

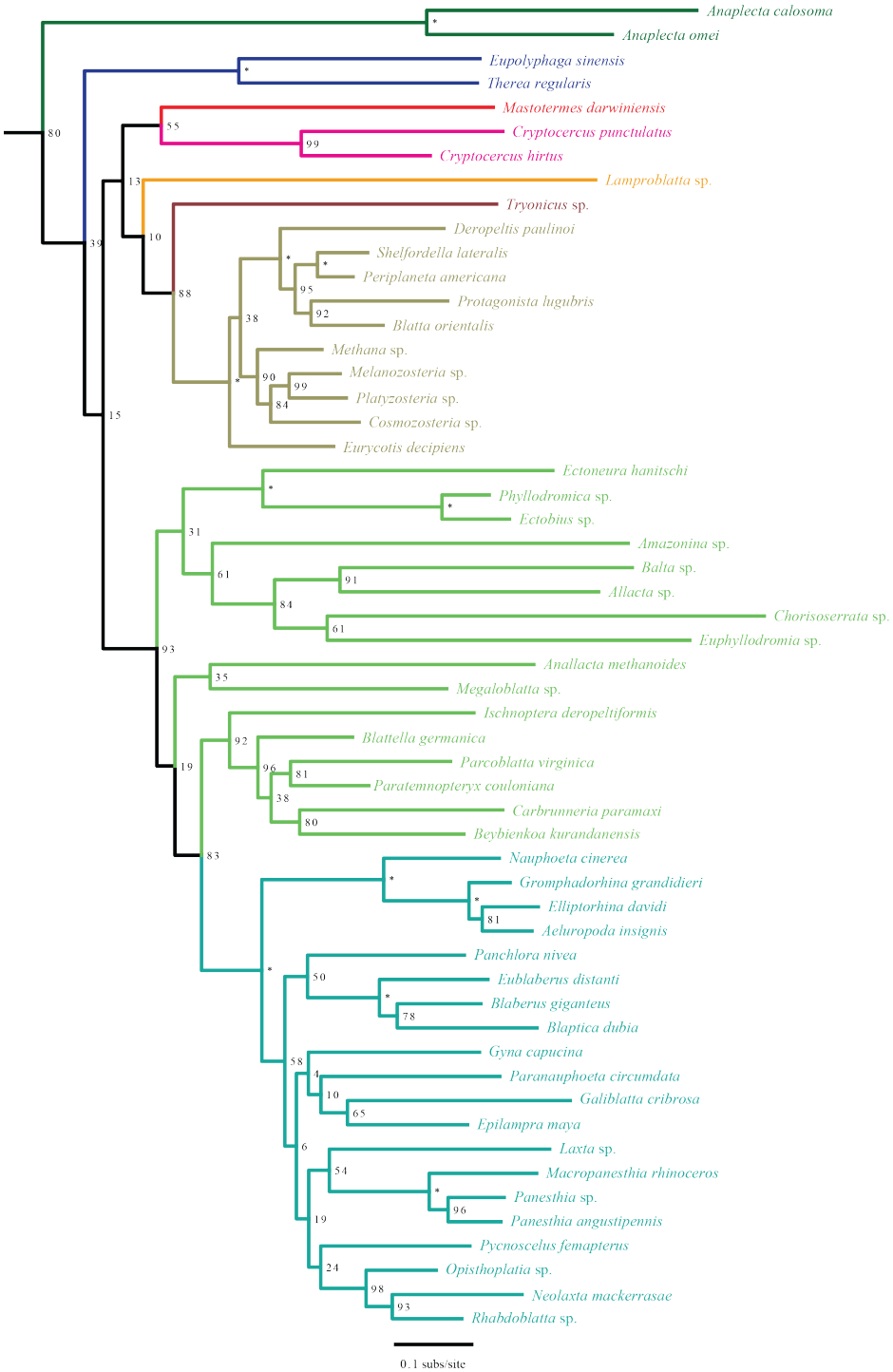
MN040648 - MN040695 30S ribosomal protein S4
MN040696 - MN040743 30S ribosomal protein S11
MN040744 - MN040791 30S ribosomal protein S13
MN040792 - MN040839 Translation initiation factor IF-1
MN040840 - MN040887 N-acetylornithine carbamoyltransferase
MN040888 - MN040935 Carbamoyl-phosphate synthate large chain
MN040936 - MN040983 Carbamoyl-phosphate synthate small chain
MN040984 - MN041031 Acetylornithine aminotransferase
MN041032 - MN041079 N-acetyl-gamma-glutanyl-phosphate reductase
MN041080 - MN041127 Argininosuccinate synthase
MN041128 - MN041175 30S ribosomal protein S1
MN041176 - MN041223 Enoyl-acyl-carrier-protein reductase NADH FabI
MN041224 - MN041271 S-adenosylmethionine synthase
MN041272 - MN041319 Phospho-2-dehydro-3-deoxyheptonate aldolase
MN041320 - MN041366 50S ribosomal protein L9
MN041367 - MN041412 30S ribosomal protein S6
MN041413 - MN041460 tRNA modification GTPase MnmE
MN041461 - MN041508 Lon protease 2
MN041509 - MN041556 Histidine--tRNA ligase
MN041557 - MN041603 Phenylalanine--tRNA ligase alpha subunit
MN041604 - MN041650 DNA gyrase subunit B
MN041651 - MN041698 30S ribosomal protein S16
MN041699 - MN041746 Aspartate aminotransferase
MN041747 - MN041794 Lysine--tRNA ligase
MN041795 - MN041840 Octanoyltransferase

MN041841 - MN041887 Methionine--tRNA ligase
MN041888 - MN041935 Histidinol dehydrogenase
MN041936 - MN041981 hypothetical protein
MN041982 - MN042029 Phosphate acetyltransferase
MN042030 - MN042076 1-deoxy-D-xylulose-5-phosphate synthase
MN042077 - MN042124 Transketolase 2
MN042125 - MN042170 SsrA-binding protein
MN042171 - MN042216 Lipoyl synthase
MN042217 - MN042264 Multifunctional CCA protein
MN042265 - MN042311 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase
MN042312 - MN042358 putative branched-chain-amino-acid aminotransferase
MN042359 - MN042405 2-oxoisovalerate dehydrogenase subunit beta
MN042406 - MN042452 Ribosomal RNA small subunit methyltransferase A
MN042453 - MN042499 Putative aminopeptase YsdC
MN042500 - MN042546 Diaminopimelate epimerase
MN042547 - MN042594 ATP synthase subunit c
MN042595 - MN042641 ATP synthase subunit beta
MN042642 - MN042688 Chaperone protein DnaJ
MN042689 - MN042735 hypothetical protein
MN042736 - MN042782 Ribose-phosphate pyrophosphokinase
MN042783 - MN042830 Imidazole glycerol phosphate synthase subunit HisF
MN042831 - MN042878 Imidazole glycerol phosphate synthase subunit HisH
MN042879 - MN042926 Ribosome recycling factor
MN042927 - MN042973 ATP synthase subunit A
MN042974 - MN043021 Bifunctional aspartokinase

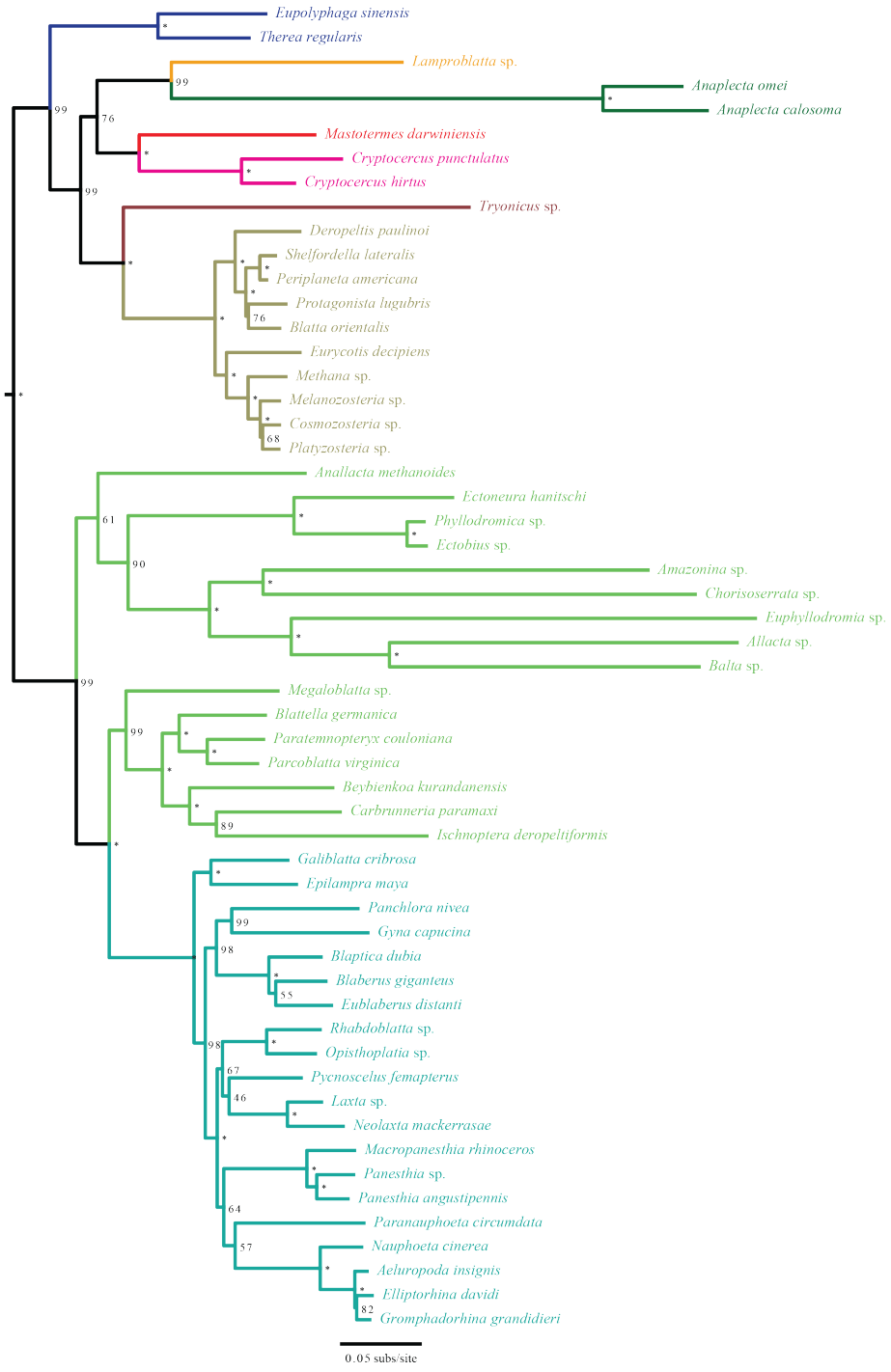
MN043022 - MN043069	hypothetical protein
MN043070 - MN043116	3-dehydroquinate synthase
MN043117 - MN043164	DNA gyrase subunit A
MN043165 - MN043211	Fructose-bisphosphate aldolase
MN043212 - MN043259	hypothetical protein
MN075834- MN075880	Elongation Factor
MN075881- MN075928	tRNA-2-methylthio-N6-dimethylallyladenosine synthase
CP003535- CP003536	*Blaberus giganteus*
CP003605- CP003606	*Blatta orientalis*
CP001487	*Blattella germanica*
CP003015- CP003016	*Cryptocercus punctulatus*
CP003000, CP003095	*Mastotermes darwiniensis*
CP005488- CP005489	*Nauphoeta cinerea*
CP001429- CP001430	*Periplaneta americana*
CP002163	*Sulcia muelleri*
AP013293	*Sulcia muelleri*
CP002165	*Sulcia muelleri*
CP017479	*Flavobacterium gilvum*
CP017478	*Lutibacter* sp.
CP013671	*Tenacibaculum dicentrachi*
KT25b	*Polaribacter* sp.



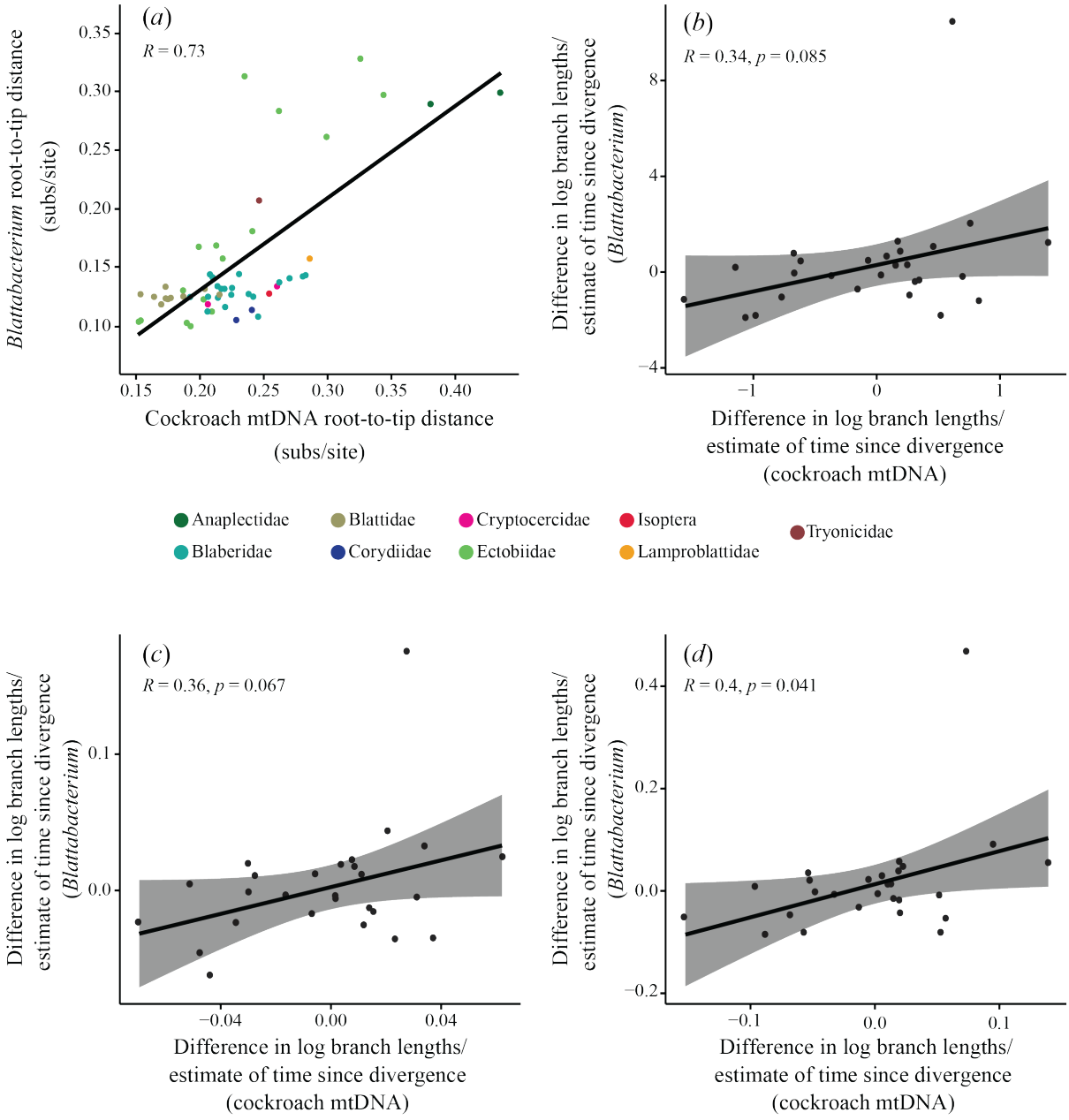
Supplementary figure S1. Phylogenetic trees inferred from (*a*) cockroach mtDNA data (protein-coding genes plus rRNAs and tRNAs) and (*b*) their *Blattabacterium* symbiont data, inferred using maximum likelihood in RAxML. A fixed topology (obtained from the *Blattabacterium* tree shown in figure 1) was used in each analysis. Twenty-seven phylogenetically independent pairs of lineages used to test for correlations of evolutionary rates are shown in red. Species names are coloured according to the family to which they belong, as shown on the left of the figure.



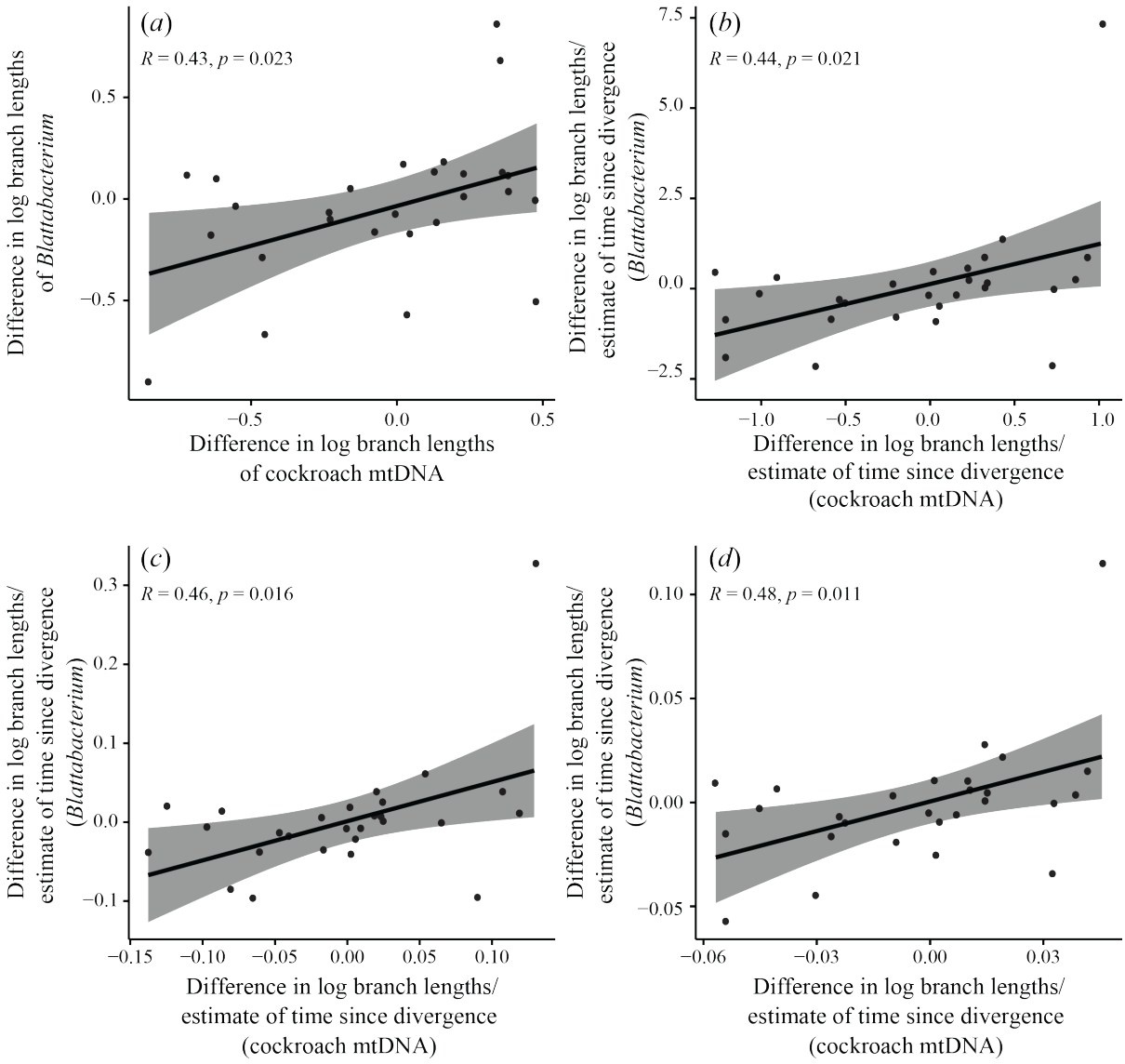
Supplementary figure S2. Host cockroach phylogenetic tree inferred using maximum likelihood, based on amino acid sequences translated from mitochondrial protein-coding genes. Support values of 100% are indicated by asterisks.



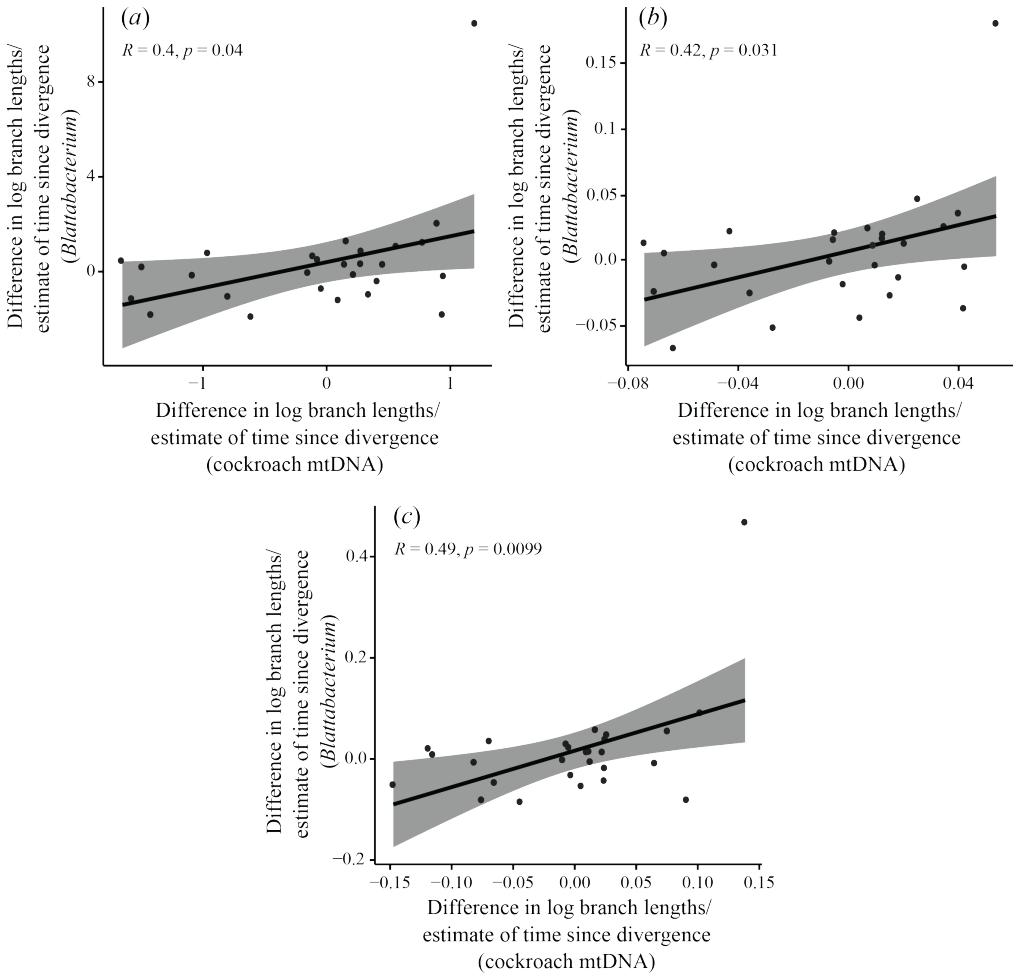
Supplementary Figure S3. *Blattabacterium* phylogenetic tree inferred using maximum likelihood, based on amino acid sequences translated from protein-coding genes. Support values of 100% are indicated by asterisks.



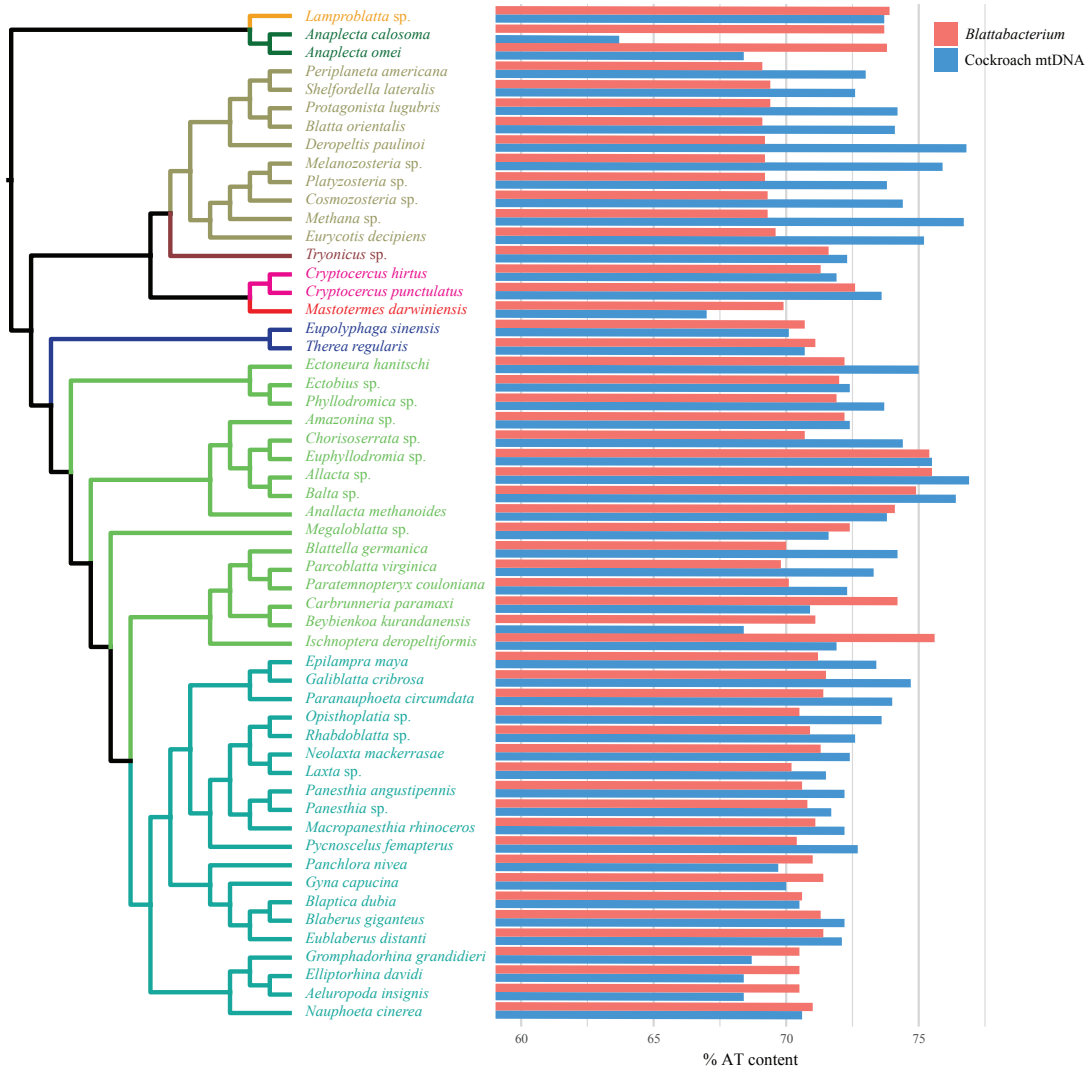
Supplementary figure S4. Comparison of evolutionary rates of *Blattabacterium* symbionts and their host cockroaches, based on protein-coding genes from host and symbiont, plus rRNAs+tRNAs from host mitochondria. (*a*) Correlation of root-to-tip distances in phylogenies of *Blattabacterium* and cockroaches. (*b–d*) Standardized tests for correlation of molecular evolutionary rates between 27 independent pairs of *Blattabacterium* and host cockroach mitochondria. Three standardizations were carried out, each based on dividing log-transformed branch-length differences by the square root of an estimate of time since divergence for the pair. In the first (*b*), time since divergence for host pairs was estimated as the average branch length of the host pair, divided by an assumed rate of 0.001.



Supplementary figure S5. Tests for correlation of molecular evolutionary rates between 27 independent pairs of *Blattabacterium* and host cockroach mitochondria, based on amino acid data translated from protein-coding genes. (*a*) Test based on comparison of log-transformed branch-length differences. Three standardizations were also carried out, each based on dividing log-transformed branch-length differences by the square root of an estimate of time since divergence for the pair. In the first standardization (*b*), time since divergence for host pairs was estimated as the average branch length of the host pair, divided by an assumed rate of 0.001 subs/site/million years, while for corresponding symbionts it was estimated as the average branch length of the symbiont pair, divided by the same assumed rate. In the second (*c*) and third (*d*) standardizations, times since divergence for both symbionts and hosts were based either on average branch lengths of host pairs only or symbiont pairs only.



Supplementary figure S6. Standardized tests for correlation of molecular evolutionary rates between 27 independent pairs of *Blattabacterium* and host cockroach mitochondria, based on protein-coding genes. Three standardizations were carried out, each based on dividing log-transformed branch-length differences by the square root of an estimate of time since divergence for the pair. In the first standardization (*a*), time since divergence for host pairs was estimated as the average branch length of the host pair, divided by an assumed rate of 0.001 subs/site/million years, while for corresponding symbionts it was estimated as the average branch length of the symbiont pair, divided by the same assumed rate. In the second (*b*) and third (*c*) standardizations, times since divergence for both symbionts and hosts were based either on average branch lengths of host pairs only or symbiont pairs only. subs/site/million years, while for corresponding symbionts it was estimated as the average branch length of the symbiont pair, divided by the same assumed rate. In the second (*c*) and third (*d*) standardizations, times since divergence for both symbionts and hosts were based either on average branch lengths of host pairs only or symbiont pairs only.



Supplementary figure S7. AT content (%) of *Blattabacterium* and mtDNA sequences for each taxon, including all codon positions.

References

1. Bourguignon T *et al.* 2018 Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Mol. Biol. Evol.* **35**, 970–983.
2. López-Sánchez MJ, Neef A, Peretó J, Patiño-Navarrete R, Pignatelli M, Latorre A, Moya A. 2009 Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLOS Genet.* **5**, e1000721.
3. Sabree ZL, Kambhampati S, Moran NA. 2009 Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc. Natl Acad. Sci. U.S.A.* **106**, 19521–19526.
4. Neef A, Latorre A, Pereto J, Silva FJ, Pignatelli M, Moya A. 2011 Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities. *Genome Biol. Evol.* **3**, 1437–1448.
5. Seemann T. 2014 Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069.
6. Abascal F, Zardoya R, Telford MJ. 2010 TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–W13.
7. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
8. Kumar S, Stecher G, Tamura K. 2016 MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874.
9. Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003 An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* **26**, 1–7.
10. Xia X, Lemey P. 2009 Assessing substitution saturation with DAMBE. In *The phylogenetic handbook: a practical approach to DNA and protein phylogeny* (eds. Lemey P, Salemi M, Vandamme A-M). Cambridge: Cambridge University Press.
11. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
12. Darriba D, Taboada GL, Doallo R, Posada D. 2012 jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772.
13. Darriba D, Taboada GL, Doallo R, Posada D. 2011 ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165.
14. R Development Core Team. 2018 R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
15. Paradis E, Schliep K. 2018 ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528.
16. Bolker B *et al.* 2015 phylobase: Base package for phylogenetic structures and comparative data. *R package version 0.8. 0*.
17. Jombart T, Balloux F, Dray S. 2010 Adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* **26**, 1907–1909.
18. Freckleton RP. 2000 Phylogenetic tests of ecological and evolutionary hypotheses: checking for phylogenetic independence. *Funct. Ecol.* **14**, 129–134.
19. Welch JJ, Waxman D. 2008 Calculating independent contrasts for the comparative study of substitution rates. *J. Theor. Biol.* **251**, 667–678.