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

Author 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 strong correlation between evolutionary rates of the two genomes, based on comparisons of root-to-tip distances and on estimates of individual branch rates. 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 could be influenced not only by factors associated with their own biology, but also those of their symbiotic partner. This is particularly the case for strictly vertically transmitted, obligate intracellular symbionts (hereafter 'symbionts'), which have a highly intimate relationship with their host [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], suggesting that host biology affects substitution rates in nuclear and cytoplasmic genomes in similar ways. In insects, nuclear genes that interact directly with mitochondrial proteins and mitochondrial data have shown rate correlation [7].

There has not yet been any study of rate correlations between hosts and bacterial symbionts. 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 [8]. 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 [9,10]. These bacteria are transovarially transmitted from the mother to the progeny. The genomes of 21 *Blattabacterium* strains sequenced to date are highly reduced compared with those of their free-living ancestors, ranging in size from 590 to 645 kb [11,12]. They contain genes encoding enzymes for DNA replication and repair, with some exceptions (*holA*, *holB*, and *mutH*) [12–14]. 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 [9]. During this process, 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 are 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 (electronic supplementary material (ESM)). We used assembled data obtained from a previous study, in which we used cockroach mitochondrial genomes to build phylogenetic trees [9]. We searched each assembly against previously published *Blattabacterium* genomes using blastn [15] to identify *Blattabacterium* contigs. Each contig was annotated using Prokka v1.12 [16]. We determined orthology among 104 genes of 55 *Blattabacterium* strains and seven Flavobacteriales outgroups with OMA v1.0.6 [17]. Further details are available in the ESM.

The 104 orthologous *Blattabacterium* genes were aligned individually using MAFFT v7.300b [18] and concatenated. The mitochondrial genome dataset included all protein coding genes from each taxon plus 12S rRNA, 16S rRNA, and the 22 tRNA genes. Third codon sites were removed from each dataset on the basis of saturation tests using Xia's method implemented in DAMBE 6 [19, 20] (see ESM). Trees were inferred using maximum likelihood in RAxML v8.2 [21]. We examined congruence between host and symbiont topologies using the distance-based ParaFit [22]. Further details of phylogenetic analyses and congruence testing are provided in the ESM.

The evolutionary timescale of hosts and symbionts, as well as their evolutionary rates, were inferred using BEAST v 1.8.4 [23], using a fixed topology from the RAxML analysis (figure 1). We calibrated the molecular clock using minimum age constraints based on four fossils (table S2, ESM). A soft maximum bound of 311 Ma was set for the root node,

representing the oldest known cockroach-like fossil [24]. To allow evolutionary rates to be estimated in a single framework, a random subset of 12 *Blattabacterium* protein-coding genes (from the larger set of 104) plus the full mitochondrial data set were concatenated and partitioned into host and symbiont subsets. These analyses were carried out a total of three times, with a novel subset of 12 randomly selected *Blattabacterium* genes for each replicate. The inferred branch rates were then compared using Pearson correlation analysis using ggpubr [25] in R [26].

Root-to-tip distances from the RAxML analysis for each host and symbiont pair were calculated using the R packages ape [27], phylobase [28], and adephylo [29]. The use of root-to-tip distances removes the confounding effects of time, because all lineages leading to terminal taxa have experienced the same amount of time since evolving from their common ancestor. We used Seq-Gen v1.3.4 [30] to simulate the evolution of sequences from both host and symbiont along the *Blattabacterium* tree topology inferred using RAxML, with evolutionary parameters obtained from our separate RAxML analyses of the original data. Tree lengths for host and symbiont were rescaled according to their relative rates, but the relative branch rates were maintained between the two trees (see ESM).

3. Results

In all analyses, there was strong support for the monophyly of each cockroach family with the exception of Ectobiidae (figure 1). The topologies inferred from the host and symbiont data sets were significantly congruent (p=0.001). Although there were some apparent disagreements between the two trees (for example, the position of Corydiidae), support at these nodes was generally low for both trees.

A molecular-clock analysis of the *Blattabacterium* data set indicated that the basal divergence occurred 314 Ma (95% credibility interval 219–420 Ma; figure S1, electronic supplementary material), giving rise to a one clade containing strains infecting Corydiidae, termites, Cryptocercidae, Blattidae, Anaplectidae, Tryonicidae, and Lamproblattidae, and a second clade containing strains infecting Ectobiidae and Blaberidae. We repeated the analysis using the mitochondrial data set and found that divergence times were markedly younger (184 Ma, 95% CI 160–212 Ma; figure S2, electronic supplementary material). In an analysis of the combined data set, divergence times were approximately midway between those from the

separate analyses (216 Ma, 95% CI 159–278 Ma; figure S3, electronic supplementary material).

The inferred substitution rates along each pair of equivalent host and symbiont branches were found to be highly correlated (R=0.88, $p<2.2\times10^{-16}$; figure 2*a*). Almost identical results were found in replicate analyses involving different sets of *Blattabacterium* genes (data not shown). The highest rates of evolution in the host and symbiont data sets (on the basis of branch lengths; figure 1) were in members of an ectobiid clade containing *Allacta*, *Amazonina*, *Balta*, *Chorisoserrata*, and *Euphyllodromia*, and a separate clade containing the two anaplectids *Anaplecta omei* and *Anaplecta calosoma*. After excluding these taxa, R was reduced to 0.47 but remained highly significant ($p=1.5\times10^{-6}$). As expected, analyses of simulated host and symbiont data yielded highly correlated estimates of branch rates (figure 2*b*).

Regression analysis of root-to-tip distances for host-symbiont pairs also indicated that these two variables were correlated (R=0.7; figure 2*c*). However, the sharing of branches between taxa in the estimation of root-to-tip distances renders the data in this plot phylogenetically non-independent and precludes statistical analysis.

4. Discussion

Our study provides evidence for a correlation in molecular evolutionary rates between *Blattabacterium* and host mitochondrial genomes, based on two different approaches (branch-rate comparisons and analysis of root-to-tip distances). To our knowledge, this is the first demonstration of such a correlation in a host-symbiont relationship. Previous studies found a correlation in evolutionary rates between mitochondrial and nuclear genes [5–7], and this relationship appears especially pronounced for nuclear genes encoding proteins that are associated with mitochondria [7]. The pattern that we found was consistent across multiple subsets of *Blattabacterium* genes, indicating that it represents a genome-wide phenomenon for this symbiont.

Similar forces acting on the underlying mutation rates of both host and symbiont genomes could translate into a relationship between their rates of substitution. This could potentially occur if symbiont DNA replication depends on the host's DNA replication and repair machinery [31]. Because the genome of *Blattabacterium* is known to possess an almost complete suite of replication and repair enzymes [31], the scope for host enzymes to

significantly influence symbiont mutation rate appears to be limited. A better understanding of the level of integration of host-encoded proteins in the metabolism of *Blattabacterium* is required to explore this issue further.

Short host generation times could potentially lead to elevated evolutionary rates in host [32] and symbiont, assuming that increased rates of symbiont replication are associated with host reproduction, as is found in *Blochmannia* symbionts of ants [33]. Variations in metabolic rate and effective population size between host taxa could also explain the rate correlations that we observed here. Unfortunately, with the exception of a few pest and other species, generation time, metabolic rates, and effective population sizes are poorly understood in cockroaches. This precludes an examination of their influence on evolutionary rates in host and symbiont.

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 [34–36], we observed a high level of agreement between the topologies inferred from cockroach mitochondrial genomes and from the 104-gene *Blattabacterium* data set. Owing to long periods of co-evolution and co-cladogenesis between cockroaches and *Blattabacterium* [9,34], 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, and our study shows 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.

Author contributions. D.A.A., T.B. generated sequence data; Z.Q. collected and provided specimens; D.A.A., T.B, N.L., and S.Y.W.H. performed data analysis and interpreted results; N.L., D.A.A., T.B. and S.Y.W.H. designed the study; N.L. and D.A.A. wrote the manuscript,

with contributions from T.B. and S.Y.W.H. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Data accessibility. Sequence data have been uploaded to GenBank (accession numbers XXXXXX–XXXXXX (TBA) and alignments have been uploaded to Github (details TBA).

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.Q. acknowledges funding from the National Natural Sciences Foundation of China (31872271).

Competing interests. There are no competing interests.

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, Penny D. 2003 The modern molecular clock. *Nat. Rev. Genet.* **4**, 216–224.
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 J, Glémin S, Chiari Y, Galtier N. 2013 The determinants of the molecular substitution process in turtles. *J. Evol. Biol.* **26**, 38–50.
7. Yan Z, Ye G-Y, Werren J. 2018 Evolutionary rate coevolution between mitochondria and mitochondria-associated nuclear-encoded proteins in insects. *bioRxiv*, 288456.
8. 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.
9. Bourguignon T, Tang Q, Ho SYW, Juna F, Wang Z, Arab DA, Cameron SL, Walker J, Rentz D, Evans TA. 2018 Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Mol. Biol. Evol.* **35**, 970–983.
10. Evangelista DA, Wipfler B, Béthoux O, Donath A, Fujita M, Kohli MK, Legendre F, Liu S, Machida R, Misof B. 2019 An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proc. R. Soc. B* **286**, 20182076.

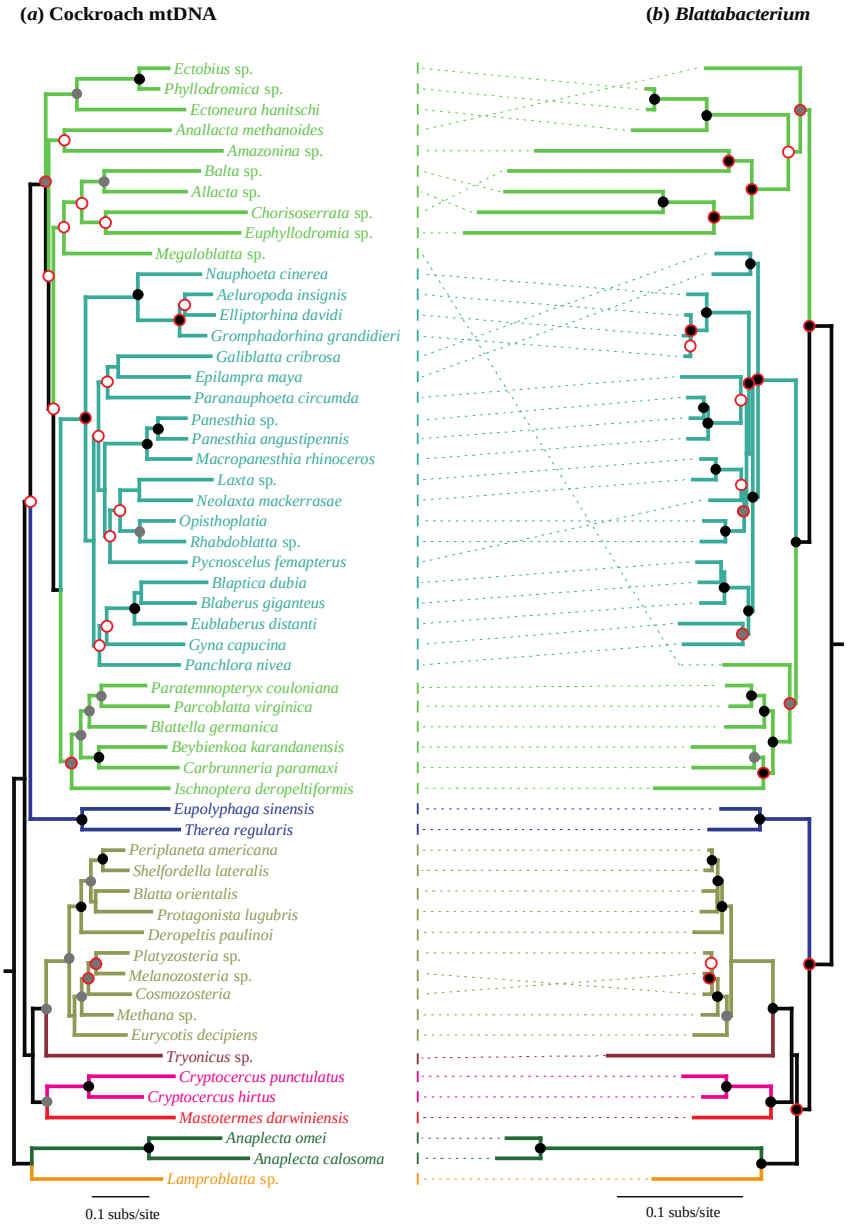
11. Kinjo Y, Bourguignon T, Tong KJ, Kuwahara H, Lim SJ, Yoon KB, Shigenobu S, Park YC, Nalepa CA, Hongoh Y. 2018 Parallel and gradual genome erosion in the *Blattabacterium* endosymbionts of *Mastotermes darwiniensis* and *Cryptocercus* wood roaches. *Genome Biol. Evol.* **10**, 1622–1630.
12. Vicente CS, 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.
13. 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.
14. 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.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
16. Seemann T. 2014 Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069.
17. Altenhoff AM, Škunca N, Glover N, Train C-M, Sueki A, Piližota I, Gori K, Tomiczek B, Müller S, Redestig H. 2014 The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. *Nucleic Acids Res.* **43**, D240–D249.
18. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
19. 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.
20. Xia X, Lemey P. 2009 Assessing substitution saturation with DAMBE. In *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny* (eds P Lemey, M Salemi, A-M Vandamme), pp. 615–630. Cambridge: Cambridge University Press.
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. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012 Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973.
24. Carpenter F. 1966 The Lower Permian insects of Kansas. Part II. The orders Protorthoptera and Orthoptera. *Psyche* **73**, 46–88.
25. Kassambara A. 2017 ggpubr: “ggplot2” based publication ready plots. R package version 0.1.
26. R Core Team. 2014 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
27. Paradis E, Schliep K. 2018 ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528.
28. Bolker B, Butler M, Cowan P, de Vienne D, Eddelbuettel D, Holder M, Jombart T, Kembel S, Michonneau F, Orme B. 2015 phylobase: Base package for phylogenetic structures and comparative data. R package version 0.8.
29. Jombart T, Balloux F, Dray S. 2010 Adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* **26**, 1907–1909.

30. Rambaut A, Grass NC. 1997 Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics* **13**, 235–238.
31. Moran NA, Bennett GM. 2014 The tiniest tiny genomes. *Annu. Rev. Microbiol.* **68**, 195–215.
32. Bromham L. 2009 Why do species vary in their rate of molecular evolution? *Biol. Lett.* **5**, 401–403.
33. 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.
34. 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.
35. 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.
36. 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.

Figure captions

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 for both datasets). Circles at nodes indicate bootstrap values (black = 100%, grey = 85–99%). Nodes without circles have bootstrap values <85%. Red outlines on circles indicate disagreement between the phylogenies, whereas red outlines on white circles indicate disagreement between the phylogenies and bootstrap values <85%. Colours represent taxa belonging to different cockroach families: light green = Ectobiidae, teal = Blaberidae, blue = Corydiidae, olive green = Blattidae, maroon = Tryonicidae, pink = Cryptocercidae, red = termites, dark green = Anaplectidae, and orange = Lamproblattidae.

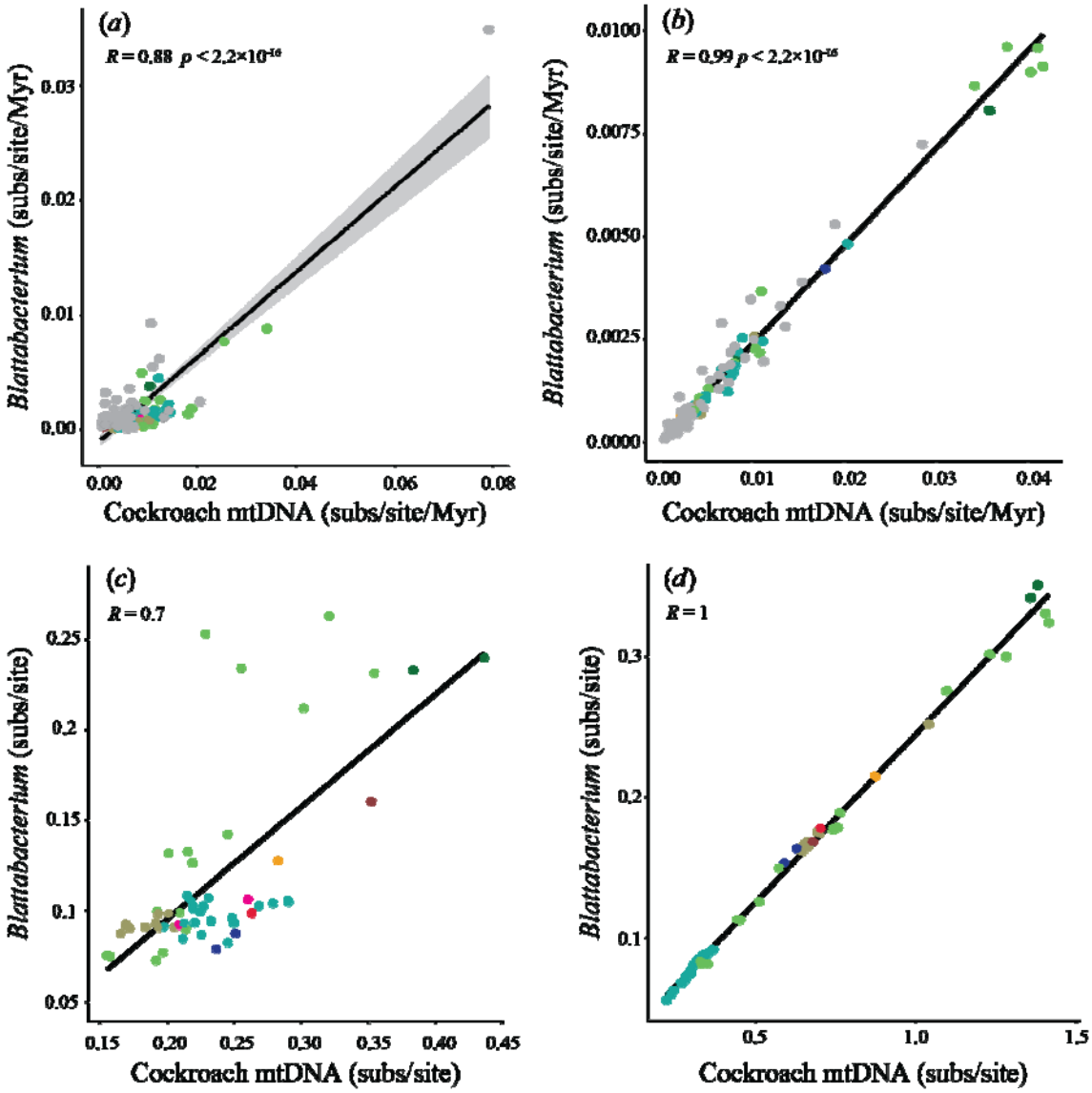
Figure 2. Comparison of evolutionary rates of *Blattabacterium* symbionts and their host cockroaches. (*a*,*b*) Correlation between branch rates in the phylogenies of *Blattabacterium* and cockroaches, obtained from a Bayesian time-calibrated tree inferred from (a) 12 *Blattabacterium* protein-coding genes and whole mitochondrial genomes from cockroaches, with 3rd codon sites excluded; (*b*) synthetic sequence data (see ESM). (*c*,*d*) Correlation of root-to-tip distances in phylogenies of *Blattabacterium* and cockroaches, inferred using maximum-likelihood analysis of (*c)*104 *Blattabacterium* protein-coding genes and whole mitochondrial genomes from cockroaches, with 3rd codon sites excluded; (*d*) synthetic sequence data (see ESM). Colours represent data from representatives of different cockroach families, as described in figure 1. Grey circles represent internal branches.



326

327 Figure 1

328



329

330

331

332

333 Figure 2