Increased mutation rate is linked to genome reduction in prokaryotes

Thomas Bourguignon[1,2,3,9,10,*], Yukihiro Kinjo[1,9,*], Paula Villa-Martin[1], Nicholas V. Coleman[3], Qian Tang[4], Daej A. Arab[3], Zongqing Wang[5], Gaku Tokuda[6], Yuichi Hongoh[7], Moriya Ohkuma[8], Simon Y. W. Ho[3], Simone Pigolotti[1], Nathan Lo[3,*]

[1]Okinawa Institute of Science & Technology Graduate University, 1919–1 Tancha, Onna-son, 904–0495 Okinawa, Japan
[2]Faculty of Tropical AgriSciences, Czech University of Life Sciences, Kamýcká 129, CZ-165 00 Prague, Czech Republic
[3]School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia
[4]Department of Biological Sciences, National University of Singapore, 117543 Singapore, Singapore
[5]College of Plant Protection, Southwest University, No. 2 Tiansheng Road, Beibei District, Chongqing, 400716, China
[6]Tropical Biosphere Research Center, Center of Molecular Biosciences, University of the Ryukyus, Nishihara, 903-0213 Okinawa, Japan
[7]School of Life Science and Technology, Tokyo Institute of Technology, Tokyo 152-8550, Japan
[8]RIKEN Bioresource Research Centre, Tsukuba 305-0074, Japan
[9]These authors contributed equally
[10]Lead contact

*Correspondence:
thomas.bourguignon@oist.jp (T.B.)
yukihiro.kinjo@oist.jp (Y.K.)
nathan.lo@sydney.edu.au (N.L.)

Keywords: endosymbionts; genetic drift; genome evolution; genome size; gene loss; Muller's Ratchet; mutation rate; DNA repair; phylogenetic tree

SUMMARY

The evolutionary processes that drive variation in genome size across the tree of life remain unresolved. Effective population size (N_e) is thought to play an important role in shaping genome size [1-3], a key example being the reduced genomes of insect endosymbionts, which undergo population bottlenecks during transmission [4]. However, the existence of reduced genomes in marine and terrestrial prokaryote species with large N_e indicate that genome reduction is influenced by multiple processes [3]. One candidate process is enhanced mutation rate, which can increase adaptive capacity, but can also promote gene loss. To investigate evolutionary forces associated with prokaryotic genome reduction, we performed molecular evolutionary and phylogenomic analyses of nine lineages from five bacterial and archaeal phyla. We found that gene loss rate strongly correlated with synonymous substitution rate (a proxy for mutation rate) in seven of the nine lineages. However, gene loss rate showed weak or no correlation with the ratio of nonsynonymous/synonymous substitution rate (d_N/d_S). These results indicate that genome reduction is largely associated with increased mutation rate, while the association between gene loss and changes in N_e is less well defined. Lineages with relatively high d_S and d_N, as well as smaller genomes, lacked multiple DNA repair genes, providing a proximate cause for increased mutation rates. Our findings suggest that similar mechanisms drive genome reduction in both intracellular and free-living prokaryotes, with implications for developing a comprehensive theory of prokaryote genome size evolution.

RESULTS AND DISCUSSION

Genome size varies dramatically across the tree of life. Among unicellular organisms, genomes differ in size by over six orders of magnitude [3, 5]. The evolutionary drivers of this variation remain unresolved [6]. One evolutionary parameter that is thought to shape genome size in both eukaryotes and prokaryotes is effective population size (N_e), which determines the rate of genetic drift [1-3]. An important example comes from the genomes of mutualistic insect endosymbionts, which are widely considered to undergo long-term degradation as a result of reductions in N_e caused by population bottlenecks during mother-to-offspring transmission [4, 7-9]. However, a number of free-living bacterial lineages with large N_e have reduced genomes [10], indicating the existence of alternative paths to genome reduction [3, 8].

Additional processes that can explain genome reduction include removal of selective constraints in the case of intracellular endosymbionts [11], and streamlining in the case of marine bacteria [8, 12]. A separate potential driver of genome reduction is enhanced mutation rate [8, 13, 14]. Increased mutation rates can facilitate rapid adaptation in organisms exposed to novel environments [15], an example being bacteria that have recently become intracellular [16]. Such increases can also lead to enhanced gene erosion and loss [8, 13]. The potential role of increased mutation rate in driving prokaryote genome reduction has received relatively little attention [17] and lacks empirical support [3, 8, 18].

The influences of different evolutionary processes on genome reduction can be disentangled in a phylogenetic framework. Because mutations at synonymous sites are selectively neutral (assuming that selection on synonymous codon usage is weak [19]), the rate of synonymous substitutions (d_S) provides a good approximation of mutation rate [20]. On the other hand, the rate of nonsynonymous substitutions (d_N) is affected both by selection and the mutation rate. By comparing rates of gene loss with d_N and d_S across a phylogeny, we can assess the relative importance of changes in N_e, mutation rate, and selection on genome degradation [21].

Previous studies of the influence of these processes on bacterial genome evolution have typically compared a few reduced-genome taxa with distantly related taxa possessing larger genomes [9, 17], or have compared several distantly related taxa [1]. We took a novel approach, performing molecular evolutionary analyses in a

phylogenetic framework on closely related strains or species with varying genome sizes. We examined nine lineages from five bacterial and archaeal phyla (Bacteroidetes, Proteobacteria, Cyanobacteria, Actinobacteria, and Euryarchaeota) that displayed notable variation in genome size among closely related taxa, and for which we were able either to generate representative genomic data or to retrieve data from GenBank. Because we compared closely related taxa, we assumed that the influence of codon usage bias on d_S was approximately equal across members of a given lineage. The intracellular endosymbiont lineages that we investigated (*Blattabacterium cuenoti* and *Buchnera aphidicola*) are not known to share their host cells with secondary symbionts that have undergone long-term co-cladogenesis with their hosts [22, 23]. Long-term secondary symbionts might cause extreme genome reduction via removal of selective constraints on redundant genes [24, 25], which could confound interpretation of the roles of mutation rates and N_e on gene loss.

Increased mutation rate is strongly associated with gene loss in Blattabacterium *and* Buchnera *endosymbionts*

We first examined the genomes of 67 *Blattabacterium cuenoti* (hereafter *Blattabacterium*) strains from cockroach and termite hosts that represent the eight dictyopteran families known to harbor this endosymbiont, including 46 sequenced for the present study. *Blattabacterium* is an obligate intracellular mutualist which participates in host nitrogen recycling [26-29] and has been strictly transmitted from mother to offspring for >200 Myr [30, 31]. Genome sizes were found to vary from 511 to 645 kb among strains. We estimated a maximum-likelihood phylogenetic tree using a set of 353 genes present in the genomes of all 67 taxa. We then reconstructed the evolution of gene loss using a model that allowed gene loss but no gene gain, as is known to occur in intracellular mutualistic endosymbionts [11] (Figure 1).

A comparison of numbers of genes lost with phylogenetic root-to-tip distances for each strain revealed a positive correlation (*rho* = 0.701, Figure 2A). To examine the relative roles of mutation rate, reduced N_e, and selection on rates of gene loss, we calculated d_S/time and d_N/time along each branch of the phylogeny using the alignment of 353 conserved genes, and performed phylogenetic generalized least-squares regression on terminal branch values (time duration of each branch was

estimated using Bayesian analysis). Because the genes used in these analyses have never been lost during the evolution of *Blattabacterium*, the removal of selective constraints is not expected to have played a major role in their evolution. We found a positive correlation between gene-loss rate (per Myr) and both d_S/time (r^2 = 0.313, $p = 10^{-4}$, Figure 2B) and d_N/time (r^2 = 0.231, p = 0.001, Figure 2C). We estimated d_N/d_S along each terminal branch across the tree and found a positive, albeit weak, correlation with per-branch gene-loss rate (r^2 = 0.036, p = 0.228, Figure 2D). We performed ranked correlation analysis across all branches, which corrected for biases associated with estimation of d_S for long branches (d_S > 1.5; as a result of substitutional saturation) and short branches (d_S < 0.2). We found a positive correlation between gene-loss rate (per Myr) and both d_S/time (*rho* = 0.467, $p = 6.81\times10^{-5}$) and d_N/time (*rho* = 0.443, $p = 1.73\times10^{-4}$). We estimated d_N/d_S across all branches of the tree and found no correlation with per-branch gene-loss rate (*rho* = -0.109, p = 0.379). Analyses in COEVOL [32], based on separate 1st+2nd and 3rd codon sites (as proxies for nonsynonymous and synonymous substitution sites, respectively), also revealed positive correlations between gene loss and evolutionary rate at these different site classes (r = 0.538, p < 0.01; r = 0.395, p < 0.01). These results indicated that gene loss is associated with increases in mutation rate, which are expected to raise both d_S and d_N [21], rather than with reductions in N_e (which are expected to lead to increases in d_N only). Increases in d_N might also be explained by positive selection, although this would not be expected to produce the genome-wide changes detected in our analyses.

We found heterogeneous GC-content across *Blattabacterium* strains, which potentially leads to biased estimates of d_N and d_S. To correct this bias, we used branch lengths estimated with nhPhyML [33] on 1st+2nd and 3rd codon sites as proxies for d_N and d_S, respectively. We found highly significant correlations between gene-loss rates and both d_S/time and d_N/time, and a marginally significant correlation between gene-loss rates and d_N/d_S (Data S1A). In the analyses described above, we used ratios as measures of evolutionary rates and gene-loss rates, an approach that might introduce spurious correlations [34]. To correct for any potential biases in our analyses, we performed partial correlation analysis, using the residual values of three linear regressions: 1) branch lengths calculated for 3rd codon sites *vs* time (as a proxy for d_S, referred to here as 'time-controlled d_S'); 2) branch lengths calculated

for 1st+2nd codon sites *vs* branch lengths calculated for 3rd codon sites (as a proxy for d_N/d_S, referred to here as 'd_S-controlled d_N'); and 3) gene loss *vs* time (referred to as 'time-controlled gene loss'). Correlations of time-controlled gene loss with time-controlled d_S and d_S-controlled d_N indicate respective associations of gene-loss rate with mutation rates and N_e. We found a positive correlation between gene-loss rate and time-controlled d_S (*rho* = 0.443, $p < 0.001$, Figure 3A), but not between gene-loss rate and d_S-controlled d_N (*rho* = 0.229, $p = 0.071$, Figure 3B), confirming that gene loss is strongly associated with mutation rate in *Blattabacterium*.

We repeated the analyses described in the previous paragraphs on 47 strains of *Buchnera aphidicola* (hereafter *Buchnera*), an obligate endosymbiont from the phylum Proteobacteria with genome sizes varying from 412 to 646 kb. *Buchnera* infected the ancestor of all aphids >150 Ma, and has been passed down from mother to offspring since that point [35], being occasionally lost in some aphid lineages [36]. We reconstructed a maximum-likelihood phylogenetic tree for *Buchnera*, inferred the evolution of gene loss, and performed correlation analyses equivalent to those described for *Blattabacterium*. We found significant correlations between gene-loss rates and both d_S/time and d_N/time, but not between gene-loss rates and d_N/d_S (see Data S1B). Partial correlation analysis confirmed these results: time-controlled gene loss was strongly correlated with time-controlled d_S (*rho* = 0.619, $p < 10^{-6}$, Figure 3C), but not with d_S-controlled d_N (*rho* = 0.196, $p = 0.080$, Figure 3D). Therefore, similar to the case for *Blattabacterium*, gene loss in *Buchnera* correlates with mutation rate, while the effect of changes in N_e on genome evolution is less clear.

Gene loss is associated with mutation rate in multiple free-living prokaryote lineages

We performed the analyses described above on seven additional free-living lineages. Because these taxa can obtain new genetic material through horizontal transfer, we estimated total gene loss per branch using a model that allowed both gene loss and gain. For estimations of d_N and d_S we used a set of 31 core genes that are unlikely to have been the subject of lateral gene transfer. We initially examined two lineages known for possessing reduced genomes: the marine cyanobacterial group *Prochlorococcus*+*Synechococcus* (genome sizes range from 1.64 to 2.79 Mb, $n = 28$), and the archaean genus *Thermococcus* (genome sizes range from 1.52 to 2.16 Mb, $n = 19$). *Prochlorococcus* and *Synechococcus* comprise some of the most

abundant bacterial species on earth [37], while *Thermococcus* is a genus of hyperthermophilic archaea found in hydrothermal vents [38]. Multiple analyses consistently revealed significant correlations between gene-loss/time and both d_S/time and d_N/time, but not d_N/d_S, in each of these groups (Data S1C–D). Partial correlation analysis revealed that time-controlled gene loss significantly correlates with time-controlled d_S but not with d_S-controlled d_N in both *Prochlorococcus*+*Synechococcus* (*rho* = 0.478, $p = 0.002$; *rho* = -0.243, $p = 0.120$) (Figures 3E–F), and *Thermococcus* (*rho* = 0.881, $p < 10^{-6}$; *rho* = 0.191, $p = 0.383$) (Figures 3G–H). These results indicate that increased mutation rate is strongly associated with genome reduction in free-living bacteria and archaea with reduced genomes. Because codon usage bias has been detected in strains of *Synechococcus*, but not *Prochlorococcus* [39], we repeated our analyses examining only members of the latter genus. We found a highly similar correlation between time-controlled d_S and time-controlled gene-loss (rho = 0.475) compared with the full data set, although significance was marginal ($p = 0.054$), possibly due to the lower number of taxa examined (n = 18). Similar to the results using the full data set, there was no correlation between time-controlled gene loss and d_S-controlled d_N (rho = -0.281, $p = 0.256$). These results indicate that codon usage bias does not have a major effect on our results.

We analysed a further five free-living lineages from a range of habitats. In three of these lineages, *Corynebacterium* (genome sizes range from 2.45 to 3.57 Mb, $n = 18$), Micrococcineae (genome sizes range from 1.43 to 5.05 Mb, $n = 22$), and Flavobacteriaceae (genome sizes range from 2.09 to 6.09 Mb, $n = 33$), we found results similar to those obtained for *Blattabacterium*, *Buchnera*, *Prochlorococcus*+*Synechococcus*, and *Thermococcus* (Figures 3K–L, 3O–R, Data S1F, H–I), although a significant correlation between gene loss and d_N/d_S was found in *Corynebacterium* (*rho* = -0.685, $p < 10^{-4}$), and between time-controlled gene loss and d_S-controlled d_N in the case of Micrococcinae (*rho* = 0.388, $p = 0.031$). In the remaining two lineages, Gammaproteobacteria (genome sizes range from 1.70 to 5.01 Mb, $n = 20$) and Mycobacteriaceae+Nocardiaceae (genome sizes range from 3.28 to 9.70 Mb, $n = 15$), we did not find consistent evidence for correlations between gene-loss/time and d_S/time or d_N/time (Data S1E, G), and similar results were found in our partial correlation analysis (Figures 3I–J, 3M–N). In the case of

Mycobacteriaceae+Nocardiaceae, a correlation was found between time-controlled gene loss and d_S-controlled d_N (*rho* = 0.47, $p = 0.028$) (Figure 3N). These results suggest a potential influence of N_e on genome reduction in Mycobacteriaceae+Nocardiaceae and Micrococcinae. Overall, these results indicate that the association between mutation rate and gene-loss rate applies to free-living bacterial groups with larger genomes, albeit not universally.

Proximate causes of increased mutation rates and genome reduction

Our results provide the first phylogenomic evidence for a link between increased mutation rate and long-term prokaryotic genome reduction, based on analyses of closely related taxa. We found evidence for this link in seven of the nine phylogenetically and ecologically divergent lineages that we tested. Previous studies have noted an inverse relationship between microbial genome size and mutation rate (per base pair, per replication) [40-42]; however, these studies examined relatively few, distantly related taxa, and did not specifically look at the process of gene loss and molecular evolution in a phylogenetic framework.

Proximate causes of the increased mutation rates that we identified are likely to include the loss of DNA repair genes [3, 8, 13] and reductions in the accuracy of replication enzymes. In *Prochlorococcus*, low-light-adapted ecotypes have lower mutation rates and have retained a larger set of DNA repair genes than high-light-adapted ecotypes [37]. The *Buchnera* strains endowed with the smallest genomes are those associated with Lachninae, Calaphidinae, and Phyllaphidinae, all of which possess reduced repair machinery in comparison with other strains of *Buchnera* [43, 44]. In *Blattabacterium*, taxa with small genomes have a significantly greater loss of genes in COG categories F (nucleotide metabolism) and L (DNA replication and repair) than do other clades (Figure S1, Table S1). Genes in these categories are thought to play key roles in reducing or removing errors that occur during DNA replication.

An inverse correlation between genome size and loss of DNA repair enzymes has been found across numerous prokaryotic taxa [45]. An increased mutation rate can lead to increased levels of gene inactivation and erosion through deletions or nonsense mutations [13, 46]. According to the “error threshold” theory, genes are

lost when the mutation rate exceeds the fitness effects of such gene loss [13, 47]. Because fitness effects vary among genes, enhanced mutation rates will remove genes that are less important in the genome.

Ultimate causes of increased mutation rates and genome reduction

Although we identified a strong correlation between mutation rate and gene loss across multiple lineages, causation may be in either direction, or there might be no causal link between the two phenomena. The ultimate causes of increases in rates of mutation and gene loss could be adaptive, neutral, or a combination of both. Below we briefly consider a number of hypotheses for the ultimate causes of genome reduction in the light of our results.

Enhanced mutation rates have been hypothesized to provide adaptive advantages in prokaryotes [48]. A 'mutator' strain that evolved via modification or loss of DNA repair genes or lower fidelity polymerases might initially be selected because of its capacity to rapidly accrue beneficial mutations in novel environments. Increased mutations in such a strain, which could be either free-living or endosymbiotic, would lead to increased gene deterioration and loss, which could lead to increased fitness due to the removal of functions with a high cost-to-benefit ratio [14, 16, 49, 50]. Under this scenario, the adaptive benefits of increased mutation rate would be the ultimate cause of genome reduction, given increased mutation rates were maintained during the evolution of the lineage.

The streamlining hypothesis for genome reduction in marine cyanobacteria proposes that strong selection acts to remove non-essential genes in ocean environments low in nitrogen and phosphorus (which are essential elements of DNA) [12, 51]. A small genome also permits small cell volume, which improves nutrient uptake [52, 53]. One interpretation of the increased mutation rates that we observed in *Prochlorococcus* spp. could be that they are a consequence of streamlining, stemming from the removal of non-essential DNA-repair genes. The streamlining hypothesis has been considered unlikely to apply to bacteria other than marine bacterioplanktons [1]. However, selection for both increased mutation rate and minimal use of DNA could provide an explanation for genome reduction in a variety of prokaryotes. For example, in hosts that persist on nutritionally restrictive diets, host-level selection for endosymbionts that consume fewer critical nutrients could

lead to reduced endosymbiont genome size. During this process, individuals with a higher mutation rate would be selected as they would be likely to lose genes more quickly than individuals with slower rates.

Hypotheses that require non-adaptive processes to explain increases in rates of mutation and gene loss include those based on removal of selective constraint and N_e. In the former, gene loss occurs because no fitness advantage is provided by retention of particular genes, while in the latter, enhanced genetic drift due to population bottlenecks leads to the fixation of deleterious mutations, ultimately resulting in gene erosion and loss [4, 9]. In each of these cases, the ultimate cause of increases in mutation rate is the non-adaptive loss or degradation of DNA repair genes. A reduction in polymerase fidelity as a result of fixation of mildly deleterious mutations via drift could also contribute to increased mutation rates. Based on the lack of correlation between d_N/d_S and gene loss, we found no evidence for an effect of reduced N_e on genome reduction during the diversification of the lineages that we examined, although we cannot rule out such an effect.

Our results show links between increased mutation rates and genome reduction in endosymbiotic and multiple free-living bacterial lineages. Our findings are consistent with previous hypotheses for genome reduction in some free-living bacterial lineages, but also suggest that currently accepted explanations for endosymbiont genome reduction require revision. The hypothesis that adaptive benefits of increased mutation rates during the early evolution of a lineage ultimately lead to long-term genome reduction should be tested in future studies, and considered in the development of a comprehensive theory of prokaryote genome-size evolution.

ACKNOWLEDGMENTS

We thank Ales Buček for assistance with phylogenetic analysis, David Rentz and James Walker for providing samples used in this study, John Alroy for providing advice regarding statistical analyses, and Eddie Holmes for comments on the manuscript. T.B. was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI 18K14767, by the subsidiary funding to OIST, by a University of Sydney Postdoctoral Fellowship, and by the Internal Grant Agency of the Faculty of Tropical AgriSciences, CULS (20205014). D.A.A. was supported by an Australian

Government Scholarship. Z.W. and GT respectively acknowledge the support of the National Natural Science Foundation of China (31672329, 31872271) and JSPS KAKENHI 17H01510. S.Y.W.H. and N.L. were supported by the Australian Research Council.

AUTHOR CONTRIBUTIONS

TB, YK, and NL conceptualized the experiments, with input from SYWH. TB, QT, ZW, and NL collected the samples. TB performed lab experiments and generated data. YK analysed the data, with significant input from TB, NVC, DAA, PVM, SM, and NL. TB wrote the first draft of the manuscript. NL and TB wrote subsequent drafts of the manuscript, with significant input from SYWH, YK, PVM, NVC, GT, and SP.

DECLARATION OF INTERESTS

The authors declare no competing interests.

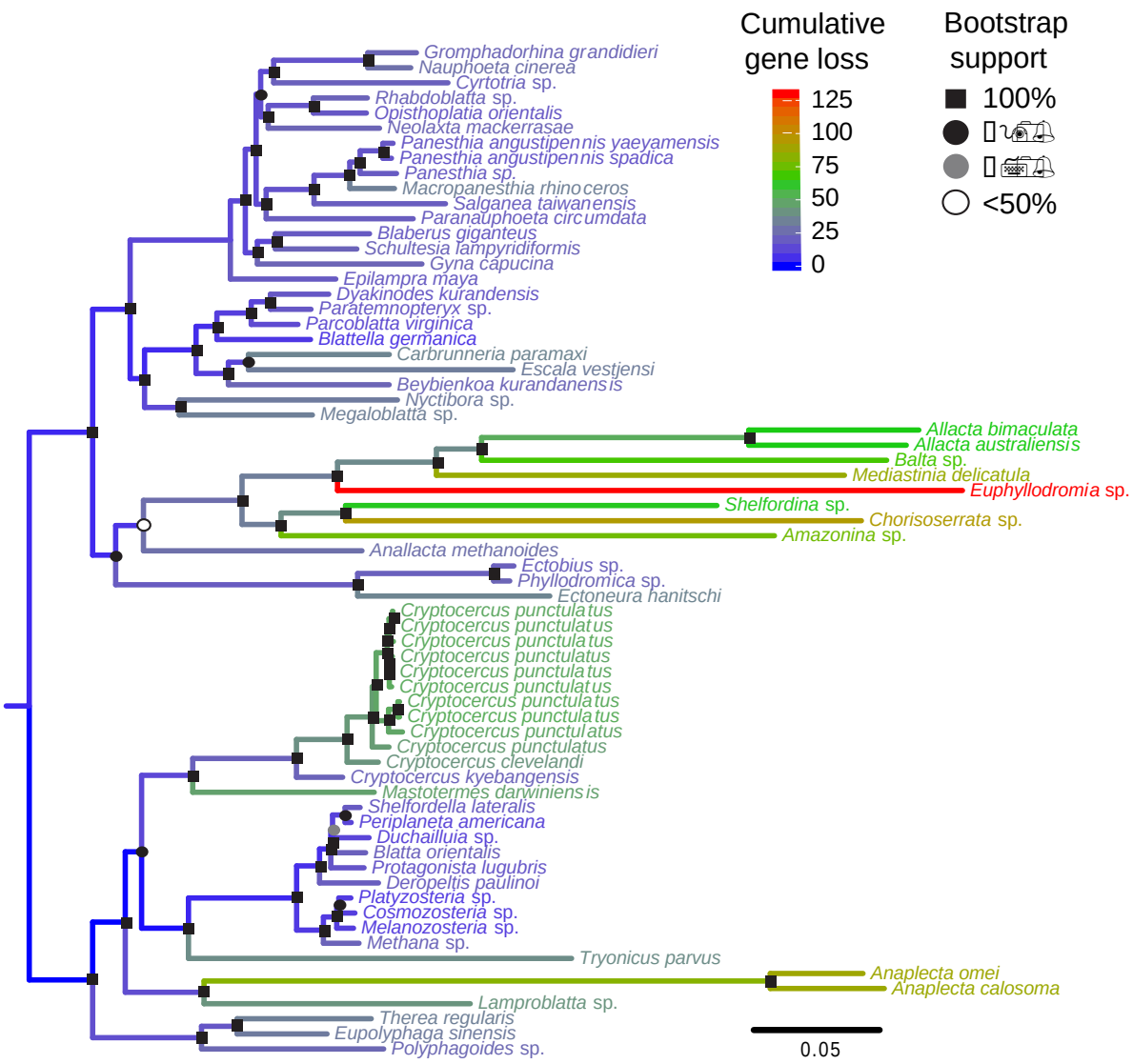


Figure 1. Phylogenetic tree of *Blattabacterium* inferred using maximum-likelihood analysis of 353 protein-coding genes, with 3rd codon sites removed. Branch color represents cumulative gene loss. Node symbols indicate bootstrap support values.

See also Figure S1, Table S1 and Table S2.

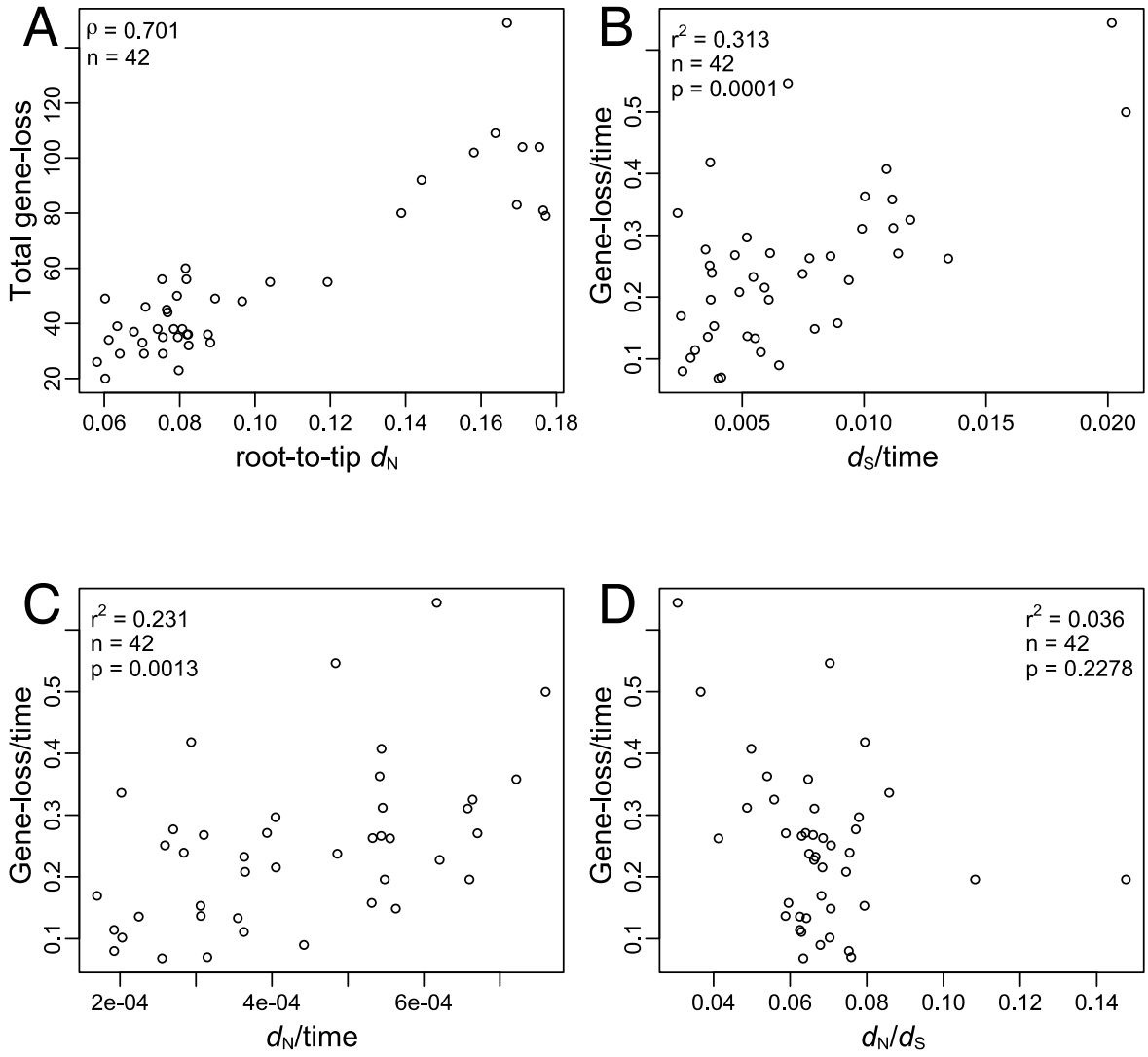


Figure 2. Evolution of genome reduction in *Blattabacterium*.

(A) Relationship between total number of gene losses and root-to-tip d_N distance (inferred from the tree represented in Figure 1) for each strain. Phylogenetic generalized least-squares regression implemented in the R package CAPER between (B) gene loss/time and d_S/time, (C) gene loss/time and d_N/time, and (D) gene loss/time and d_N/d_S, per terminal branch.

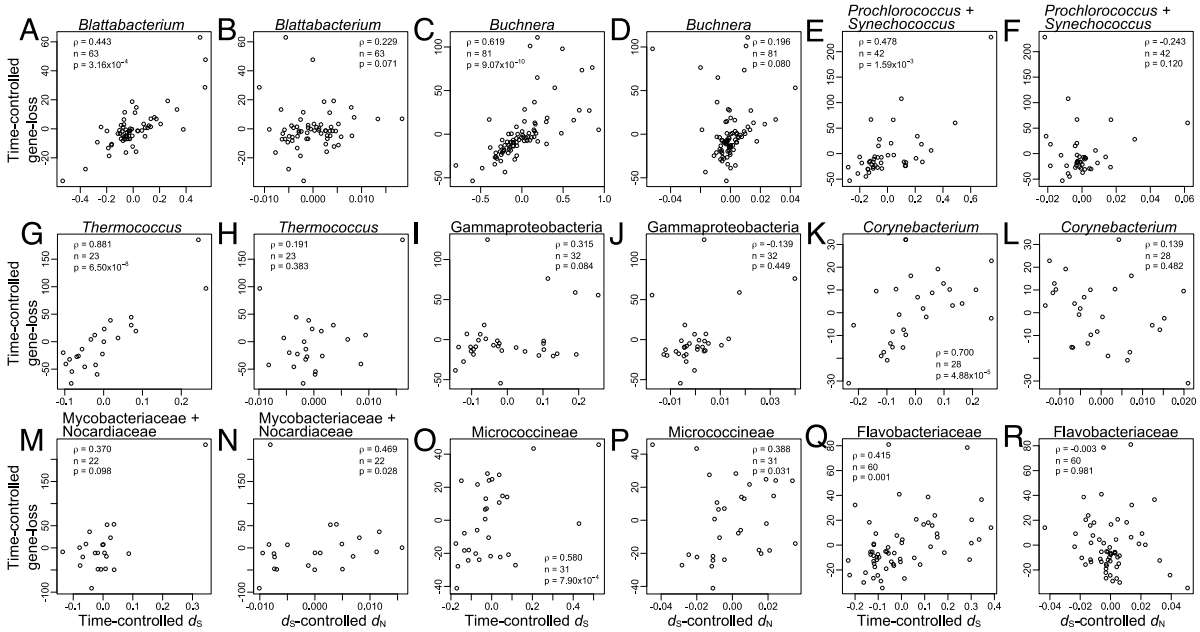


Figure 3. Partial correlation analysis of time-controlled gene loss with time-controlled d_S and d_S-controlled d_N for nine prokaryote lineages (see text for details).

Correlation between time-controlled gene loss and time-controlled d_S for (A) *Blattabacterium*, (C) *Buchnera*, (E) *Prochlorococcus* + *Synechococcus*, (G) *Thermococcus*, (I) Gammaproteobacteria, (K) *Corynebacterium*, (M) Mycobacteriaceae + Nocardiaceae, (O) Micrococcineae, and (Q) Flavobacteriaceae. Correlation between time-controlled gene loss and d_S-controlled d_N for (B) *Blattabacterium*, (D) *Buchnera*, (F) *Prochlorococcus* + *Synechococcus*, (H) *Thermococcus*, (J) Gammaproteobacteria, (L) *Corynebacterium*, (N) Mycobacteriaceae + Nocardiaceae, (P) Micrococcineae, and (R) Flavobacteriaceae. See also Data S1.

STAR METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Cockroach samples used for RNA isolation	This Study, see Table S2	N/A
Chemicals, Peptides, and Recombinant Proteins		
RNA-later®	ThermoFisher Scientific	Cat#AM7021
Fisherbrand™ Disposable Pestle System	Fisher Scientific	Cat#03-392-103
DNeasy Blood and Tissue extraction kit	Qiagen	Cat#69506
Qubit	ThermoFisher Scientific	Cat#Q32854
Deposited data		

Blattabacterium genomes associated with 46 cockroach species	This Study, see Table S2	N/A
Software and Algorithms		
COEVOL	[32]	https://megasun.bch.umontreal.ca/People/lartillot/www/
nhPhyML	[33]	http://pbil.univ-lyon1.fr/software/nhphyml/
BLAST+ package	[54]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
TCSF and IMRA	[55]	https://github.com/Yukihirokinjo/TCSF_IMRA
GapFiller	[56]	https://sourceforge.net/projects/gapfiller/
Pilon	[57]	https://github.com/broadinstitute/pilon
Prodigal	[58]	https://github.com/hyattpd/Prodigal
COG database	[59]	https://www.ncbi.nlm.nih.gov/COG/
RNAmmer	[60]	http://www.cbs.dtu.dk/services/RNAmmer/
tRNAscan-SE	[61]	http://lowelab.ucsc.edu/tRNAscan-SE/

Infernal	[62]	http://eddylab.org/infernal/
Proteinortho ver. 5.16	[63]	https://www.bioinf.uni-leipzig.de/Software/proteinortho/manual5.html
FIGfam	[64]	http://blog.theseed.org/servers/presentations/t1/figfams.html
CD-search	[65]	https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi
MAFFT v7.300b	[66]	http://mafft.cbrc.jp/alignment/software/
pal2nal v14	[67]	https://github.com/HajkD/orthologr/tree/master/inst/pal2nal
IQ-TREE 1.6.7	[68]	http://www.iqtree.org
BEAST1.8.4	[70]	https://beast.community
AMPORA2	[71]	http://wolbachia.biology.virginia.edu/WuLab/Software.html.
trimAl	[72]	http://trimal.cgenomics.org
Tracer 1.5	[74]	http://tree.bio.ed.ac.uk/software/tracer/
Paleobiology		https://www.paleobiodb.org/#/

Database		
RAxML version 8.2	[77]	https://cme.h-its.org/exelixis/web/software/raxml/
PAML4	[78]	http://evomics.org/resources/software/molecular-evolution-software/paml/
ape	[79]	https://cran.r-project.org/web/packages/ape/index.html
phytools	[81]	https://cran.r-project.org/web/packages/phytools/
CAPER	[82]	https://cran.r-project.org/web/packages/caper/index.html
ppcor	[83]	https://cran.r-project.org/web/packages/ppcor/index.html

RESOURCE AVAILABILITY

Lead Contact

Further information and requests may be directed to and will be fulfilled by the lead contact Thomas Bourguignon (thomas.bourguignon@oist.jp). Yukihiro Kinjo (yukihiro.kinjo@oist.jp), and Nathan Lo (nathan.lo@sydney.edu.au) may also be contacted for further information.

Material Availability

This study did not generate new unique reagents.

Data and Code Availability

The assembled genomes of *Blattabacterium* generated in this study are freely available on NCBI under the accession PRJNA643811.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We obtained samples of 46 cockroach species preserved in RNA-later®. Cockroaches were shipped at room temperature to Sydney, Australia, where they were stored at -80 ºC until DNA extraction. Details on individual sample collection can be found in Table S2.

METHOD DETAILS

Blattabacterium *sequencing*

Fat bodies of a single cockroach specimen were dissected using a sterile scalpel and DNA was extracted with a DNeasy Blood and Tissue kit (Qiagen). DNA extraction was performed according to the manufacturer's protocol. Cockroach fat-body DNA, which includes *Blattabacterium* DNA, was sequenced during multiple Illumina runs. For the first run, DNA samples of 23 cockroach specimens were tagged with unique barcode combinations, mixed in equimolar concentration, and 150 bp paired-end-reads-sequenced on an Illumina HiSeq4000. From this initial sequencing run, 10 *Blattabacterium* genomes were each assembled in a single circular chromosome (Table S2), while the remaining 13 genomes were each split into several contigs. In the second run, we used the same procedure and sequencing platform and sequenced fat-body DNA of 18 cockroach species, two of which were specimens re-sequenced from the first run (Table S2). In total, this sequencing run yielded four *Blattabacterium* genomes, each assembled in a single circular chromosome (Table S2).

To improve the assembly of fragmented genomes, we re-sequenced specimens over 11 runs of Illumina HiSeq X Ten. The fat-body DNA from two to six species, belonging to different cockroach families or subfamilies (i.e., divergent taxa), was mixed prior to library preparation and sequenced in one run of Illumina HiSeq X Ten. Therefore, the reads obtained from these sequencing runs included DNA from several *Blattabacterium* strains, which were assembled together. We observed no interaction between *Blattabacterium* genomes during the assembling

steps. Each *Blattabacterium* contig could be unambiguously attributed to a single cockroach species using blastn searches, implemented in the BLAST+ package [54].

QUANTIFICATION AND STATISTICAL ANALYSIS

Blattabacterium *genome assembly and annotation*

We assembled high-quality reads using the "TCSF and IMRA" pipeline as previously described [55]. Unknown regions within scaffolds were determined using GapFiller [56]. For each species, we evaluated the final assembly and corrected erroneous regions using Pilon [57]. Regions of low quality, characterized by a high probability of being misassembled, were removed and masked with "N".

We annotated a total of 67 *Blattabacterium* genomes, 46 of which were sequenced in this study. The remaining 21 genomes were downloaded from RefSeq (Table S2). We predicted protein-coding regions using Prodigal [58] with a cut-off score of 0.6. In addition to the Prodigal prediction, we carried out homology-based open reading frame prediction using blastp search, implemented in the BLAST+ package [54], against the COG database [59]. Predictions for rRNAs, tRNAs, and other non-coding RNAs were carried out using RNAmmer [60], tRNAscan-SE [61], and Infernal [62], respectively.

Pseudogenes were identified by checking for fragmentation and truncation of open reading frames. Briefly, we used blastp to search each predicted gene against the predicted orthologous protein sequences of eight published *Blattabacterium* genomes. We used an e-value of 10^{-30} as the threshold. Genes with fragmented open reading frames and with disrupted conserved functional motifs or domains were regarded as pseudogenes. We used CDD searches to identify functional motifs and domains. Truncated genes missing more than 30% of typical mean gene length, and missing complete functional motifs or domains, were also considered as pseudogenes.

We determined all sets of orthologous genes from all genomes used in this study using Proteinortho ver. 5.16 [63] with the parameter -cov = 35. All orthologous gene sets were further curated manually, and only those shared among at least five strains were used for our evolutionary analyses. In addition, to remove uncertainties from the prediction of orthologous gene sets, orthologous gene sets with low clustering confidence scores (<0.6) were removed from the analyses.

Functional annotation of each predicted orthologous gene set was carried out using FIGfam [64]. Annotation was further curated using CD-search [65] against COG database.

Phylogenetic analysis of Blattabacterium

We carried out phylogenetic analyses on 67 strains of *Blattabacterium* using 353 orthologous protein-coding genes that were present across all strains, and did not consider further the genes that were absent in one or more strains. We aligned the amino acid complement of each gene with MAFFT v7.300b using the option "--maxiterate 1000 --globalpair" for maximum accuracy [66]. Amino acid sequence alignments were back-translated to nucleotides using pal2nal v14 [67], and stop codons were masked as "NNN".

The concatenated sequence alignment was partitioned into three subsets, one for each codon position of the protein-coding genes. We removed the 3rd codon sites from subsequent phylogenetic analyses, and partitioned our data set into two subsets: one containing the 1st codon sites and one containing the 2nd codon sites. A maximum-likelihood phylogenetic tree was reconstructed with IQTREE version 1.6.7 [68] using ultrafast bootstrapping and 1000 replicates [69].

Molecular dating of Blattabacterium

We inferred a time-calibrated phylogenetic tree for *Blattabacterium* using BEAST 1.8.4 [70]. Because BEAST analyses are computationally intensive, we ran the analyses with a subset of 31 genes from the 353 genes used for our maximum-likelihood phylogenetic analysis with IQ-TREE. The 31 selected genes were standard bacterial phylogenetic marker genes used in AMPORA2 [71]. Each gene was aligned independently and the 31 gene alignments were concatenated as described above. We further trimmed the concatenated alignment matrix, removed the 3rd codon sites, and removed each column containing gaps using trimAl [72]. The final sequence alignment included 14,100 nucleotide sites.

We partitioned our data set into two subsets: one containing 1st codon sites and one containing 2nd codon sites. An independent GTR+G model of nucleotide substitution was assigned to each subset. We implemented an uncorrelated lognormal relaxed clock to account for rate variation across branches [73]. For each

analysis, Markov chain Monte Carlo (MCMC) sampling was used to estimate the tree and the posterior distributions of parameters. Each MCMC analysis was performed in duplicate. The MCMC chains were run for 10^8 steps and the parameter values were sampled every 10^4 steps. Following inspection of the MCMC samples in Tracer 1.5 [74], we discarded the samples from the first 10^7 steps as burn-in. The marginal log-likelihood of the tree inferred with a birth-death tree prior was -174,154, whereas that of the tree inferred with a Yule process was -174,712. Therefore, we only present the tree inferred using a birth-death tree prior [75].

The molecular clock was calibrated using seven minimum age constraints (Table S3). Each calibration was based on the fossil record and we systematically selected the youngest possible age for each fossil, as mentioned on the Paleobiology Database (www.paleobiodb.org; last accessed on 27 July 2018). Fossil calibrations were implemented as exponential priors on node times [76]. In each case the 97.5% soft maximum bound was determined using a combination of phylogenetic bracketing and absence of fossil evidence (Table S3).

Phylogenetic analysis of Buchnera *and free-living bacteria*

We obtained genomic data from the RefSeq database. For each lineage, we used the CD-HIT program to remove redundant genomes, which we defined as genomes with nucleotide identity, on the marker gene alignment without 3rd codon positions, upwards of 96%. As a result, we obtained 46 genomes of *Buchnera*, 28 genomes of *Prochlorococcus* and *Synechococcus*, 19 genomes of *Thermococcus*, 18 genomes of *Corynebacterium*, 22 genomes of Micrococcineae, 33 genomes of Flavobacteriaceae, 20 genomes of Gammaproteobacteria, and 15 genomes of Mycobacteriaceae and Nocardiaceae (Table S4). We predicted gene orthology and carried out alignments as described above. We inferred phylogenetic trees using maximum-likelihood analysis of 30 bacterial phylogenetic marker genes for *Buchnera*, and 31 bacterial phylogenetic marker genes for all other lineages. The marker genes were those used in AMPORA2 [71]. The alignment was then recoded into RY (A/G to R, T/C to Y) to avoid bias caused by heterogenous nucleotide composition in the alignments. Phylogenetic trees were reconstructed using RAxML version 8.2 [77] with the BINGAMMA binary character substitution model.

Timetree reconstruction of Buchnera *and free-living bacteria*

We used MCMCtree implemented in the PAML4 package [78] to estimate divergence times, using the alignment generated for the maximum-likelihood phylogenetic analysis. We used the GTR+G nucleotide substitution model and the log-normal correlated clock model to model rate variation across branches with the following priors: rgene_gamma = 1, 15; sigma2_gamma = 1, 10. The MCMC chains were run for 5.05×10^5 steps and the parameter values were sampled every 50 steps. The first 5,000 steps were discarded as burn-in. We ran two independent MCMC chains with different random seed values, and confirmed convergence. The molecular clocks were calibrated using two minimum age constraints for *Buchnera* and one minimum age constraint for *Prochlorococcus* and *Synechococcus* (Table S3). For other free-living prokaryote lineages, we set the root of the tree to an arbitrary depth of 1 to obtain time-related branch lengths.

Reconstruction of gene loss

We reconstructed the evolution of gene loss using the function “ace” from the R package ape [79]. The presence or absence of each gene was treated independently as a discrete binary character and the ancestral state was estimated using maximum likelihood [80]. For *Blattabacterium* and *Buchnera*, the model was specified using the option “model= matrix(c(0, 1, 0, 0), 2)” which assumes no gene gain. For the seven lineages of free-living bacteria, including *Prochlorococcus*+*Synechococcus*, *Thermococcus*, *Corynebacterium*, Micrococcineae, Flavobacteriaceae, Gammaproteobacteria, and Mycobacteriaceae+Nocardiaceae, we selected the all-rates-different model, which allows unequal rates of gene loss and gain. We ran these analyses on each maximum-likelihood tree. The result of each reconstruction was visualized using the function “plotTree” in the R package phytools [81]. We also used the cumulative maximum-likelihood estimate of gene loss to plot the rate of gene loss across each tree.

Correlation of gene loss with evolutionary rate and d_N/d_S

We investigated the relationship between gene loss and evolutionary rate in *Blattabacterium* using a combination of methods. First, we calculated the Spearman’s rank correlation coefficient between the total number of genes lost by

each strain and phylogenetic root-to-tip distances. To correct for the phylogenetic non-independence of data points in our analyses of root-to-tip distances, we calculated for each branch: a) gene-loss rate per unit of time (based on a subset of genes analysed in BEAST); and b) d_N and d_S per unit of time, and d_N/d_S. As we did for *Buchnera*, we used CD-HIT to remove genomes with upwards of 96% nucleotide identity on the marker gene alignment without 3rd codon positions. We calculated d_N, d_S and d_N/d_S using codeml implemented in PAML4 [78] with the F3x4 codon substitution model on a concatenated alignment of 30 core protein-coding genes shared by all *Blattabacterium* strains. We then carried out phylogenetic generalized least-squares regression using the pgls function implemented in the R package CAPER [82] with lambda value estimation (lambda="ML"). Other parameters were set to default values. To avoid possible bias caused by over-/under-estimation of d_S for short/long branches, we performed Spearman's rank correlation on all branches of the tree. We used the software COEVOL [32] to test for correlation between mutation rates and gene loss. We ran COEVOL twice: once with 1st+2nd codon sites, and once with 3rd codon sites (as proxies for nonsynonymous and synonymous substitution sites, respectively).

In addition to the above-described analyses carried out on *Blattabacterium* only, we carried out two more analyses on all lineages, including *Blattabacterium*. In these analyses, we used nhPhyML [33] on 1st+2nd and 3rd codon sites to correct for potential rate-estimation bias associated with heterogeneous GC-content, which we found to be present in all lineages. The branch lengths estimated by nhPhyML for 1st+2nd and 3rd codon sites were used as proxies for d_N and d_S, respectively. In the first analysis, we used d_N and d_S values calculated with nhPhyML to estimate d_S/time and d_N/d_S for each branch, except for short branches, whose length estimations are imprecise, and which were removed from the analyses. We then carried out Spearman's rank correlation on gene-loss/time vs nhPhyML-estimated d_S/time and d_N/d_S. In the second analysis, we carried out partial correlation analyses with the ppcor R package [83]. In the partial correlation analyses, we did not use ratios, such as gene-loss/time, d_S/time, and d_N/d_S, because comparisons of fractions can generate spurious correlations [34]. Instead, we used residual values obtained from linear regressions that we refer to as controlled variables. We carried out three linear regressions: 1) time against branch lengths calculated for 3rd codon positions (as a

proxy for d_S/time, referred to here as 'time-controlled d_S'); 2) branch lengths calculated for 3rd codon positions against branch lengths calculated for 1st+2nd codon position (as a proxy for d_N/d_S, referred to here as 'd_S-controlled d_N'), and 3) time against gene loss (referred to as 'time-controlled gene-loss'). We then carried out Spearman's rank correlation on time-controlled gene-loss vs time-controlled d_S and time-controlled gene-loss vs d_S-controlled d_N.

REFERENCES

1. Kuo, C.H., Moran, N., and Ochman, H. (2009). The consequences of genetic drift for bacterial genome complexity. Genome Res. *19*, 1450–1454.
2. Lynch, M. *et al.* (2016). Genetic drift, selection and the evolution of the mutation rate. Nat. Rev. Genet. *17*, 704–714.
3. Batut, B., Knibbe, C., Marais, G., and Daubin, V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. Nat. Rev. Microbiol. *12*, 841–850.
4. Moran, N.A., McCutcheon, J.P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. *42*, 165–190.
5. McGrath, C.L., and Katz, L.A. (2004). Genome diversity in microbial eukaryotes. Trends Ecol. Evol. *19*, 32–38.
6. Sela, I., Wolf, Y.I., and Koonin, E.V. (2016). Theory of prokaryotic genome evolution. Proc. Natl. Acad. Sci. *113*, 11399–11407.
7. Wernegreen, J.J. (2002). Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. *3*, 850–861.
8. Martínez-Cano, D.J. *et al.* (2015). Evolution of small prokaryotic genomes. Front. Microbiol. *6*, 1–23.
9. Moran, N.A. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. *93*, 2873–2878.
10. Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015). *Prochlorococcus*: the structure and function of collective diversity. Nat. Rev. Microbiol. *13*, 13–27.

11. McCutcheon, J.P., and Moran, N.A. (2012). Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. *10*, 13–26.
12. Giovannoni, S.J. *et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. Science *309*, 1242–1246.
13. Wernegreen, J.J. (2017). In it for the long haul: evolutionary consequences of persistent endosymbiosis. Curr. Opin. Genet. Dev. *47*, 83–90.
14. Marais, G.B., Calteau, A., and Tenaillon, O. (2008). Mutation rate and genome reduction in endosymbiotic and free-living bacteria. Genetica *134*, 205–210.
15. Giraud, A., Radman, M., Matic, I., and Taddei, F. (2001). The rise and fall of mutator bacteria. Curr. Opin. Microbiol. *4*, 582–585.
16. Clayton, A.L., Jackson, D.G., Weiss, R.B., and Dale, C. (2016). Adaptation by deletogenic replication slippage in a nascent symbiont. Mol. Biol. Evol. *33*, 1957–1966.
17. Itoh, T., Martin, W., and Nei, M. (2002). Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. Proc. Natl. Acad. Sci. U. S. A. *99*, 12944–12948.
18. Canbäck, B., Tamas, I., and Andersson, S.G.E. (2004). A phylogenomic study of endosymbiotic bacteria. Mol. Biol. Evol. *21*, 1110–1122.
19. Hershberg, R., and Petrov, D. (2008). Selection on codon bias. Annu. Rev. Genet. *42*, 287–299.
20. Bromham, L. (2011). The genome as a life-history character: why rate of molecular evolution varies between mammal species. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *366*, 2503–2513.
21. Bromham, L., Cowman, P.F., and Lanfear, R. (2013). Parasitic plants have increased rates of molecular evolution across all three genomes. BMC Evol. Biol. *13*, 126.
22. Bandi, C., Damiani, G., Magrassi, L., Grigolo, A., Fani, R., and Sacchi, L. (1994). Flavobacteria as intracellular symbionts in cockroaches. Proc. R. Soc. B Biol. Sci. *257*, 43–48.
23. Burke, G.R., Normark, B.B., Favret, C., and Moran, N.A. (2009). Evolution and diversity of facultative symbionts from the aphid subfamily Lachninae. Appl. Environ. Microbiol. *75*, 5328–5335.

24. McCutcheon, J.P., and Moran, N.A. (2007). Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc. Natl. Acad. Sci. U. S. A. *104*, 19392–19397.
25. Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R.P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A.C.C., *et al.* (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell *153*, 1567–1578.
26. Bandi, C. *et al.* (1995). The establishment of intracellular symbiosis in an ancestor of cockroaches and termites. Proc. R. Soc. B *259*, 293–299.
27. Cochran, D. (1985). Nitrogen excretion in cockroaches. Annu. Rev. Entomol. *30*, 29–49.
28. López-Sánchez, M.J. *et al.* (2009). Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. PLOS Genet. *5*, e1000721.
29. Sabree, Z.L., Kambhampati, S., and Moran, N.A. (2009). Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. Proc. Natl. Acad. Sci. U. S. A. *106*, 19521–19526.
30. Bourguignon, T., Tang, Q., Ho, S.Y.W., Juna, F., Wang, Z., Arab, D.A., Cameron, S.L., Walker, J., Rentz, D., Evans, T.A., *et al.* (2018). Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. Mol. Biol. Evol. *35*, 970–983.
31. Evangelista, D.A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M.K., Legendre, F., Liu, S., Machida, R., Misof, B., *et al.* (2019). An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). Proc. R. Soc. B Biol. Sci. *286*, 20182076.
32. Lartillot, N., and Poujol, R.A (2011). phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol. Biol. Evol. *28*, 729–744.
33. Boussau, B., and Gouy, M. (2006). Efficient likelihood computations with nonreversible models of evolution. Syst. Biol. *55*, 756–768.

34. Kronmal, R.A. (1993). Spurious correlation and the fallacy of the ratio standard revisited. J. R. Stat. Soc. *156*, 379–392.
35. Baumann, P., Moran, N.A., and Baumann, L. (1997). The evolution and genetics of aphid endosymbionts. Bioscience *47*, 12–20.
36. Chong, R. A., and Moran, N. A. (2018). Evolutionary loss and replacement of *Buchnera*, the obligate endosymbiont of aphids. ISME J. *12*, 898–908.
37. Partensky, F., and Garczarek, L. (2010). *Prochlorococcus*: advantages and limits of minimalism. Ann. Rev. Mar. Sci. *2*, 305–331.
38. Zillig, W., Holz, I., Janekovic, D., and Reiter, W.D. (1983). The archaebacterium *Thermococcus celer* represents, a novel genus within the thermophilic branch of the archaebacteria. Syst. Appl. Microbiol. *4*, 88–94.
39. Yu, T., Li, J., Yang, Y., Qi, L., Chen, B., Zhao, F., Bao, Q., and Wu, J. (2012). Codon usage patterns and adaptive evolution of marine unicellular cyanobacteria *Synechococcus* and *Prochlorococcus*. Mol. Phylogenet. Evol. *62*, 206–213.
40. Drake, J.W. (1991). Constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. U. S. A. *88*, 7160–7164.
41. Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F. (1998). Estimation of spontaneous mutation rates. Genetics *148*, 1667–1686.
42. Massey, S.E. (2008). The proteomic constraint and its role in molecular evolution. Mol. Biol. Evol. *25*, 2557–2565.
43. Pérez-Brocal, V. *et al.* (2006). A small microbial genome: the end of a long symbiotic relationship? Science *314*, 312–313.
44. Chong, R.A., Park, H., and Moran, N.A. (2019). Genome evolution of the obligate endosymbiont *Buchnera aphidicola*. Mol. Biol. Evol. *36*, 1481–1489.
45. Acosta, S., Carela, M., Garcia-Gonzalez, A., Gines, M., Vicens, L., Cruet, R., and Massey, S.E. (2015). DNA Repair Is Associated with Information Content in Bacteria, Archaea, and DNA Viruses. J. Hered. *106*, 644–659.
46. Moran, N.A., McLaughlin, H.J., and Sorek, R. (2009). The dynamics and time scale of ongoing genome erosion in symbiotic bacteria. Science *323*, 379–382.

47. Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften *58*, 465–523.
48. Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., and Godelle, B. (1997). Role of mutator alleles in adaptive evolution. Nature *387*, 700–702.
49. Funchain, P., Yeung, A., Stewart, J.L., Lin, R., Slupska, M.M., and Miller, J.H. (2000). The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. Genetics *154*, 959–970.
50. Koskiniemi, S., Sun, S., Berg, O.G., and Andersson, D.I. (2012). Selection-driven gene loss in bacteria. PLoS Genet. *8*, e1002787.
51. Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R., *et al.* (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature *424*, 1042–1047.
52. Dufresne, A., Garczarek, L., and Partensky, F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol. *6*, R14.
53. Button, D.K. (1991). Biochemical basis for whole-cell uptake kinetics: specific affinity, oligotrophic capacity, and the meaning of the Michaelis constant. Appl. Environ. Microbiol. *57*, 2033–2038.
54. Camacho, C. *et al.* (2009). Madden, BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.
55. Kinjo, Y., Saitoh, S., and Tokuda, G. (2015). An efficient strategy developed for next-generation sequencing of endosymbiont genomes performed using crude DNA isolated from host tissues: a case study of *Blattabacterium cuenoti* inhabiting the fat bodies of cockroaches. Microbes Environ. *30*, 208–220.
56. Nadalin, F., Vezzi, F., and Policriti, A. (2012). GapFiller: A *de novo* assembly approach to fill the gap within paired reads. BMC Bioinformatics. *13*, S8.

57. Walker, B.J. *et al.* (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS One. *9*, e112963.
58. Hyatt, D. *et al.* (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. *11*, 119.
59. Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. *43*, D261–D269.
60. Lagesen, K. *et al.* (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108.
61. Lowe, T.M., and Eddy, S.R. (1997). TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 9550964.
62. Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. *29*, 2933–2935.
63. Lechner, M. *et al.* (2011). Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics. *12*, 124.
64. Meyer, F., Overbeek, R., and Rodriguez, A. (2009). FIGfams: yet another set of protein families. Nucleic Acids Res. *37*, 6643–6654.
65. Marchler-Bauer, A., and Bryant, S. (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Res. *32*, W327–W331.
66. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
67. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, 609–612.
68. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
69. Minh, B.Q., Nguyen, M.A.T., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. *30*, 1188–1195.

70. Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214.
71. Wu, M., and Scott, A.J. (2012). Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics. *28*, 1033–1034.
72. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. *25*, 1972–1973.
73. Drummond, A.J., Ho, S.Y.W., Phillips, M.J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. PLOS Biol. *4*, e88.
74. Rambaut, A., and Drummond, A.J. Tracer (2007), (available at http://www.beast.bio.ed.ac.uk/Tracer).
75. Gernhard, T. (2008). The conditioned reconstructed process. J. Theor. Biol. *253*, 769–778.
76. Ho, S.Y.W., and Phillips, M.J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. *58*, 367–380.
77. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
78. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591 (2007).
79. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. Bioinformatics. *20*, 289–290.
80. Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. London B. *255*, 37–45.
81. Revell, L.J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223.
82. Orme, D., Freckleton, R., Thomas, G., and Petzoldt, T. (2013). The caper package: comparative analysis of phylogenetics and evolution in R. *R Package*.
83. Kim, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. Commun. Stat. Appl. Methods *22*, 665–674.

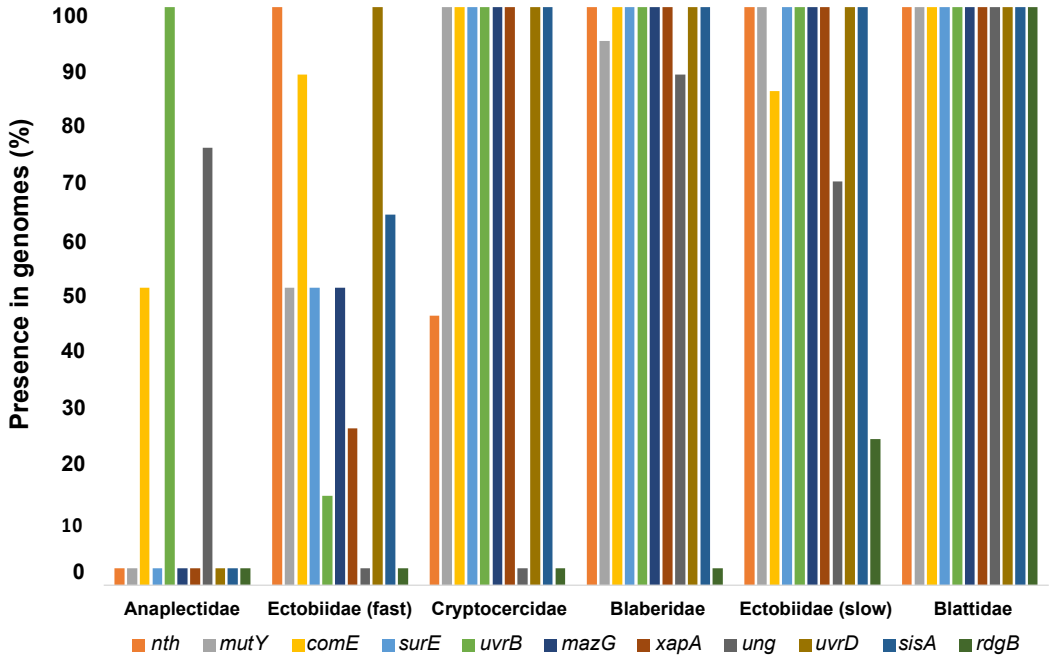


Figure S1. Loss of genes associated with DNA repair across six clades of *Blattabacterium* species. Related to Figure 1.

'Fast' ectobiid lineages are *Blattabacterium* from the host taxa *Allacta* spp., *Balta* sp., *Mediastinia delicatula, Euphyllodromia* sp., *Shelfordina* sp. *Chorisoserrata* sp., and *Amazonina* sp. 'Slow' ectobiid lineages are *Blattabacterium* strains from the remaining ectobiid hosts (see Figure 1) (see Table S1 for further details).

Gene	Putative protein	Comments	Gene presence in different Blattabacterium lineages:					
			Anaplectidae (2 genomes)	**'Fast-evolving' Ectobiidae** (8 genomes)	**'slow-evolving' Ectobiidae** (13 genomes)	**Cryptocercidae** (11 genomes)	**Blaberidae** (17 genomes)	**Blattidae** (10 genomes)
BPLAN_RS00265	Nth, Endonuclease III	excision of bases that have oxidative damage	0	100	100	45	100	100
BPLAN_RS00490	MutY, DNA glycosylase	repairs G-A mispairs and 8-oxo-GTP lesions	0	50	100	100	94	100
BPLAN_RS00675	ComE, dCMP deaminase	hydrolyses dCMP into dUTP	50	88	85	100	100	100
BPLAN_RS00750	SurE, phosphatase	dephosphorylates various ribo- and deoxyribo-nucleoside monophosphates	0	50	100	100	100	100
BPLAN_RS00965	UvrB, DNA helicase	nucleotide excision repair e.g. after UV damage	100	13	100	100	100	100
BPLAN_RS01190	MazG, nucleotide pyrophosphohydrolase	house-cleaning of non-canonical NTPs	0	50	100	100	100	100
BPLAN_RS01330	XapA, purine nucleoside phosphorylase	phosphorolysis of xanthosine, inosine and guanosine	0	25	100	100	100	100
BPLAN_RS02465	Ung, uracil DNA glycosylase	Removal of uracil from DNA; initiates base excision repair pathway	75	0	69	0	88	100
BPLAN_RS02510	UvrD, helicase	repair of DNA damage caused by UV radiation or other causes	0	100	100	100	100	100
BPLAN_RS02695	DisA, DNA integrity scanning protein	scans genome for lesions in DNA	0	63	100	100	100	100
BPLAN_RS03040	RdgB, Inosine/xanthosine triphosphate pyrophosphatase	protects against mutagenesis by 6-N-hydroxylaminopurine	0	0	23	0	0	100

Table S1. List of genes potentially involved in enhanced mutation rates in *Blattabacterium*. Related to Figure 1.

Only genes that are differentially present across the 67 *Blattabacterium* genomes are considered here (genes present in all genomes are not considered). Gene numbers reference to strain BPLAN genome, Genbank NC_013418. Gene presence is given as percentage of genomes in each group that contain the particular gene. The 'fast-evolving' Ectobiidae (>50 genes lost) include *Allacta*, *Amazonina*, *Shelfordina*, *Balta*, *Mediastinia*, *Chorioserata*, *Euphyllodromica* spp., and the 'slow-evolving' Ectobiidae (<50 genes lost) include *Dyakinodes, Paratemnopteryx, Parcoblatta, Blatella, Carbrunneria, Escala, Beybienkoa, Nyctibora, Megaloblatta, Anallacta, Ectobius, Phyllodromica, Ectoneura* spp.

Species	Family	Sample ID	Accession number	Collecting locality	Collector	Date	First run	Second run	Third run
Allacta bimaculata	Ectobiidae	Allacta	SAMN15428893	Next to G213 National Road, Menglun, Yunnan Province, China	Zong Qing Wang	25-Apr-14	Assembled	-	-
Anaplecta omei	Anaplectidae	Anaplecta omei	SAMN15428897	Mt. Emeishan, Sichuan Province, China	Zong Qing Wang	01-Jul-13	Fragmented assembly	Fragmented assembly	Assembled
Allacta australiensis	Ectobiidae	AUS Allacta	SAMN15428892	James Cook University, Rainforest site, Queensland, Australia	David Rentz	22-Jun-15	Fragmented assembly	Fragmented assembly	Assembled
Methana sp.	Blattidae	AUS1	SAMN15428920	North Manly, New South Wales, Australia	Nathan Lo	01-Aug-15	Assembled	-	-
Platyzosteria sp.	Blattidae	AUS3	SAMN15428929	Olney State Forest, New South, Wales, Australia	Nathan Lo and Thomas Bourguignon	25-Aug-15	Fragmented assembly	-	Assembled
Gromphadorhina grandidieri	Blaberidae	B030	SAMN15428913	Breeding of Kyle Kandilian	N/A	N/A	Assembled	-	-
Escala vestjensi	Ectobiidae	B053	SAMN15428910	Breeding of Kyle Kandilian	N/A	N/A	-	Fragmented assembly	Assembled
Schultesia lampyridiformis	Blaberidae	B055	SAMN15428933	Breeding of Kyle Kandilian	N/A	N/A	Fragmented assembly	-	Assembled
Anallacta methanoides	Ectobiidae	B057	SAMN15428895	Breeding of Kyle Kandilian	N/A	N/A	Fragmented assembly	-	Assembled
Paratemnopteryx	Ectobiidae	B061	SAMN15428926	Breeding of Kyle Kandilian	N/A	N/A	Fragmented assembly	-	Assembled
Deropeltis paulinoi	Blattidae	B069	SAMN15428904	Breeding of Kyle Kandilian	N/A	N/A	Assembled	-	-
Shelfordella lateralis	Blattidae	B080	SAMN15428934	Breeding of Kyle Kandilian	N/A	N/A	Fragmented assembly	-	Assembled
Eupolyphaga sinensis	Corydiidae	B081	SAMN15428912	Breeding of Kyle Kandilian	N/A	N/A	Assembled	-	-
Therea regularis	Corydiidae	B091	SAMN15428936	Palm plantation between Puducherry and Auroville, India	Kyle Kandilian	N/A	-	-	Assembled
Macropanesthia rhinoceros	Blaberidae	B092	SAMN15428916	Breeding of Kyle Kandilian	N/A	N/A	-	Fragmented assembly	Assembled
Epilampra maya	Blaberidae	B095	SAMN15428909	Arcadia, Florida, USA	Kyle Kandilian	07-Jul-09	Assembled	-	-

Parcoblatta virginica	Ectobiidae	B102	SAMN15428927	Breeding of Kyle Kandilian	N/A	N/A	Assembled	-	-
Neolaxta mackerrasae	Blaberidae	B107	SAMN15428921	Paluma Range, Australia		15-Oct-15	-	Fragmented assembly	Assembled
Cosmozosteria sp.	Blattidae	B117	SAMN15428902	Cape Upstart, Australia	James Walker	13-Oct-15	-	Fragmented assembly	Assembled
Balta sp.	Ectobiidae	Balta_sp.	SAMN15428898	Cairns, Australia	David Rentz	18-Dec-15	-	Fragmented assembly	Assembled
Beybienkoa kurandanensis	Ectobiidae	Beybienkoa_karandanensis	SAMN15428899	Cairns, Australia	David Rentz	18-Dec-15	-	Fragmented assembly	Assembled
Blattella germanica	Ectobiidae	BGE	CP001487	GenBank - [S1]	N/A	N/A	N/A	N/A	N/A
Blaberus giganteus	Blaberidae	BGIGA	CP003535-CP003536	GenBank - [S2]	N/A	N/A	N/A	N/A	N/A
Nauphoeta cinerea	Blaberidae	BNCIN	CP005488-CP005489	GenBank - [S3]	N/A	N/A	N/A	N/A	N/A
Blatta orientalis	Blattidae	BOR	CP003605-CP003606	GenBank - [S4]	N/A	N/A	N/A	N/A	N/A
Panesthia angustipennis spadica	Blaberidae	BPAA	NC_020510.1	GenBank - [S5]	N/A	N/A	N/A	N/A	N/A
Panesthia angustipennis yaeyamensis	Blaberidae	BPAY	NZ_AP014609.1	GenBank - [S5]	N/A	N/A	N/A	N/A	N/A
Periplaneta americana	Blattidae	BPLAN	CP001429-CP001430	GenBank - [S6]	N/A	N/A	N/A	N/A	N/A
Carbrunneria paramaxi	Ectobiidae	Carbru	SAMN15428900	James Cook University, Rainforest site, Queensland, Australia	David Rentz	05-Oct-15	Fragmented assembly	-	Assembled
Cryptocercus clevelandi	Cryptocercidae	CCLhc	CP029844-CP029845	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Chorisoserrata sp	Ectobiidae	CHORI	SAMN15428901	China	Zong Qing Wang	N/A	Fragmented assembly	-	Assembled
Cryptocercus kyebangensis	Cryptocercidae	CKYod	CP029820-CP029821	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Anaplecta calosoma	Anaplectidae	Cockroach contig 1688	SAMN15428896	Kuranda, Queensland, Australia	David Rentz	17-Nov-15	-	-	Assembled
Protagonista lugubris	Blattidae	Cockroach contig 4907	SAMN15428931	Mt. Diaoluoshan, Hainan, China	Zong Qing Wang	25-May-15	-	-	Assembled

Cryptocercus punctulatus	Cryptocercidae	CPU	CP003015-CP003016	GenBank - [S8]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUbr	CP029816-CP029817	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUbt	CP029813	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUmc	CP029815	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUml	AP014610	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUmp	CP029814	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUpc	CP029811	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUsm	CP029810	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUsv	CP029812	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Cryptocercus punctulatus	Cryptocercidae	CPUwf	CP029818-CP029819	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Dyakinodes kurandensis	Ectobiidae	Dyakinodes	SAMN15428906	Queensland, Australia	David Rentz	18-Dec-15	-	Fragmented assembly	Assembled
Megaloblatta sp.	Ectobiidae	ECMD1	SAMN15428918	Ecuador, *Podocarpus* National Park	Frantisek Juna	Apr-2016	-	Assembled	-
Ectoneura hanitschi	Ectobiidae	Ectoneura_hanitschi	SAMN15428908	Queensland, Australia	David Rentz	18-Dec-15	-	Fragmented assembly	Assembled
Lamproblatta sp.	Lamproblattidae	LA male	SAMN15428915	Petit Saut, French Guiana	Frantisek Juna	08-Jul-09	-	-	Assembled
Mastotermes darwiniensis	Isoptera	MADAR	CP003000, CP003095	GenBank [S9]	N/A	N/A	N/A	N/A	N/A
Melanozosteria sp.	Blattidae	Melanozosteria_sp.	SAMN15428919	Cairns, Australia	David Rentz	18-Dec-15	-	Assembled	-
Paranauphoeta circumdata	Blaberidae	PARA	SAMN15428925	N/A	N/A	N/A	Assembled	-	-
Phyllodromica sp.	Ectobiidae	Phil	SAMN15428928	Czech Republic		01-Aug-15	Fragmented assembly	-	Assembled
Polyphagoides sp.	Corydiidae	POLY	SAMN15428930	Cairns, Australia	David Rentz	18-Dec-15	-	Fragmented assembly	Assembled
Rhabdoblatta sp.	Blaberidae	RHA	SAMN15428932	Kuranda, Queensland, Australia	David Rentz	16-Sep-15	Assembled	-	-

Panesthia sp.	Blaberidae	Salganea	SAMN15428924	Bubeng, Yunnan province, China	N/A	08-Jul-09	Assembled	-	-
Salganea taiwanensis	Blaberidae	STAT	AP014608	GenBank - [S7]	N/A	N/A	N/A	N/A	N/A
Tryonicus parvus	Tryonicidae	Tryonicus parvus	SAMN15428937	Olney State Forest, New South, Wales, Australia	NL and TB	10-Mar-16	-	-	Assembled
Cyrtotria	Blaberidae	Z137	SAMN15428903		Frantisek Juna	N/A	Fragmented assembly	-	Assembled
Gyna capucina	Blaberidae	Z139GY	SAMN15428914	Ebogo, Cameroon	Frantisek Juna	08-Sep-15	Fragmented assembly	-	Assembled
Opisthoplatia orientalis	Blaberidae	Z15100	SAMN15428923	breeds J. Hromádka	N/A	N/A	Fragmented assembly	-	Assembled
Ectobius sp.	Ectobiidae	Z254C	SAMN15428907	Slovenia	Frantisek Juna	Apr-2016	-	Assembled	-
Amazonina sp.	Ectobiidae	Z256E	SAMN15428894	Ecuador, Bosque Protector del Alto Nangaritza	Frantisek Juna	Apr-2016	-	Assembled	-
Euphyllodromia sp.	Ectobiidae	Z257	SAMN15428911	Ecuador, Podocarpus National Park	Frantisek Juna	Apr-2016	-	Fragmented assembly	Assembled
Nyctibora	Ectobiidae	Z258E	SAMN15428922	Ecuador	Frantisek Juna	Apr-2016	-	-	Assembled
Duchailluia sp.	Blattidae	Z299	SAMN15428905	Cameroon	Frantisek Juna	Dec-16	-	-	Assembled
Mediastinia delicatula	Ectobiidae		SAMN15428917	Queensland, Australia	David Rentz	2015	-	Fragmented assembly	Assembled
Shelfordina sp.	Ectobiidae		SAMN15428935	Queensland, Australia	David Rentz	2015	-	Fragmented assembly	Assembled

Table S2. List of cockroach samples used in this study. Related to STAR Methods and Figure 1.

Group	*Species*	Age (Ma) / min. age constraint for group	Calibration group	Soft max. bound (97.5% probability)	Reference	Comments on soft max. bound
Blattabacterium	*Baissatermes lapideus*	137	*Cryptocercus* + Isoptera	235	[S10]	*Triassoblatta argentina, earliest fossil of Mesoblattinidae [S19]*
	Balatronis libanensis	125	Blattidae + Tryonicidae	235	[S11]	*Triassoblatta argentina, earliest fossil of Mesoblattinidae [S19]*
	Pseudoplecta krassilovi	89.3	Anaplectidae + Lamproblatidae	235	[S12]	*Triassoblatta argentina, earliest fossil of Mesoblattinidae [S19]*
	Blattella lengleti	94.3	Blaberidae + sister group	235	[S13]	*Triassoblatta argentina, earliest fossil of Mesoblattinidae [S19]*
	Periplaneta houlberti	56	*Periplaneta* + *Shelfordella* + *Blatta* + *Neostylopyga* + *Deropeltis*	145	[S14]	First mordern cockroach: Zhujiblatta [S20]
	Gyna obesa	56	Gyninae + Panchlorinae + Blaberinae	145	[S14]	First mordern cockroach: Zhujiblatta [S20]
	Ectobius kohlsi	46.2	*Philodromica* + *Ectobius* + *Ectoneura*	145	[S15]	First mordern cockroach: Zhujiblatta [S20]
Buchnera	NA	50	APS - Sg	70	[S16]	NA
	NA	80	Root of the tree	100	[S17]	NA
Prochloraceae	NA	NA	Root of the tree	1,500	[S18]	NA

Table S3. List of fossils used in this study to calibrate the timetrees of *Blattabacterium*, *Buchnera*, and *Prochlorococcus+Synechococcus.* Related to STAR Methods.

Organism name	Lineage	Phylum	Assembly level	Genome size	GC content	RefSeq assembly accession
Blattabacterium spp. str. BPAA	Blattabacterium	Bacteroidetes	Complete	0.632	26.4	GCF_000348805.1
Blattabacterium spp. str. BPAY	Blattabacterium	Bacteroidetes	Complete	0.632	26.3	GCF_002355135.1
Blattabacterium spp. str. STAT	Blattabacterium	Bacteroidetes	Complete	0.632	24.8	GCF_003573915.1
Blattabacterium spp. str. CKYod	Blattabacterium	Bacteroidetes	Complete	0.637	25.7	GCF_003226855.1
Blattabacterium spp. str. CCLhc	Blattabacterium	Bacteroidetes	Complete	0.621	24.5	GCF_003268615.1
Blattabacterium spp. str. CPUpc	Blattabacterium	Bacteroidetes	Complete	0.614	23.8	GCF_003226715.1
Blattabacterium spp. str. CPUsv	Blattabacterium	Bacteroidetes	Complete	0.614	23.8	GCF_003226775.1
Blattabacterium spp. str. CPUsm	Blattabacterium	Bacteroidetes	Complete	0.614	23.8	GCF_003226755.1
Blattabacterium spp. str. CPUmc	Blattabacterium	Bacteroidetes	Complete	0.613	23.9	GCF_003226815.1
Blattabacterium spp. str. CPUbt	Blattabacterium	Bacteroidetes	Complete	0.613	23.9	GCF_003226795.1
Blattabacterium spp. str. CPUmp	Blattabacterium	Bacteroidetes	Complete	0.613	23.8	GCF_003226835.1
Blattabacterium spp. str. CPUml	Blattabacterium	Bacteroidetes	Complete	0.616	24.1	GCF_003226695.1
Blattabacterium spp. str. CPUwf	Blattabacterium	Bacteroidetes	Complete	0.611	23.8	GCF_003226875.1
Blattabacterium spp. str. CPUbr	Blattabacterium	Bacteroidetes	Complete	0.609	23.8	GCF_003226735.1
Blattabacterium spp. str. Cpu	Blattabacterium	Bacteroidetes	Complete	0.610	23.8	GCF_000236405.1
Blattabacterium spp. str. MADAR	Blattabacterium	Bacteroidetes	Complete	0.590	27.5	GCF_000233435.1
Blattabacterium spp. str. BGIGA	Blattabacterium	Bacteroidetes	Complete	0.633	25.7	GCF_000262715.1
Blattabacterium spp. str. BOR	Blattabacterium	Bacteroidetes	Complete	0.638	28.2	GCF_000334405.1
Blattabacterium spp. str. BGE	Blattabacterium	Bacteroidetes	Complete	0.641	27.1	GCF_000022605.2
Blattabacterium spp. str. BNCIN	Blattabacterium	Bacteroidetes	Complete	0.627	26.1	GCF_000471965.1
Blattabacterium spp. str. BPLAN	Blattabacterium	Bacteroidetes	Complete	0.640	28.2	GCF_000093165.1
Buchnera aphidicola str. Aar	Buchnera	Proteobacteria	Chromosome	0.641	24.5	GCF_005082365.1
Buchnera aphidicola str. Acr	Buchnera	Proteobacteria	Complete	0.642	24.4	GCF_005082145.1
Buchnera aphidicola str. Ahe	Buchnera	Proteobacteria	Complete	0.645	24.1	GCF_005083845.1

Buchnera aphidicola str. Ak	Buchnera	Proteobacteria	Complete	0.653	25.7	GCF_000225445.1
Buchnera aphidicola str. Ana	Buchnera	Proteobacteria	Chromosome	0.641	24.8	GCF_005083345.1
Buchnera aphidicola str. Ane	Buchnera	Proteobacteria	Chromosome	0.642	24.2	GCF_005083105.1
Buchnera aphidicola str. Aoe	Buchnera	Proteobacteria	Chromosome	0.549	22.7	GCF_005080765.1
Buchnera aphidicola str. APS	Buchnera	Proteobacteria	Complete	0.656	26.4	GCF_000009605.1
Buchnera aphidicola str. BAg	Buchnera	Proteobacteria	Complete	0.639	25.6	GCF_001280225.1
Buchnera aphidicola str. Bbr	Buchnera	Proteobacteria	Chromosome	0.657	25.0	GCF_005082825.1
Buchnera aphidicola str. Bca	Buchnera	Proteobacteria	Chromosome	0.655	25.3	GCF_005081945.1
Buchnera aphidicola str. BCc	Buchnera	Proteobacteria	Complete	0.422	20.2	GCF_000090965.1
Buchnera aphidicola str. Bciconfinis	Buchnera	Proteobacteria	Complete	0.452	24.1	GCF_900128735.1
Buchnera aphidicola str. Bcifornacula	Buchnera	Proteobacteria	Complete	0.457	28.3	GCF_900128725.1
Buchnera aphidicola str. Bcipseudotaxifoliae	Buchnera	Proteobacteria	Complete	0.455	24.2	GCF_900128595.1
Buchnera aphidicola str. Bp	Buchnera	Proteobacteria	Complete	0.618	25.3	GCF_000007725.1
Buchnera aphidicola str. BTI	Buchnera	Proteobacteria	Complete	0.643	25.3	GCF_003671935.1
Buchnera aphidicola str. BTs	Buchnera	Proteobacteria	Complete	0.430	21.6	GCF_900016785.1
Buchnera aphidicola str. BuCicuneomaculata	Buchnera	Proteobacteria	Chromosome	0.443	23.8	GCF_900698865.1
Buchnera aphidicola str. BuCicurtihirsuta	Buchnera	Proteobacteria	Chromosome	0.442	21.3	GCF_900698895.1
Buchnera aphidicola str. BuCicurvipes	Buchnera	Proteobacteria	Chromosome	0.445	21.3	GCF_900698915.1
Buchnera aphidicola str. BuCikochiana	Buchnera	Proteobacteria	Chromosome	0.442	23.4	GCF_900698905.1
Buchnera aphidicola str. BuCilaricifoliae	Buchnera	Proteobacteria	Chromosome	0.446	22.4	GCF_900698945.1
Buchnera aphidicola str. BuCipiceae	Buchnera	Proteobacteria	Chromosome	0.444	22.0	GCF_900699035.1
Buchnera aphidicola str. BuCisplendens	Buchnera	Proteobacteria	Chromosome	0.453	24.0	GCF_900698845.1
Buchnera aphidicola str. BuCistrobi	Buchnera	Proteobacteria	Chromosome	0.449	24.0	GCF_900560745.1
Buchnera aphidicola str. Hla	Buchnera	Proteobacteria	Chromosome	0.653	26.1	GCF_005081705.1
Buchnera aphidicola str. Hta	Buchnera	Proteobacteria	Complete	0.645	27.0	GCF_005081445.1
Buchnera aphidicola str. Lps	Buchnera	Proteobacteria	Chromosome	0.652	25.0	GCF_005081185.1
Buchnera aphidicola str. LSU	Buchnera	Proteobacteria	Complete	0.637	25.3	GCF_003096055.1

Buchnera aphidicola str. Meu	Buchnera	Proteobacteria	Complete	0.645	25.6	GCF_005237295.1
Buchnera aphidicola str. Mga	Buchnera	Proteobacteria	Complete	0.654	25.9	GCF_005080965.1
Buchnera aphidicola str. Mrh	Buchnera	Proteobacteria	Chromosome	0.616	25.7	GCF_005080745.1
Buchnera aphidicola str. Msa	Buchnera	Proteobacteria	Chromosome	0.633	24.4	GCF_005080885.1
Buchnera aphidicola str. Mst	Buchnera	Proteobacteria	Complete	0.631	25.7	GCF_005080865.1
Buchnera aphidicola str. Nmo	Buchnera	Proteobacteria	Complete	0.600	22.3	GCF_006741185.1
Buchnera aphidicola str. Rpa	Buchnera	Proteobacteria	Chromosome	0.654	25.2	GCF_005080845.1
Buchnera aphidicola str. SAM	Buchnera	Proteobacteria	Complete	0.636	25.5	GCF_001700895.1
Buchnera aphidicola str. Sav	Buchnera	Proteobacteria	Complete	0.648	26.0	GCF_005082585.1
Buchnera aphidicola str. SC	Buchnera	Proteobacteria	Complete	0.608	25.8	GCF_001648115.1
Buchnera aphidicola str. Sg	Buchnera	Proteobacteria	Complete	0.641	25.3	GCF_000007365.1
Buchnera aphidicola str. Ska	Buchnera	Proteobacteria	Chromosome	0.428	24.8	GCF_005080725.1
Buchnera aphidicola str. Ssp	Buchnera	Proteobacteria	Chromosome	0.412	23.0	GCF_005080785.1
Buchnera aphidicola str. Tca	Buchnera	Proteobacteria	Chromosome	0.534	22.6	GCF_005080825.1
Buchnera aphidicola str. Tma	Buchnera	Proteobacteria	Complete	0.419	20.2	GCF_005080705.1
Buchnera aphidicola str. Ua	Buchnera	Proteobacteria	Complete	0.628	24.1	GCF_000225465.1
Prochlorococcus marinus subsp. marinus str. CCMP1375	Synechococcales	Cyanobacteria	Complete	1.75	36.4	GCF_000007925.1
Prochlorococcus marinus subsp. pastoris str. CCMP1986	Synechococcales	Cyanobacteria	Complete	1.66	30.8	GCF_000011465.1
Prochlorococcus sp. MIT 0604	Synechococcales	Cyanobacteria	Complete	1.78	31.2	GCF_000757845.1
Prochlorococcus sp. MIT 0801	Synechococcales	Cyanobacteria	Complete	1.93	34.9	GCF_000757865.1
Prochlorococcus marinus str. MIT 9211	Synechococcales	Cyanobacteria	Chromosome	1.69	38.0	GCF_000018585.1
Prochlorococcus marinus str. MIT 9215	Synechococcales	Cyanobacteria	Complete	1.74	31.1	GCF_000018065.1
Prochlorococcus marinus str. MIT 9301	Synechococcales	Cyanobacteria	Complete	1.64	31.3	GCF_000015965.1
Prochlorococcus marinus str. MIT 9303	Synechococcales	Cyanobacteria	Complete	2.68	50.0	GCF_000015705.1
Prochlorococcus marinus str. MIT 9312	Synechococcales	Cyanobacteria	Complete	1.71	31.2	GCF_000012645.1
Prochlorococcus marinus str. MIT 9313	Synechococcales	Cyanobacteria	Complete	2.41	50.7	GCF_000011485.1

Prochlorococcus marinus str. MIT 9515	Synechococcales	Cyanobacteria	Complete	1.70	30.8	GCF_000015665.1
Prochlorococcus marinus str. NATL1A	Synechococcales	Cyanobacteria	Complete	1.86	35.0	GCF_000015685.1
Prochlorococcus marinus str. NATL2A	Synechococcales	Cyanobacteria	Complete	1.84	35.1	GCF_000012465.1
Prochlorococcus sp. RS01	Synechococcales	Cyanobacteria	Chromosome	1.66	31.4	GCF_001989435.1
Synechococcus sp. CB0101	Synechococcales	Cyanobacteria	Complete	2.79	64.1	GCF_000179235.2
Synechococcus sp. CC9311	Synechococcales	Cyanobacteria	Complete	2.61	52.4	GCF_000014585.1
Synechococcus sp. CC9605	Synechococcales	Cyanobacteria	Complete	2.51	59.2	GCF_000012625.1
Synechococcus sp. CC9902	Synechococcales	Cyanobacteria	Complete	2.23	54.2	GCF_000012505.1
Synechococcus sp. KORDI-100	Synechococcales	Cyanobacteria	Complete	2.79	57.5	GCF_000737535.1
Synechococcus sp. KORDI-49	Synechococcales	Cyanobacteria	Complete	2.59	61.4	GCF_000737575.1
Synechococcus sp. KORDI-52	Synechococcales	Cyanobacteria	Complete	2.57	59.1	GCF_000737595.1
Synechococcus sp. RCC307	Synechococcales	Cyanobacteria	Complete	2.22	60.8	GCF_000063525.1
Synechococcus sp. SynAce01	Synechococcales	Cyanobacteria	Complete	2.75	63.9	GCF_001885215.1
Synechococcus sp. WH 7803	Synechococcales	Cyanobacteria	Complete	2.37	60.2	GCF_000063505.1
Synechococcus sp. WH 8020	Synechococcales	Cyanobacteria	Chromosome	2.66	53.1	GCF_001040845.1
Synechococcus sp. WH 8101	Synechococcales	Cyanobacteria	Complete	2.63	63.3	GCF_004209775.1
Synechococcus sp. WH 8102	Synechococcales	Cyanobacteria	Complete	2.43	59.4	GCF_000195975.1
Synechococcus sp. WH 8109	Synechococcales	Cyanobacteria	Complete	2.11	60.1	GCF_000161795.2
Thermococcus sp. P6	Thermococcus	Euryarchaeota	Complete	1.52	54.9	GCF_002214525.1
Thermococcus gorgonarius	Thermococcus	Euryarchaeota	Complete	1.67	51.7	GCF_002214385.1
Thermococcus pacificus	Thermococcus	Euryarchaeota	Complete	1.79	54.2	GCF_002214485.1
Thermococcus onnurineus NA1	Thermococcus	Euryarchaeota	Complete	1.85	51.3	GCF_000018365.1
Thermococcus sp. 5-4	Thermococcus	Euryarchaeota	Complete	1.85	55.7	GCF_002197185.1
Thermococcus celer Vu 13 = JCM 8558	Thermococcus	Euryarchaeota	Complete	1.87	56.4	GCF_002214365.1
Thermococcus radiotolerans	Thermococcus	Euryarchaeota	Complete	1.87	55.6	GCF_002214565.1
Thermococcus guaymasensis DSM 11113	Thermococcus	Euryarchaeota	Complete	1.92	52.9	GCF_000816105.1
Thermococcus barossii	Thermococcus	Euryarchaeota	Complete	1.92	54.7	GCF_002214465.1

Thermococcus piezophilus	Thermococcus	Euryarchaeota	Complete	1.93	51.1	GCF_001647085.1
Thermococcus cleftensis	Thermococcus	Euryarchaeota	Complete	1.95	55.8	GCF_000265525.1
Thermococcus nautili	Thermococcus	Euryarchaeota	Complete	1.98	54.8	GCF_000585495.1
Thermococcus sp. 4557	Thermococcus	Euryarchaeota	Complete	2.01	56.1	GCF_000221185.1
Thermococcus siculi	Thermococcus	Euryarchaeota	Complete	2.03	55	GCF_002214505.1
Thermococcus gammatolerans EJ3	Thermococcus	Euryarchaeota	Complete	2.05	53.6	GCF_000022365.1
Thermococcus thioreducens	Thermococcus	Euryarchaeota	Complete	2.07	53.5	GCF_002214545.1
Thermococcus sp. AM4	Thermococcus	Euryarchaeota	Complete	2.09	54.8	GCF_000151205.2
Thermococcus kodakarensis KOD1	Thermococcus	Euryarchaeota	Complete	2.09	52	GCF_000009965.1
Thermococcus sp. EXT12c	Thermococcus	Euryarchaeota	Complete	2.16	54.6	GCF_900198835.1
Actinobacillus succinogenes 130Z	Gammaproteobacteria	Proteobacteria	Complete	2.32	44.9	GCF_000017245.1
Actinobacillus suis ATCC 33415	Gammaproteobacteria	Proteobacteria	Complete	2.50	40.2	GCF_000739435.1
Aggregatibacter actinomycetemcomitans	Gammaproteobacteria	Proteobacteria	Complete	2.37	44.2	GCF_001594265.1
Bibersteinia trehalosi USDA-ARS-USMARC-192	Gammaproteobacteria	Proteobacteria	Complete	2.41	41	GCF_000347595.1
Escherichia coli str. K-12 substr. MG1655	Gammaproteobacteria	Proteobacteria	Complete	4.64	50.8	GCF_000005845.2
Frischella perrara	Gammaproteobacteria	Proteobacteria	Complete	2.69	34.1	GCF_000807275.1
Gallibacterium anatis UMN179	Gammaproteobacteria	Proteobacteria	Complete	2.69	39.9	GCF_000209675.1
Gilliamella apicola	Gammaproteobacteria	Proteobacteria	Complete	3.14	33.6	GCF_000599985.1
Haemophilus ducreyi 35000HP	Gammaproteobacteria	Proteobacteria	Complete	1.70	38.2	GCF_000007945.1
Haemophilus influenzae Rd KW20	Gammaproteobacteria	Proteobacteria	Complete	1.83	38.2	GCF_000027305.1
Histophilus somni 2336	Gammaproteobacteria	Proteobacteria	Complete	2.26	37.4	GCF_000019405.1
Mannheimia haemolytica M42548	Gammaproteobacteria	Proteobacteria	Complete	2.73	41.0	GCF_000376645.1
Morganella morganii subsp. morganii KT	Gammaproteobacteria	Proteobacteria	Complete	3.80	51.1	GCF_000286435.2
Obesumbacterium proteus	Gammaproteobacteria	Proteobacteria	Complete	5.01	49.1	GCF_001586165.1
Pectobacterium carotovorum subsp. carotovorum PC1	Gammaproteobacteria	Proteobacteria	Complete	4.86	51.9	GCF_000023605.1
Plautia stali symbiont	Gammaproteobacteria	Proteobacteria	Complete	4.09	56.9	GCF_000180175.2

Pragia fontium	Gammaproteobacteria	Proteobacteria	Complete	4.09	45.4	GCF_001026985.1
Proteus mirabilis HI4320	Gammaproteobacteria	Proteobacteria	Complete	4.10	38.9	GCF_000069965.1
Serratia fonticola	Gammaproteobacteria	Proteobacteria	Complete	6.00	53.6	GCF_001006005.1
Xenorhabdus hominickii	Gammaproteobacteria	Proteobacteria	Complete	4.52	43.4	GCF_001721185.1
Corynebacterium aquilae DSM 44791	Corynebacterium	Actinobacteria	Complete	2.93	60.9	GCF_001941445.1
Corynebacterium callunae DSM 20147	Corynebacterium	Actinobacteria	Complete	2.93	52.5	GCF_000344785.1
Corynebacterium casei LMG S-19264	Corynebacterium	Actinobacteria	Complete	3.13	55.7	GCF_000550785.1
Corynebacterium efficiens YS-314	Corynebacterium	Actinobacteria	Complete	3.22	62.9	GCF_000011305.1
Corynebacterium falsenii DSM 44353	Corynebacterium	Actinobacteria	Complete	2.72	63.2	GCF_000525655.1
Corynebacterium flavescens	Corynebacterium	Actinobacteria	Complete	2.76	59.9	GCF_001941465.1
Corynebacterium glyciniphilum AJ 3170	Corynebacterium	Actinobacteria	Complete	3.57	64.7	GCF_000626675.1
Corynebacterium halotolerans YIM 70093 = DSM 44683	Corynebacterium	Actinobacteria	Complete	3.22	68.3	GCF_000341345.1
Corynebacterium jeikeium K411	Corynebacterium	Actinobacteria	Complete	2.48	61.4	GCF_000006605.1
Corynebacterium kroppenstedtii DSM 44385	Corynebacterium	Actinobacteria	Complete	2.45	57.5	GCF_000023145.1
Corynebacterium lactis RW2-5	Corynebacterium	Actinobacteria	Complete	2.77	60.5	GCF_001274895.1
Corynebacterium marinum DSM 44953	Corynebacterium	Actinobacteria	Complete	2.73	67.8	GCF_000835165.1
Corynebacterium resistens DSM 45100	Corynebacterium	Actinobacteria	Complete	2.60	57.1	GCF_000177535.2
Corynebacterium simulans	Corynebacterium	Actinobacteria	Complete	2.74	59.0	GCF_001586215.1
Corynebacterium sphenisci DSM 44792	Corynebacterium	Actinobacteria	Complete	2.59	74.7	GCF_001941505.1
Corynebacterium terpenotabidum Y-11	Corynebacterium	Actinobacteria	Complete	2.75	67.0	GCF_000418365.1
Corynebacterium testudinoris	Corynebacterium	Actinobacteria	Complete	2.72	63.1	GCF_001021045.1
Corynebacterium vitaeruminis DSM 20294	Corynebacterium	Actinobacteria	Complete	2.93	65.5	GCF_000550805.1
Mycobacterium avium subsp. paratuberculosis K-10	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	4.83	69.3	GCF_000007865.1
Mycobacterium leprae TN	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	3.27	57.8	GCF_000195855.1
Mycobacteroides chelonae CCUG 47445	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	5.03	63.9	GCF_001632805.1

Mycolicibacter sinensis	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	4.64	68.4	GCF_000214155.1
Mycolicibacterium neoaurum VKM Ac-1815D	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	5.42	66.9	GCF_000317305.3
Mycolicibacterium phlei	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	5.35	69.4	GCF_001583415.1
Mycolicibacterium rhodesiae NBB3	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	6.42	65.5	GCF_000230895.3
Mycolicibacterium smegmatis MC2 155	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	6.99	67.4	GCF_000015005.1
Mycolicibacterium vaccae 95051	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	6.24	68.6	GCF_001655245.1
Nocardia cyriacigeorgica GUH-2	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	6.19	68.4	GCF_000284035.1
Nocardia mangyaensis	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	7.31	68.1	GCF_001886715.1
Nocardia nova SH22a	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	8.35	67.8	GCF_000523235.1
Rhodococcus fascians D188	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	5.50	64.6	GCF_001620305.1
Rhodococcus jostii RHA1	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	9.70	67.0	GCF_000014565.1
Rhodococcus pyridinivorans SB3094	Mycobacteriaceae + Nocardiaceae	Actinobacteria	Complete	5.59	67.8	GCF_000511305.1
Agromyces aureus	Micrococcineae	Actinobacteria	Complete	4.37	70.4	GCF_001660485.1
Arthrobacter alpinus	Micrococcineae	Actinobacteria	Complete	4.33	60.6	GCF_001445575.1
Clavibacter michiganensis subsp. sepedonicus	Micrococcineae	Actinobacteria	Complete	3.40	72.4	GCF_000069225.1
Cnuibacter physcomitrellae	Micrococcineae	Actinobacteria	Complete	4.35	70.8	GCF_002096055.1
Cryobacterium arcticum	Micrococcineae	Actinobacteria	Complete	4.35	68.4	GCF_001679725.1
Glutamicibacter halophytocola	Micrococcineae	Actinobacteria	Complete	3.92	60.0	GCF_001302565.1
Kocuria flava	Micrococcineae	Actinobacteria	Complete	3.64	73.9	GCF_001482365.1
Kocuria palustris	Micrococcineae	Actinobacteria	Complete	2.85	70.5	GCF_001275345.1
Kocuria rhizophila DC2201	Micrococcineae	Actinobacteria	Complete	2.70	71.2	GCF_000010285.1
Leifsonia xyli subsp. cynodontis DSM 46306	Micrococcineae	Actinobacteria	Complete	2.69	68.3	GCF_000470775.1
Microbacterium aurum	Micrococcineae	Actinobacteria	Complete	3.42	69.9	GCF_001974985.1

Microbacterium paludicola	Micrococcineae	Actinobacteria	Complete	3.41	70.1	GCF_001887285.1
Microbacterium testaceum StLB037	Micrococcineae	Actinobacteria	Complete	3.98	70.3	GCF_000202635.1
Micrococcus luteus NCTC 2665	Micrococcineae	Actinobacteria	Complete	2.50	73.0	GCF_000023205.1
Microterricola viridarii	Micrococcineae	Actinobacteria	Complete	3.69	68.7	GCF_001542775.1
Neomicrococcus aestuarii	Micrococcineae	Actinobacteria	Complete	2.67	59.1	GCF_001887245.1
Pseudarthrobacter sulfonivorans	Micrococcineae	Actinobacteria	Complete	5.06	64.7	GCF_001484605.1
Rathayibacter toxicus	Micrococcineae	Actinobacteria	Complete	2.35	61.5	GCF_001465855.1
Renibacterium salmoninarum ATCC 33209	Micrococcineae	Actinobacteria	Complete	3.16	56.3	GCF_000018885.1
Rhodoluna lacicola	Micrococcineae	Actinobacteria	Complete	1.43	51.5	GCF_000699505.1
Rothia mucilaginosa DY-18	Micrococcineae	Actinobacteria	Complete	2.26	59.6	GCF_000011025.1
Sinomonas atrocyanea	Micrococcineae	Actinobacteria	Complete	4.49	71.4	GCF_001577305.1
Aequorivita sublithincola DSM 14238	Flavobacteriaceae	Bacteroidetes	Complete	3.52	36.2	GCF_000265385.1
Algibacter alginicilyticus	Flavobacteriaceae	Bacteroidetes	Complete	3.99	31.8	GCF_001310225.1
Arenibacter algicola	Flavobacteriaceae	Bacteroidetes	Complete	5.86	39.8	GCF_002234495.1
Capnocytophaga canimorsus Cc5	Flavobacteriaceae	Bacteroidetes	Complete	2.57	36.1	GCF_000220625.1
Capnocytophaga ochracea DSM 7271	Flavobacteriaceae	Bacteroidetes	Complete	2.61	39.6	GCF_000023285.1
Cellulophaga baltica 18	Flavobacteriaceae	Bacteroidetes	Complete	4.64	34.7	GCF_000468615.2
Cellulophaga lytica DSM 7489	Flavobacteriaceae	Bacteroidetes	Complete	3.77	32.1	GCF_000190595.1
Chryseobacterium indologenes	Flavobacteriaceae	Bacteroidetes	Complete	5.31	35.9	GCF_002025665.1
Croceibacter atlanticus HTCC2559	Flavobacteriaceae	Bacteroidetes	Complete	2.95	33.9	GCF_000196315.1
Dokdonia donghaensis DSW-1	Flavobacteriaceae	Bacteroidetes	Complete	3.29	38.2	GCF_001653755.1
Elizabethkingia anophelis NUHP1	Flavobacteriaceae	Bacteroidetes	Complete	4.37	35.6	GCF_000495935.2
Flavobacterium crassostreae	Flavobacteriaceae	Bacteroidetes	Complete	3.03	36.0	GCF_001831475.1
Flavobacterium indicum GPTSA100-9 = DSM 17447	Flavobacteriaceae	Bacteroidetes	Complete	2.99	31.4	GCF_000455605.1
Flavobacterium johnsoniae UW101	Flavobacteriaceae	Bacteroidetes	Complete	6.10	34.1	GCF_000016645.1
Gramella flava JLT2011	Flavobacteriaceae	Bacteroidetes	Complete	4.01	42.1	GCF_001951155.1

Gramella forsetii KT0803	Flavobacteriaceae	Bacteroidetes	Complete	3.80	36.6	GCF_000060345.1
Lacinutrix venerupis	Flavobacteriaceae	Bacteroidetes	Complete	3.19	30.6	GCF_001971745.1
Lutibacter profundi	Flavobacteriaceae	Bacteroidetes	Complete	2.97	29.8	GCF_001543325.1
Muricauda lutaonensis	Flavobacteriaceae	Bacteroidetes	Complete	3.27	45.0	GCF_000963865.1
Muricauda ruestringensis DSM 13258	Flavobacteriaceae	Bacteroidetes	Complete	3.84	41.4	GCF_000224085.1
Myroides profundi	Flavobacteriaceae	Bacteroidetes	Complete	4.06	33.8	GCF_000833025.1
Nonlabens dokdonensis DSW-6	Flavobacteriaceae	Bacteroidetes	Complete	3.91	35.3	GCF_000332115.1
Nonlabens sediminis	Flavobacteriaceae	Bacteroidetes	Complete	2.84	35.5	GCF_002117085.1
Ornithobacterium rhinotracheale DSM 15997	Flavobacteriaceae	Bacteroidetes	Complete	2.40	37.2	GCF_000265465.1
Polaribacter vadi	Flavobacteriaceae	Bacteroidetes	Complete	3.81	29.6	GCF_001761365.1
Psychroflexus torquis ATCC 700755	Flavobacteriaceae	Bacteroidetes	Complete	4.32	34.5	GCF_000153485.2
Riemerella anatipestifer ATCC 11845 = DSM 15868	Flavobacteriaceae	Bacteroidetes	Complete	2.16	35.0	GCF_000252855.1
Robiginitalea biformata HTCC2501	Flavobacteriaceae	Bacteroidetes	Complete	3.53	55.3	GCF_000024125.1
Siansivirga zeaxanthinifaciens CC-SAMT-1	Flavobacteriaceae	Bacteroidetes	Complete	3.30	33.5	GCF_000941055.1
Tenacibaculum dicentrarchi	Flavobacteriaceae	Bacteroidetes	Complete	2.92	31.5	GCF_001483385.1
Wenyingzhuangia fucanilytica	Flavobacteriaceae	Bacteroidetes	Complete	3.43	31.6	GCF_001697185.1
Zobellia galactanivorans	Flavobacteriaceae	Bacteroidetes	Complete	5.52	42.8	GCF_000973105.1
Zunongwangia profunda SM-A87	Flavobacteriaceae	Bacteroidetes	Complete	5.13	36.2	GCF_000023465.1

Table S4. List of published genomes used in this study. Related to STAR Methods.

SUPPLEMENTAL REFERENCES

S1 López-Sánchez, M.J. *et al.* (2009). Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. PLOS Genet. *5*, e1000721.

S2 Huang, C.Y., Sabree, Z.L., and Moran, N.A. (2012). Genome sequence of *Blattabacterium* sp. strain BGIGA, endosymbiont of the *Blaberus giganteus* cockroach. J. Bacteriol. *194*, 4450–4451.

S3 Kambhampati, S., Alleman, A., and Park, Y. (2013). Complete genome sequence of the endosymbiont *Blattabacterium* from the cockroach *Nauphoeta cinerea* (Blattodea: Blaberidae). Genomics *102*, 479–483.

S4 Patiño-Navarrete, R., Moya, A., Latorre, A., and Peretó, J. (2013). Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. Genome Biol. Evol. *5*, 351–361.

S5 Tokuda, G. *et al.* (2013). Maintenance of essential amino acid synthesis pathways in the *Blattabacterium cuenoti* symbiont of a wood-feeding cockroach. Biol. Lett. *9*, 20121153.

S6 Sabree, Z.L., Kambhampati, S., and Moran, N.A. (2009). Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. Proc. Natl. Acad. Sci. U. S. A. *106*, 19521–19526.

S7 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.

S8 Neef, A. *et al.* (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.

S9 Sabree, Z.L. *et al.* (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.

S10 Engel, M.S., Grimaldi, D.A., and Krishna, K. (2007). Primitive termites from the Early Cretaceous. Stuttgarter Beitrage zur Naturkd. 371, 1–32.

S11 Sendi, H., and Azar, D. (2017). New aposematic and presumably repellent bark cockroach from Lebanese amber. Cretac. Res. *72*, 13–17.

S12 Anisyutkin, L.N., Rasnitsyn, A.P., and Vršanský, P. (2008). Cockroaches and mantises. Orders Blattodea (= Blattida) and Mantodea (= Mantida) descriptions. in *Plant-Arthopod interactions in the early angiosperm history: evidence from the Cretaceous of Israel* (eds. Krassilov, V., and Rasnitsyn, A.), Pensoft Publishers, pp. 199–209.

S13 Vršanský, P. (2008). Mesozoic relative of the common synanthropic German cockroach (Blattodea). Mitteilungen aus dem Museum fur Naturkd. Berlin - Dtsch. Entomol. Zeitschrift *55*, 215–221.

S14 Piton, L.E. (1940). Paléontologie du gisement Éocène de Menat. Paris, Lechevalier.

S15 Vršanský, P., Oružinský, R., Barna, P., Vidlička, L., and Labandeira, C.C. (2014). Native *Ectobius* (Blattaria: Ectobiidae) from the Early Eocene green river formation of Colorado and its reintroduction to North America 49 million years later. Ann. Entomol. Soc. Am. *107*, 28–36.

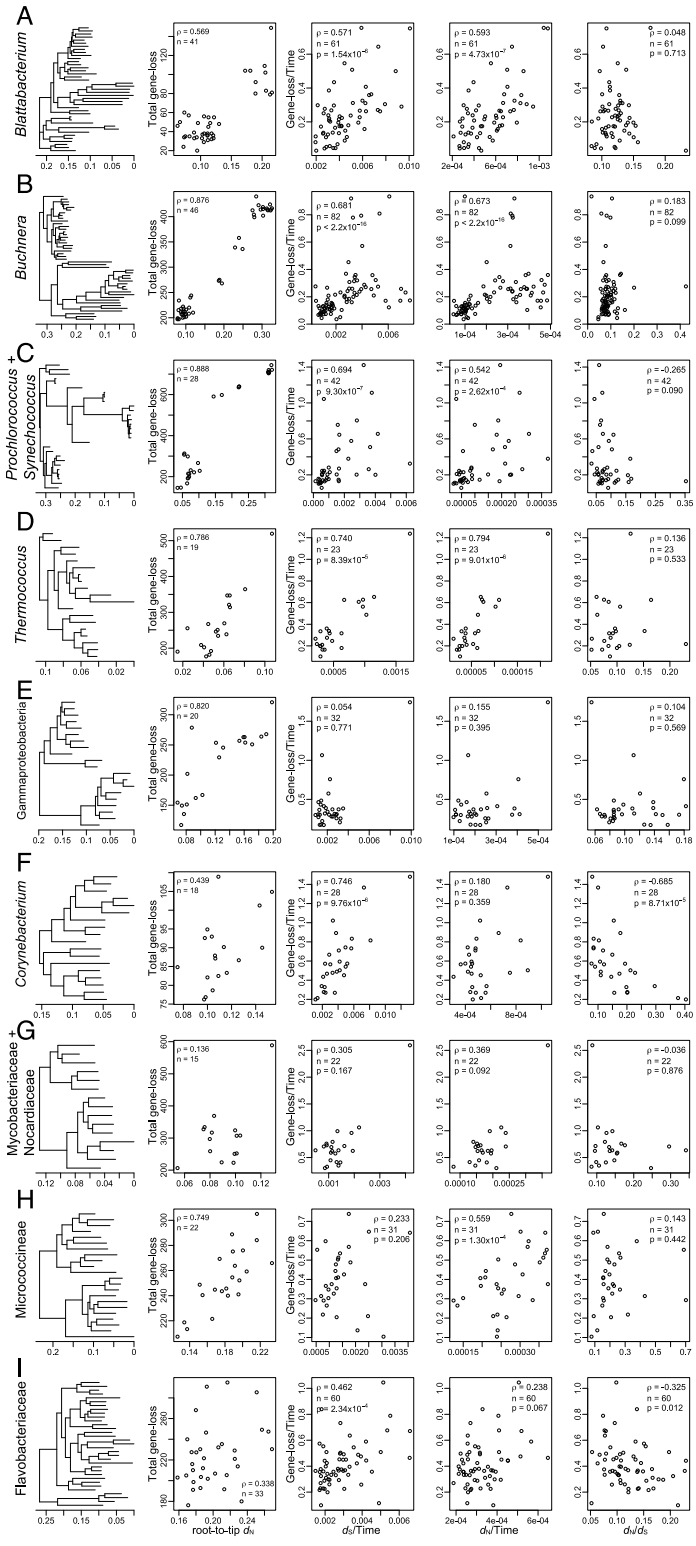
S16 Clark, M. A., Moran, N. A., and Baumann, P. (1999). Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol. Biol. Evol. *16*, 1586–1598.

S17 Kim, H., Lee, S., and Jang, Y. (2011). Macroevolutionary patterns in the Aphidini aphids (hemiptera: Aphididae): Diversification, host association, and biogeographic origins. PLOS One *6*, e24749.

S18 Dvořák, P. *et al.* (2014). *Synechococcus*: 3 billion years of global dominance. Mol. Ecol. *23*, 5538–5551.

S19 Martins-Neto, R.G., Mancuso, A., and Gallego, O.F. (2005). The Triassic insect fauna from Argentina. Blattoptera from the Los Rastros formation (Bermejo Basin), La Rioja Province. Ameghiniana *42*, 705–723.

S20 Lin, Q.B. (1980). Fossil insects from the Mesozoic of Zhejiang and Anhui provinces. Div. Correl. Mesozoic volcano-sedimentary Strat. Zhejiang Anhui Prov. 211–234.



Data S1. Evolution of genome reduction by gene loss in multiple prokaryote lineages. Related to Figure 3.

For each of nine prokaryote lineages (A-I), the following are provided from left to right: (i) phylogenetic trees inferred using maximum-likelihood analysis of 30–31 marker genes, with 3rd codon sites removed; (ii) relationship between total number

of gene losses and root-to-tip distance; (iii) Spearman's rank correlation between gene-loss/time and d_S/time; (iv) Spearman's rank correlation between gene-loss/time and d_N/time; and (v) Spearman's rank correlation between gene-loss/time and d_N/d_S, per terminal branch.