



Coevolution of Metabolic Pathways in Blattodea and Their *Blattabacterium* Endosymbionts, and Comparisons with Other Insect-Bacteria Symbioses

Yukihiro Kinjo,^{a,b,c} Thomas Bourguignon,^{a,d} Yuichi Hongoh,^{b,c} Nathan Lo,^e Gaku Tokuda,^f Moriya Ohkuma^b

^aOkinawa Institute of Science and Technology Graduate University, Okinawa, Japan

^bRIKEN Bioresource Research Centre, Tsukuba, Japan

^cSchool of Life Science and Technology, Tokyo Institute of Technology, Tokyo, Japan

^dFaculty of Tropical AgriSciences, Czech University of Life Sciences, Kamýcká, Prague, Czech Republic

^eSchool of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia

^fTropical Biosphere Research Center, University of the Ryukyus, Okinawa, Japan

ABSTRACT Many insects harbor bacterial endosymbionts that supply essential nutrients and enable their hosts to thrive on a nutritionally unbalanced diet. Comparisons of the genomes of endosymbionts and their insect hosts have revealed multiple cases of mutually-dependent metabolic pathways that require enzymes encoded in 2 genomes. Complementation of metabolic reactions at the pathway level has been described for hosts feeding on unbalanced diets, such as plant sap. However, the level of collaboration between symbionts and hosts that feed on more variable diets is largely unknown. In this study, we investigated amino acid and vitamin/cofactor biosynthetic pathways in Blattodea, which comprises cockroaches and termites, and their obligate endosymbiont *Blattabacterium cuenoti* (hereafter *Blattabacterium*). In contrast to other obligate symbiotic systems, we found no clear evidence of “collaborative pathways” for amino acid biosynthesis in the genomes of these taxa, with the exception of collaborative arginine biosynthesis in 2 taxa, *Cryptocercus punctulatus* and *Mastotermes darwiniensis*. Nevertheless, we found that several gaps specific to *Blattabacterium* in the folate biosynthetic pathway are likely to be complemented by their host. Comparisons with other insects revealed that, with the exception of the arginine biosynthetic pathway, collaborative pathways for essential amino acids are only observed in phloem-sap feeders. These results suggest that the host diet is an important driving factor of metabolic pathway evolution in obligate symbiotic systems.

IMPORTANCE The long-term coevolution between insects and their obligate endosymbionts is accompanied by increasing levels of genome integration, sometimes to the point that metabolic pathways require enzymes encoded in two genomes, which we refer to as “collaborative pathways”. To date, collaborative pathways have only been reported from sap-feeding insects. Here, we examined metabolic interactions between cockroaches, a group of detritivorous insects, and their obligate endosymbiont, *Blattabacterium*, and only found evidence of collaborative pathways for arginine biosynthesis. The rarity of collaborative pathways in cockroaches and *Blattabacterium* contrasts with their prevalence in insect hosts feeding on phloem-sap. Our results suggest that host diet is a factor affecting metabolic integration in obligate symbiotic systems.

KEYWORDS *Blattabacterium*, Blattodea, co-evolution, co-metabolism, endosymbiosis, insects

Many insects have established intimate relationships with bacterial symbionts that have helped them to exploit new niches. In many cases, intracellular symbionts provide nutrients to their insect host, while the host provides a stable environment

Editor Konstantinos Aristomenis Kormas, University Of Thessaly

Copyright © 2022 Kinjo et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yukihiro Kinjo, yukihirokinjo@gmail.com.

The authors declare no conflict of interest.

[This article was published on 12 September 2022 with the publication date incorrectly given as “13 September 2022.” The publication date was corrected in the current version, posted on 14 September 2022.]

Received 11 August 2022

Accepted 18 August 2022

Published 12 September 2022

favorable to symbiont growth (1, 2). These symbiotic relationships are often obligate, as the host and its symbionts depend on products encoded in the genome of their partners. Moreover, obligate symbioses are known to involve host-symbiont complementarity, whereby each partner relies on metabolic products and enzymes of the other to complete some of their metabolic pathways.

One example is branched-chain amino acid (BCAA) biosynthesis in aphids and their endosymbiotic *Buchnera aphidicola* (hereafter *Buchnera*). Aphids depend on *Buchnera* to supply many essential amino acids (EAAs) deficient in their phloem-sap diet (3, 4). However, the genome of *Buchnera* lacks the genes coding for branched-chain amino acid aminotransferase (*ilvE*), the enzyme involved in the last step of BCAA biosynthesis (5), which is present in the aphid genome (6). The genomes of both symbiotic partners are therefore complementary, sometimes at the level of genes within pathways, and together encode all genes involved in EAA biosynthesis (4, 7–9).

Such intimate metabolic interactions, which we term “collaborative pathways” (10), are not unique to aphids and have also been found in several other host-symbiont systems (e.g., 11, 12). Collaborative pathways typically evolve from hosts and symbionts endowed with complete functional pathways at the origin of symbiogenesis, but may depend on genes horizontally transferred from other bacteria to the host. In addition to the preexisting gene set of the symbiont pair, collaborative pathways require the evolution of unusual transport mechanisms for metabolites or enzymes between members of the pair. Collaborative pathways have been reported from obligate symbiotic systems that have a long history of coevolution (10, 13).

To date, the existence of collaborative pathways between hosts and symbionts has only been investigated in sap-feeding insects (13). Although a collaborative arginine biosynthetic pathway between carpenter ants and their obligate endosymbiont, *Blochmannia*, has been proposed based on the gene content of the symbiont genome (14), comparisons between the genomes of host insects with alternative diets and that of their symbionts have not yet been reported. A model for insect symbioses that do not involve sap-feeding taxa is that between Blattodea, which comprises cockroaches and termites (15), and their intracellular symbiont *Blattabacterium cuenoti* (hereafter *Blattabacterium*). Blattodea have evolved a wide range of feeding habits and are sometimes known as “garbage collectors,” recycling dead plants, dead animals, and excrement (16). Almost all cockroaches, as well as the most primitive termite *Mastotermes darwiniensis*, host *Blattabacterium* endosymbionts in specialized cells (bacteriocytes) located in their fat body. *Blattabacterium* appears to have been lost in the lineage leading to all termites other than *M. darwiniensis*, as it has not been found in these taxa. The association between Blattodea and *Blattabacterium* is ancient and was established >200 million years ago (17–19), which is comparable with the time frame of other symbiotic systems found in insects (e.g. 160–280 MYA in aphids-*Buchnera*, [20]). Physiological experiments and genomic analyses have suggested that *Blattabacterium* is involved in nitrogen recycling and essential amino acid biosynthesis (21–23). Unlike sap feeders, cockroaches have adapted to various diets, and many species are known to engage in coprophagy (16). Their diet is therefore more nutritionally balanced than plant sap. However, because their food sources are often transient (16), *Blattabacterium* is likely to contribute to host fitness through nitrogen recycling. Thus, metabolic collaborations between the two genomes are expected to be different from that of sap-feeding hosts and their symbionts.

The evolution of metabolic pathways in *Blattabacterium* is largely unknown. In contrast to the endosymbionts belonging to the phylum *Pseudomonadota* (previously *Proteobacteria*), to which *Escherichia coli* belongs, the enzyme repertoire of the order *Flavobacteriales* (phylum *Bacteroidota*), to which *Blattabacterium* belongs, is not well-characterized. Unknown enzymes and pathways are still frequent in lineages that do not include model organisms, affecting prokaryote genome annotation (24). Indeed, many bacteria have genomes annotated with incomplete amino acid biosynthetic pathways and are incorrectly assumed to be heterotrophic while they can be cultured without the given amino acids (25). For this reason, it is unclear whether the apparent gaps present within the pathways of *Blattabacterium* are genuine gaps, representing

gene losses during the evolution of *Blattabacterium* and complemented by their host Blattodea, or whether these genes are absent from the genomes of *Blattabacterium* and their free-living relatives because of the existence of yet-uncharacterized genes replacing the functions. Genes absent in all *Blattabacterium* genomes as well as the genomes of their free-living relatives are indicative of the presence of uncharacterized genes completing the pathways and are not expected to be compensated by the host genome.

In this study, we conducted a careful reconstruction of the metabolic pathways involved in amino acid and vitamin/cofactor biosynthesis of *Blattabacterium* and their host Blattodea. We investigated the genomic evidence of metabolic collaboration between host insects and *Blattabacterium*. To this end, we reconstructed the metabolic pathways of seven representative genomes of *Blattabacterium* and compared them to the metabolic pathways of other *Flavobacteriales* bacteria. This approach allowed us to distinguish genes that were lost in *Blattabacterium* and are complemented by their insect host from genes commonly absent in *Flavobacteriales* (whose inferred function is carried out by the products of yet-uncharacterized genes or pathways). We then inferred the metabolic pathways of Blattodea using the genome and transcriptome of 2 Blattodean species harboring *Blattabacterium* (*Periplaneta americana* and *Cryptocercus punctulatus*, respectively) and the genome of 1 Blattodean species devoid of *Blattabacterium* (*Zootermopsis nevadensis*). Using this approach, we obtained robust inference of the pathways common across Blattodea. We compared the inferred metabolic pathways of Blattodea and *Blattabacterium* and identified potential collaborative pathways. Finally, we compared the profile of collaborative pathways in *Blattabacterium* with those of other host-symbiont systems with variable diets to see how diet affects the profile of collaborative pathways.

RESULTS

Evolution of metabolic pathways in *Blattabacterium*. We analyzed the metabolic pathways present in 42 genomes of non-*Blattabacterium* *Flavobacteriales* bacteria and the genome of a representative of the genus *Bacteroides* of the Bacteroidales order, which we used as an outgroup (Table S1). We compared these with the metabolic pathways found in 7 *Blattabacterium* genomes. The presence/absence of each enzyme involved in amino acid and vitamin/cofactor biosynthesis is summarized for all genomes in Fig. 1. One unique characteristic of all *Blattabacterium* strains is the presence of urease (EC 3.5.1.5) in their genomes, a gene that is generally absent from the genomes of other strains of *Flavobacteriales*. Phylogenetic analysis of the urease large subunit (UreC) revealed that the sequences of *Blattabacterium* form a monophyletic cluster which is sister to sequences from members of the genus *Paenibacillus* (*Firmicutes*) (Fig. 2).

The biosynthetic pathways of four amino acids not essential in insects (hereafter non-EAAs), glutamine, asparagine, glycine, and proline, and two vitamins/cofactors, pantothenate (B5) and protoheme, were lost almost exclusively in *Blattabacterium* (Fig. 1). In contrast, the almost complete biosynthetic pathways of 10 amino acids essential in insects (including arginine (26); hereafter EAAs) were retained in all *Blattabacterium* strains, except in the strains hosted by *Cryptocercus punctulatus* (CPUbr) and *Mastotermes darwiniensis* (MADAR), which have lost several of these pathways (Fig. 1). At the within-pathway level, we identified several gaps (missing genes) in biosynthetic pathways. We defined gaps as missing enzymes (genes) in pathways otherwise complete. We recognized 2 categories of gaps: (i) "common gaps," which are shared among *Blattabacterium* and over 70% of other *Flavobacteriales* genomes (EC numbers highlighted in yellow in Fig. 1); and (ii) "specific gaps," which are generally specific to *Blattabacterium* (EC numbers highlighted in red in Fig. 1). Since common gaps are found in other strains of *Flavobacteriales*, regardless of their ecological habitats, and the other enzymes involved in the pathway are highly conserved, we assume gaps are complemented by alternative enzymes in the genomes of *Blattabacterium* and other *Flavobacteriales* rather than by metabolites obtained from the host (see Materials and Methods). On the contrary, specific gaps found only in *Blattabacterium* are likely to be complemented by host-derived metabolic intermediates, assuming that the host genome

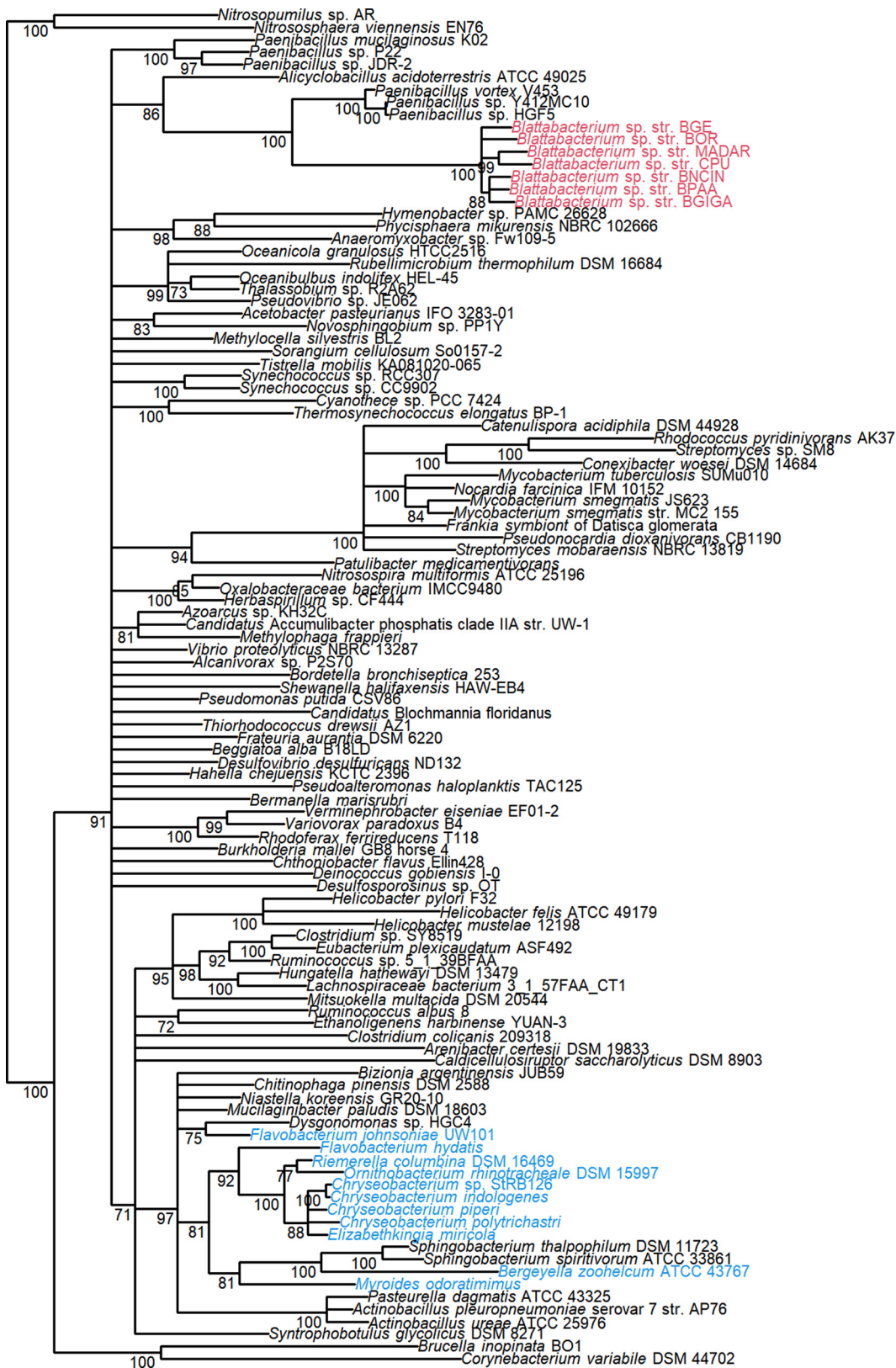


FIG 2 Phylogenetic analysis of the ureC gene in bacteria. Maximum likelihood phylogenetic tree of ureC gene reconstructed using RAxML with PROTCATLG amino acid substitution model. Blattabacterium and Flavobacteriales strains are colored in red and blue, respectively. Nodes with low (<70%) bootstrap support values were collapsed.

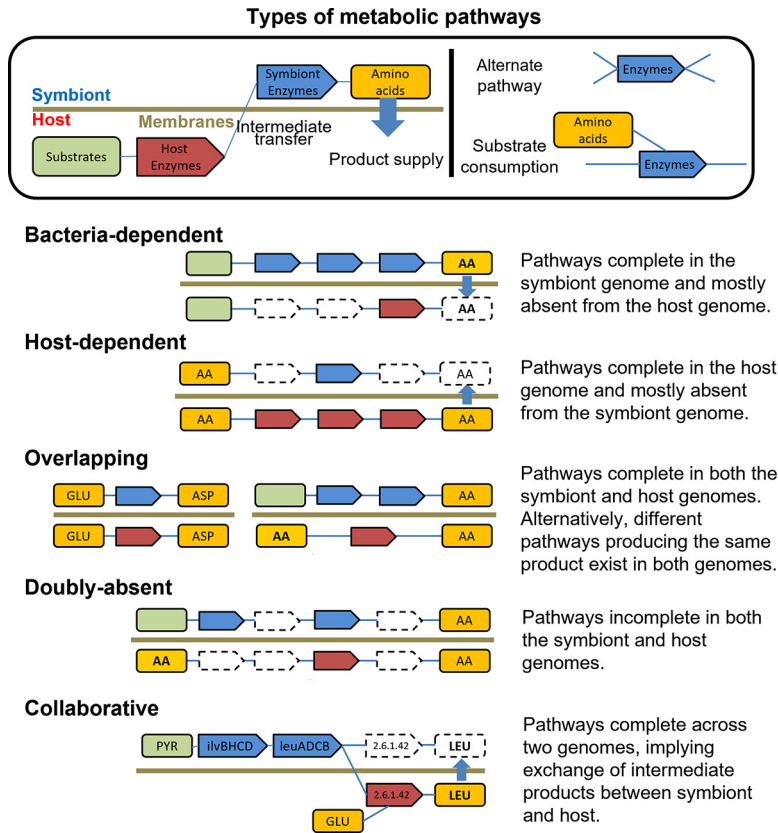


FIG 3 Types of metabolic interactions between *Blattabacterium* and their hosts. Boxes with connections represent biosynthetic pathways. Host enzymes and symbiont enzymes are represented by red and blue boxes, respectively. Blue arrows indicate products supplied to the symbiotic partner. Blue bars connecting box edges represent the routes of the pathways.

encodes genes complementing these gaps. Missing genes in the methionine, arginine, and riboflavin (B2; EC 3.5.4.25) biosynthetic pathways were assigned to common gaps. On the other hand, missing genes in the pyridoxine (B6), riboflavin (B2; EC 3.1.3.104), folate (B9), and menaquinone (K2; EC 6.2.1.26) biosynthetic pathways were assigned to the specific gap category. In addition, the enzyme involved in the final step of arginine biosynthesis, argininosuccinate lyase (ArgH; EC 4.3.2.1), and 2 enzymes involved in the synthesis of para-amino-benzoate (pabA and trpE; EC 2.6.1.85 and EC 4.1.3.38) from the folate biosynthetic pathway, were absent from the genomes of the *Blattabacterium* strains CPUbr and MADAR, and were therefore considered as specific gaps (Fig. 1). A gap in the serine biosynthetic pathway (EC 3.1.3.3) was not classified into any category as it is neither common across *Flavobacteriales* genomes nor specific to *Blattabacterium*.

Metabolic relationships between *Blattabacterium* strains and their hosts. To investigate the molecular basis of *Blattabacterium*-host metabolic collaboration, we compared the reconstructed metabolic pathways of *Blattabacterium* to those of their hosts, with the assumption that gaps common across *Flavobacteriales* and present in *Blattabacterium* probably do not need to be compensated by their hosts. We inferred the common metabolic pathways of Blattodea representatives by reconstructing consensus pathways from the cockroaches *P. americana* and *C. punctulatus* and the termite *Z. nevadensis*. The metabolic relationships between *Blattabacterium* and their hosts were classified into five categories (Fig. 3): *Blattabacterium*-dependent, host-dependent, overlapping, doubly-absent, and collaborative pathways.

The biosynthetic pathways of the 10 EAAs were all *Blattabacterium*-dependent except for CPUbr and MADAR. No horizontal gene transfers were detected for amino acid and cofactor biosynthetic pathways in *P. americana* and *Z. nevadensis* genomes. Conversely,

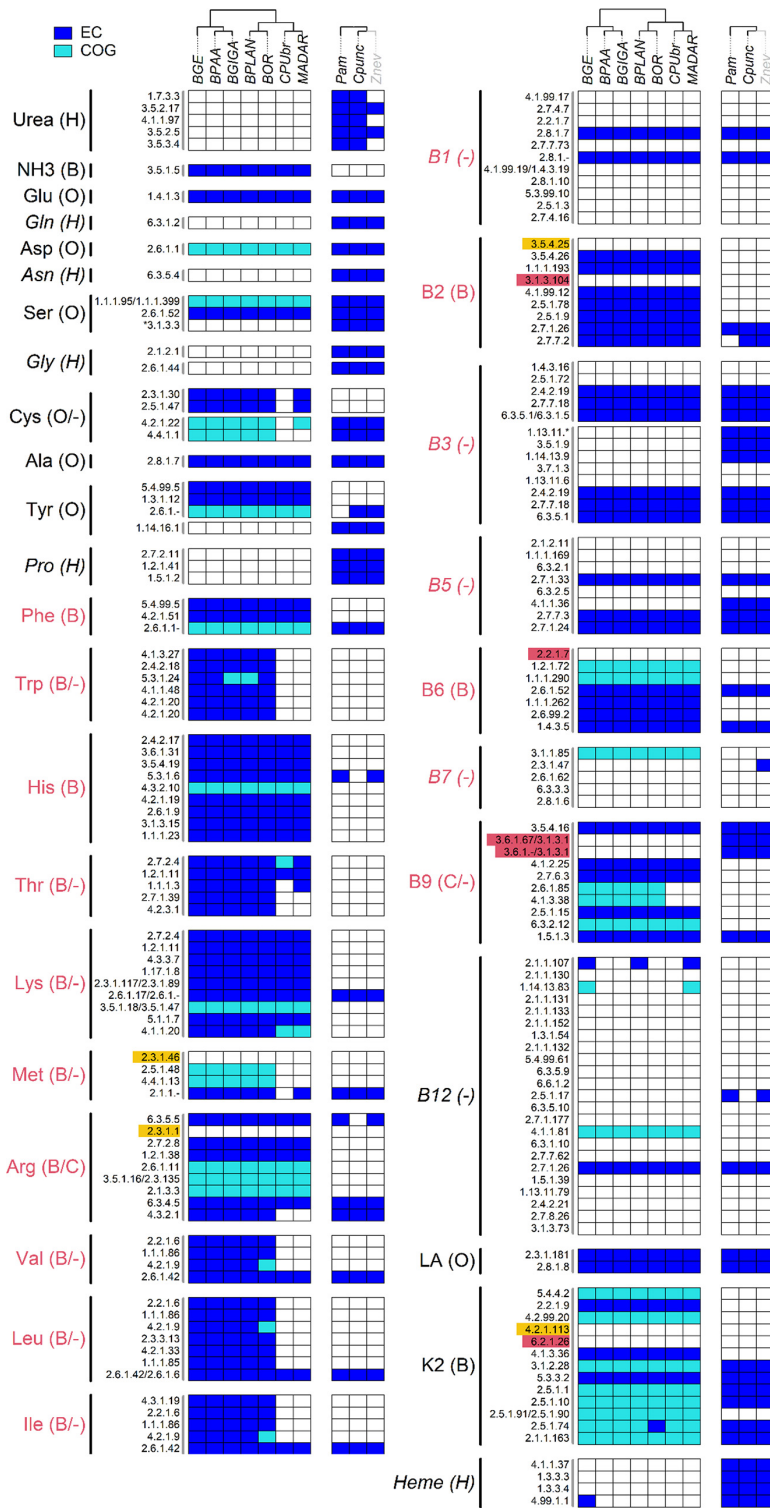


FIG 4 Pathway comparisons between *Blattabacterium* strains and their Blattodea host. Characters within parentheses next to the final products indicate: H, host-dependent; B, *Blattabacterium*-dependent; O, overlapping; -, doubly-absent; C, collaborative pathways.

non-EAA pathways were all host-dependent or overlapping. While the biosynthesis of glycine, asparagine, glutamine, and proline was categorized as host-dependent, the biosynthesis of serine, cysteine, alanine, tyrosine, glutamate, and aspartate was categorized as overlapping (Fig. 4). For the serine biosynthetic pathway, it is possible that the loss of SerB

in *Blattabacterium* genomes is complemented by phosphoserine phosphatase (EC 3.1.3.3) derived from the host genome. However, SerB is also absent from the genomes of several *Flavobacteriales* strains and an alternative enzyme catalyzing the same reaction, HisB (EC 3.1.3.-) (27), is present in *Blattabacterium* genomes and may compensate the lack of SerB, as previously suggested (23). We therefore categorized the serine biosynthetic pathway as overlapping. The initial steps of the host-*Blattabacterium* metabolic pathway for nitrogen metabolism, the uricolytic pathways, are complete in the two cockroach transcriptomes investigated in this study. On the other hand, the lower termite *Z. nevadensis*, which does not harbor *Blattabacterium*, lacks three enzymes in the uricolytic pathway: urate oxidase (EC 1.7.3.3), 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazole decarboxylase (EC 4.1.1.97), and allantoinase (EC 3.5.3.4) (Fig. 4).

Blattabacterium CPUbr and MADAR have undergone further genome reduction, being ~50kbp smaller than other strains (28, 29). The additional genome reduction in these strains is mostly due to the loss of many genes involved in EAA biosynthesis. It has been hypothesized that the loss of these genes occurred as a result of relaxed selection due to the provision of EAAs either by the gut symbiotic system established in the common ancestor of *C. punctulatus* and *M. darwiniensis* (28–31), and/or through social transfer of nutrients (proctodeal trophallaxis). One of these gene losses is present as a single gap in the arginine biosynthetic pathway, and consists of the loss of the gene encoding the enzyme involved in the final step of arginine biosynthesis, ArgH (EC 4.3.2.1). Although the products of the two *Blattabacterium* genes *purB* and *fumC* possibly replace the function of ArgH (30), the host genome also encodes arginosuccinate lyase (ASL: EC 4.3.2.1), an enzyme with a catalytic function similar to ArgH. The loss of ArgH has occurred in parallel in the ancestors of CPUbr and MADAR (29). A collaborative pathway involving the loss of ArgH by the endosymbiont *Carsonella* and complementation by a psyllid host gene was reported by Sloan et al. (2014), based on transcriptomic evidence. Therefore, we hypothesize that the arginine biosynthetic pathways of CPUbr and MADAR are collaborative (Fig. 4).

Among cofactor biosynthetic pathways, the riboflavin (B2), pyridoxine (B6), and menaquinone (K2) pathways were *Blattabacterium*-dependent (Fig. 4). The protoheme pathway was host-dependent, and the lipoate pathway was overlapping between *Blattabacterium* and their host (Fig. 4). The thiamine (B1), niacin (B3), pantothenate (B5), biotin (B7), and cobalamin (B12) biosynthetic pathways were doubly-absent (Fig. 4). Although we found several gaps in the riboflavin, pyridoxine, and folate pathways, we found no enzyme in the host genomes to compensate for these gaps, except for alkaline phosphatase (EC 3.1.3.1) of the folate pathway, suggesting that the pathways for pyridoxine and riboflavin biosynthesis use alternative enzymes, such as multifunctional enzymes, to catalyze these reactions. For the *Blattabacterium*-specific gap in the folate biosynthetic pathway, which comprises 2 reactions (EC 3.6.1.67 and EC 3.6.1.-), we identified a host enzyme, alkaline phosphatase (EC 3.1.3.1), which can bypass these reactions (32). Although we could not find any experimental evidence for this bypassed reaction, transcriptome-based evidence for this collaboration has been described in the symbiotic system composed of whiteflies and their secondary symbionts, *Hamiltonella* (33). We thus hypothesize that this pathway is a collaborative pathway in Blattodea, as is the case in whiteflies. *Blattabacterium* CPUbr and MADAR have additionally lost 2 enzymes in the folate biosynthetic pathway, PabA (EC 2.6.1.85) and TrpE (EC 4.1.3.38), which are involved in the *p*-aminobenzoate (PABA) synthesis. We found no enzymes in the host genomes compensating for the loss of the PABA branch. We therefore hypothesize that the folate biosynthetic pathway is doubly-absent in CPUbr and MADAR.

Comparison with other host-symbiont partnerships. Inferred metabolic relationships between insects and their symbionts (11, 14, 22, 23, 28, 29, 34–68) are summarized in Table 1. Pathway predictions for other insect-symbiont pairs were performed using the same procedure used for *Blattabacterium*. Note that pathways present in secondary symbionts are omitted due to high variability in genome reduction levels. The data shown in Table 1 thus only include the relationship between the host and its primary symbiont. Overall, collaborative pathways are non-uniformly distributed across

insect-symbiont pairs. While phloem-sap feeders tend to have many collaborations in EAA pathways, host-symbiont collaborative pathways in Blattodea and carpenter ants were only found in the arginine and folate biosynthetic pathways, and no collaborative pathways were found in the only xylem-sap feeder examined, Auchenorrhyncha-*Sulcia*. Both Blattodea and carpenter ants have predicted collaborative arginine biosynthetic pathways; however, the enzymes involved are different. Carpenter ants require ornithine carbamoyltransferase (ArgI: EC 2.1.3.3) from their *Blochmannia* endosymbionts to complement their arginine biosynthetic pathway (part of the urea cycle), while in Blattodea, CPUbr and MADAR require argininosuccinate lyase (ASL: EC 4.3.2.1) from their hosts to complement their *de novo* arginine biosynthetic pathway. In phloem-sap feeders, collaborative EAA, and in particular BCAA biosynthetic pathways, are especially common, regardless of the symbiont's taxonomic position. The commonness of collaborative BCAA biosynthetic pathways in these insects is likely due to the evolutionary stability of the interaction between host and symbiont. This stability is also found in standard biosynthetic pathways. For example, *Sulcia* strain OLIH, which has the genome among *Sulcia* strains used in this study, has lost genes for all essential amino acids and vitamins/cofactors except for those for BCAA biosynthesis (65). The *Blattabacterium* strains MADAR and CPUbr are in striking contrast to this pattern, as they have lost BCAA biosynthetic pathways together with several other EAA biosynthetic pathways. Another exception to this trend is found in partnerships between blood-sucking insects and their obligate endosymbionts, which have lost almost all EAA biosynthetic pathways while retaining almost all vitamin/cofactor pathways. For vitamins/cofactors, two collaborative biosynthetic pathways for pantothenate (B5) and folate (B9) biosynthesis were detected (Table 1). The former collaboration (B5) has been reported in aphids (69). In this study, Blattodea and carpenter ants (which feed on wood or are omnivorous) were predicted to have the latter (B9), and blood-feeders were predicted to have both (Fig. 5).

DISCUSSION

In this study, we analyzed the amino acid and vitamin/cofactor metabolic pathways of *Blattabacterium* endosymbionts from Blattodea and compared them to those of other *Flavobacteriales* genomes. We found that pathways for EAA biosynthesis are enriched in *Blattabacterium* genomes, while many non-EAA and vitamin/cofactor pathways have been lost. We further identified two types of gaps present in these pathways: (i) gaps shared by *Blattabacterium* and other *Flavobacteriales* strains, and (ii) gaps specific to *Blattabacterium*. We then analyzed the metabolic pathways of 3 species of Blattodea and compared them to those of *Blattabacterium* in order to examine potential metabolic collaborations in this symbiotic system. The product-level complementation (exchange of amino acids and vitamins between the host and its symbiont) shown in this study is consistent with results from previous studies, which performed genome-based flux balance analysis on *Blattabacterium* (70, 71). The pathways for several non-EAAs (Gln, Gly, Asn, and Pro) absent from *Blattabacterium* genomes were present in the transcriptome of *B. germanica* (72). While we found several examples of product-level complementation, we only found 2 candidate examples of within-pathway mutual dependence (defined as “collaborative pathways” in this study), in the folate biosynthetic pathway for all strains (except for CPUbr and MADAR), and in the arginine biosynthetic pathway for CPUbr and MADAR. We also compared metabolic pathways across 11 insect-symbiont systems comprised of 55 genomes belonging to various taxonomic groups and having diverse diets (Table 1, Table S3), and found that there is a clear pattern of putative collaboration related to the host diet. In addition to patterns related to diet, we also found that BCAA biosynthetic pathways have been retained by the endosymbionts of all sap-feeding insects examined here despite their significant levels of reductive genome evolution. In sharp contrast to this trend in sap-feeding insects, genome reduction in CPUbr and MADAR *Blattabacterium* involves the loss of BCAA pathways.

We found several unique characteristics in *Blattabacterium* metabolic pathways. Firstly,

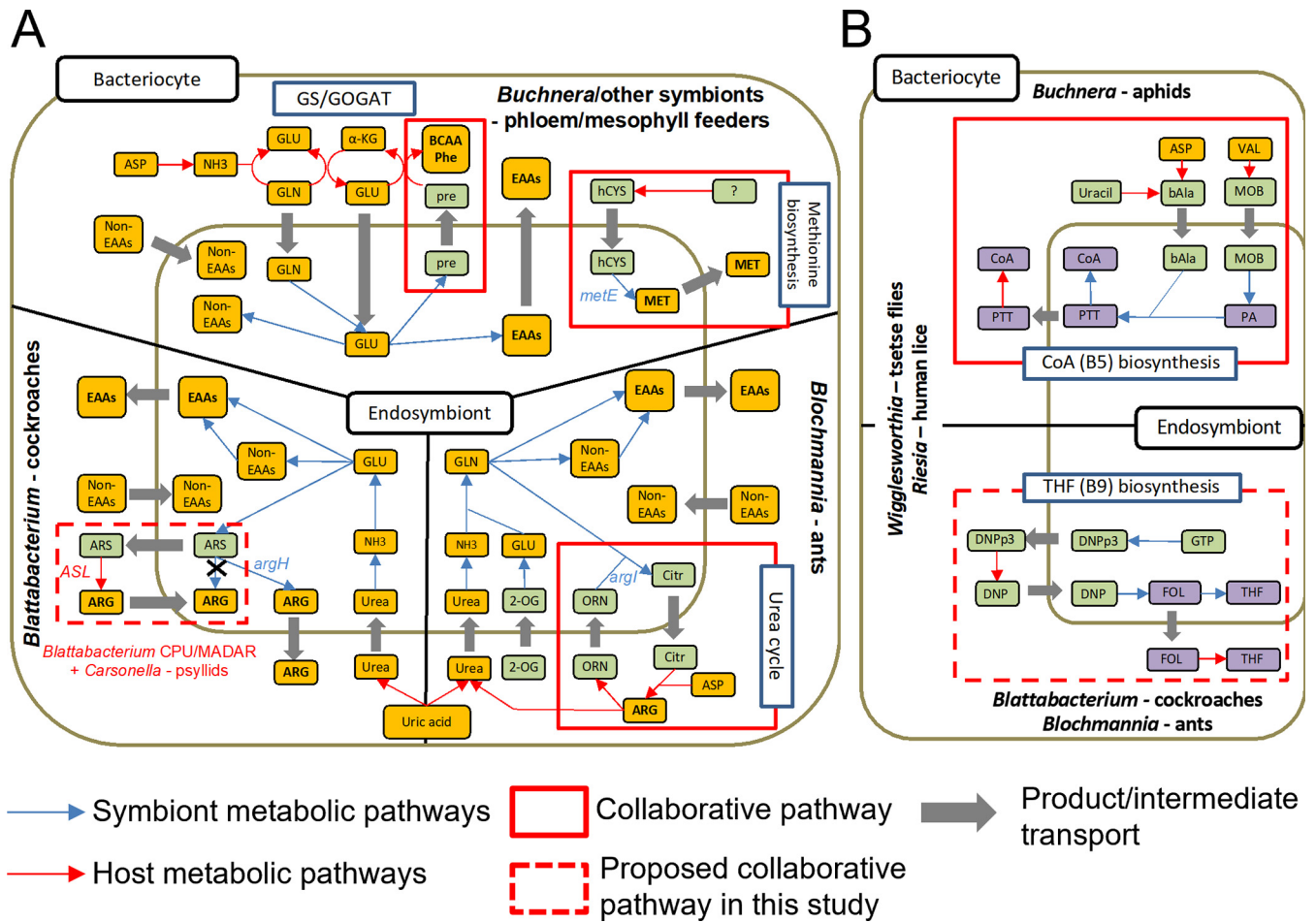


FIG 5 Schematic representations of metabolic interactions and distribution of collaborative pathways for amino acids biosynthesis among host-endosymbiont systems in insects. Narrow blue and red arrows indicate the metabolic pathways of the symbionts and hosts, respectively. Thick gray arrows indicate the transport of metabolites across membranes. Yellow, purple, and green boxes represent amino acids, vitamins/cofactors, and intermediate metabolites, respectively. (A) metabolic pathways for amino acid biosynthesis. (B) collaborative pathways for vitamin/cofactor biosynthesis. ARS, argininosuccinate; ORN, ornithine; Citr, citrulline; 2-OG, 2-Oxoglutarate; pre, precursors for BCAAs and phenylalanine; hCys, homocysteine; bAla, beta-alanine; PA, pantothenate; PTT, pantotheine; MOB, 3-methyl-2-oxobutanoate; DNPP3, 7,8-dihydroneopterin 3'-triphosphate; DNP, dihydroneopterin; FOL, folate; THF, tetrahydrofolate.

although urease is present in *Blattabacterium* and in the pathogenic strains of the genera *Weeksella*, *Ornithobacterium*, *Riemerella*, and *Chryseobacterium*, it is generally absent from the genomes of other *Flavobacteriales* genera. A previous study suggested that the presence of a urea transporter and urease accessory proteins is highly correlated with non-marine pathogenic *Flavobacteriales* ecotypes (73). Interestingly, our phylogenetic analysis of UreC protein suggests that *Blattabacterium* has acquired urease via horizontal gene transfer while almost all ureases from other *Flavobacteriales* strains are monophyletic (Fig. 2). The universal presence of this gene among all known *Blattabacterium* strains indicates it was present in their last common ancestor. The conversion of uric acid to usable nitrogen in the form of NH₃ occurs via collaboration between the host (which performs the first steps [72]) and *Blattabacterium*. This process is an essential trigger for the biosynthesis of many nutrients (22, 23) and forms a key part of the symbiotic relationship between these partners. Other specific characteristics of the early stage of *Blattabacterium* genome evolution include losses of several non-EAA pathways and the retention of EAA pathways, which suggest that *Blattabacterium* has evolved with the strict requirement of supplying EAAs to their hosts.

Blattodea are known to contain a high level of cobalamin (B12) in their hindgut (74, 75). Since both host and *Blattabacterium* lack biosynthetic pathway of B12, the reported higher cobalamin concentration in the hindgut of Blattodea is likely due to

the presence of B12-producing microbes in their hindgut, such as *Shimwellia blattae* (76). Unlike other animals, insects are not thought to require cobalamin as none of the cobalamin-dependent enzymes were detected from 19 publicly available genomes (77). We searched for the 3 cobalamin-dependent eukaryotic enzymes and could not detect them in the Blattodea genomes used in this study. We expect that the high level of B12 in the hindgut of Blattodea is not required by the host or its *Blattabacterium*.

Our study shows that the metabolic host-symbiont relationships in the Blattodea differ from those of other insects. With respect to collaborative pathways for amino acids biosynthesis, we propose 3 categories: (i) phloem-sap feeders, which commonly have collaborative BCAAs, phenylalanine, and methionine biosynthetic pathways; (ii) carpenter ants, which have a collaborative arginine biosynthetic pathway; (iii) and the 2 blattodeans *C. punctulatus* and *M. darwiniensis* and psyllids, which have another type of collaborative arginine biosynthesis pathway (Fig. 5A). Blood-feeders do not have any biosynthetic pathways for EAAs. Interestingly, while phloem-sap feeders have evolved similar patterns of collaboration, xylem-sap feeders do not appear to possess any collaborative pathways. With respect to vitamin/cofactor pathways, we propose 3 categories: (i) aphids, which have a collaborative pantothenate biosynthetic pathway; (ii) carpenter ants and Blattodea, which have collaborative folate biosynthetic pathways; (iii) blood-feeders, which have collaborative pathways for both pantothenate and folate (Fig. 5B).

A previous study suggested that host gene content can be a constraint for collaborative pathway evolution (10). In the case of BCAA biosynthesis, the branched-chain-amino-acid transaminase (BCAT: EC 2.6.1.42) is ubiquitous in insects, indicating that collaborative BCAA biosynthesis has the potential to evolve. Similarly, the 2 collaborative pathways in vitamin/cofactor biosynthesis we describe here are supported by the universal presence of the beta-alanine biosynthetic pathway and the alkaline phosphatase gene in insect genomes. The type of cell that gave rise to bacteriocytes is also thought to potentially constrain the evolution of collaborative pathways because the cells require adequate levels of gene expression to complement gaps in the symbiont's pathways (10). The potential for forming collaborative pathways is thus expected to vary among different insect hosts, depending on gene expression levels in particular cells, even when the hosts possess the same set of genes. Previous studies have reported that within a host individual, different (primary and secondary) symbionts are harbored within morphologically distinct bacteriocytes (57, 78, 79), and the bacteriocytes have distinct gene expression profiles (80).

Although patterns of collaborative pathways across varying symbiotic systems are potentially explained by the "bacteriocyte cell type origin" hypothesis, our results suggest that collaborative pathways are driven by the host diet, and particularly the EAA and vitamin/cofactor profile required by the host. Previous studies have suggested that collaborative EAA biosynthesis is controlled by the host through control over amino donor supply (especially the supply of glutamate for BCAAs and phenylalanine) (3). The production rate of BCAAs and phenylalanine in aphid bacteriocytes was estimated to exceed 60% of total amino acid production (3). The glutamine synthetase/glutamine oxoglutarate aminotransferase (GS/GOGAT) cycle in the host cytoplasm is also thought to have a central role in efficient ammonia assimilation, which keeps toxic ammonia at low levels in bacteriocytes (8). Generally, phloem-sap is rich in asparagine and sugars, and thus symbiotic systems can obtain sufficient ammonia and ATP to activate the GS/GOGAT cycle. In contrast, xylem-sap is generally poorer in both carbon and nitrogen sources than phloem-sap, which might explain why Auchenorrhyncha-*Sulcia* systems lack collaborative pathways based on GS/GOGAT cycle. In this context, Blattodea do not need to use the GS/GOGAT cycle system because they use uric acid to store nitrogen wastes, and *Blattabacterium* utilizes host-provided urea as a nitrogen source (Fig. 5). Because Blattodea have various feeding habits, the EAA profile of their food source is less specific than plant sap. Therefore, control over the EAA production profile is likely to be less important in this system. The collaborative arginine biosynthetic pathway in the carpenter ant-*Blochmannia* system is believed to be linked to the use of arginine as a nitrogen storage compound in ants (14). This potentially

explains the absence of the same type of collaborative arginine biosynthetic pathways found in carpenter ants and Blattodea. In termites, uric acid is transported into the gut via the malpighian tubules and degraded by gut microbiota (81). This may explain the absence of uricolytic ability in the genome of the lower termite *Z. nevadensis*, which has lost *Blattabacterium*.

Our study has revealed several general trends of metabolic pathway evolution within and across insects and their symbionts. We found genomic evidence that symbiont partnerships vary in their amino acid and vitamin/cofactor biosynthetic collaborations. Experimental evidence, such as differential gene expression analysis or enzymatic assays, is required to identify the precise mechanisms by which these collaborations occur. For example, in the case of gaps unique to one or several strains of symbionts, complementation by unknown host enzymes or the emergence of new multifunctional enzymes should be considered as potential replacements (69). Our analyses revealed several gaps specific to *Blattabacterium* and several gaps common across *Flavobacteriales*. However, since our approach focused on interpreting gaps rather than reconstructing whole system metabolic networks, it is unclear how the metabolic collaborations affect the total metabolic flux within the symbiotic system. Although multi-organism scale modeling methods are needed to answer such questions, these modeling methods require experimental evidence, such as metabolite analyses, chemically-defined diets, or differential gene expression data, to make accurate predictions. Future studies are needed to determine the precise mechanisms and adaptive significance of pathway complementation in insect-symbiont systems and to provide a clear picture of the evolution of their genomes.

MATERIALS AND METHODS

Genome and transcriptome data of hosts. We used the genomes of 2 species of Blattodea, *Periplaneta americana* (82) and *Zootermopsis nevadensis* (83), and the transcriptome of a cockroach, *Cryptocercus punctulatus* (84), to infer the common metabolic pathways of Blattodean hosts. Although *Z. nevadensis* does not harbor *Blattabacterium*, we assumed that the metabolic pathways shared among these three distantly related species are common across Blattodean hosts. We did not use the published genome of the cockroach *Blattella germanica* (85) because its completeness, evaluated with BUSCO ver. 4 (86), was markedly lower than that of the 2 genomes and the transcriptome publicly available (Table S1). Because gene prediction was not available for the genome of *P. americana*, we conducted gene prediction using a transcriptome data set of *P. americana* generated by another study (87). Details concerning gene prediction and transcriptome assembly are available in the supplementary methods.

Genomes of *Blattabacterium* and related *Flavobacteriales* bacteria. The genomes of 7 strains of *Blattabacterium* were used for metabolic reconstruction. We also reconstructed the metabolic pathways of 42 genomes of other *Flavobacteriales* bacteria that are not obligate mutualists and one genome from the order *Bacteroidales*. All bacterial genomes used for the metabolic reconstruction are listed in Table S2. Phylogenetic positions of *Flavobacteriales* bacterial genomes are shown in Fig. S1.

Phylogenetic tree inference. Orthologous genes present in all studied genomes of *Flavobacteriales* (Table S2) were determined using Proteinortho5 (88) with >0.9 cluster quality threshold. Thirty-one representative marker genes for bacteria were extracted from orthologous sets using AMPHORA2 (89). Extracted sequences were aligned using MAFFT ver. 7 (90) in “linsi” mode, and ambiguously aligned sites were removed with Gblocks (91) using default parameters. Trimmed amino acid alignments were checked for gene conversion using Genconv under default parameters and concatenated with FASconCAT-G (92). All sequences of the concatenated alignment were re-coded into a Dayhoff6 matrix to reduce potential biases due to base composition heterogeneity between bacterial species (93). A maximum likelihood phylogenetic tree was calculated using RAxML ver. 8 (94) with the GTR-MULTIGAMMA substitution model. Branch support was estimated using 100 bootstrap replicates. For the UreC protein, all available bacterial and archaeal sequences were obtained from OrthoDB v10 (95), and the sequences were filtered with respect to length (sequences with <550 and >650 amino acids were removed). The remaining sequences, except for *Flavobacteriales*, were further filtered using the VSEARCH clustering method (96) with a 75% identity threshold. Alignments were prepared as described above. A maximum likelihood phylogenetic tree was reconstructed using RAxML ver. 8 with the PROTGAMMALG amino acid substitution model.

Metabolic reconstruction. We carried out metabolic reconstructions for the hosts, *Blattabacterium*, and several strains of free-living bacteria belonging to *Flavobacteriales*. Because annotations based solely on sequence homology occasionally contain inaccurate assignments (97), enzyme-coding genes were identified and re-curated using a “profile homology” information approach. First, 2 profile-based annotation programs, InterProScan (98), and SPARCLE (99), were run against all CDSs of the genome. Each CDS with an E-value <1E-10 was given an enzyme commission (EC) number using EC2GO mapping (100). We used EC assignments with four and 3 digits (i.e., EC X.X.X.X or EC X.X.X.-) to reduce ambiguous EC assignments. This domain-based method is known to have high accuracy (100); nonetheless, some enzyme families and subfamilies were not appropriately annotated using this method (101). Therefore, the sequences with no EC assignment were annotated using RPSBLAST searches against the COG database (102). The final set of EC numbers and COG IDs were mapped onto MetaCyc pathways (103),

and the direction of each reaction was inferred, when possible, using information from the BRENDA database (104).

We investigated whether gaps or missing enzymes in *Blattabacterium* metabolic pathways were also found in related bacteria. Gaps conserved across related bacterial lineages imply the presence of alternative/unknown enzymes complementing the pathways. Phylogenetic comparative methods have previously been used to identify novel genes completing gaps in metabolic pathways (105–107). In this study, we focused on determining whether pathways with gaps are complemented. We did not attempt to identify the enzymes filling the gaps. We investigated all biosynthetic pathways having up to 2 gaps. We assumed that gaps are complemented by alternative enzymes if they are shared by over 70% of other *Flavobacteriales* genomes. We refer to these gaps as “common gaps,” which all share three characteristics: (i) evolutionary stability; (ii) lineage-specificity, implying they are reactions which are complemented by alternative/unknown enzymes (25, 108); and (iii) representative of the ecological diversity of *Flavobacteriales* bacteria (109). The former 2 characteristics strongly support that the pathways with common gaps are functional, and the third characteristic supports that the dependence on the same metabolic intermediates is unlikely. Gaps were considered specific to *Blattabacterium* when shared by the 7 *Blattabacterium* genomes analyzed here but present in less than 10% of other *Flavobacteriales*. We refer to these gaps as “specific gaps”. Specific gaps were assumed to be complemented by the host if enzymes catalyzing the same reaction were found in the host genome. In case no such enzymes were found in the host genome, we assume the presence of alternative enzymes specifically found in *Blattabacterium*.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

SUPPLEMENTAL FILE 2, XLS file, 0.2 MB.

ACKNOWLEDGMENTS

This work was supported by KAKENHI 21K15071 to Y.K., 17H01447 and 19H05689 to M.O., 18K14767 to T.B., 17H01510, 26292177, 26117722 to G.T., and by OIST core funding.

Y.K. was supported by RIKEN Junior Research Associate program. N.L. was supported by an Australian Research Council Future Fellowship.

Accession numbers for the genome and transcriptome data sets analyzed in this study are listed in online supplementary material (Table S1 and Table S2).

REFERENCES

- Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190. <https://doi.org/10.1146/annurev.genet.41.110306.130119>.
- Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote–animal symbioses. *Nat Rev Genet* 9: 218–229. <https://doi.org/10.1038/nrg2319>.
- Macdonald SJ, Lin GG, Russell CW, Thomas GH, Douglas AE. 2012. The central role of the host cell in symbiotic nitrogen metabolism. *Proc Biol Sci* 279:2965–2973. <https://doi.org/10.1098/rspb.2012.0414>.
- Shigenobu S, Wilson ACC. 2011. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cell Mol Life Sci* 68: 1297–1309. <https://doi.org/10.1007/s00018-011-0645-2>.
- Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 407:81–86. <https://doi.org/10.1038/35024074>.
- The International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol* 8:e1000313. <https://doi.org/10.1371/journal.pbio.1000313>.
- Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan PF, Macdonald SJ, Schwartz JF, Thomas GH, Douglas AE. 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*: Amino acid metabolism genes in the pea aphid symbiosis. *Insect Mol Biol* 19: 249–258. <https://doi.org/10.1111/j.1365-2583.2009.00942.x>.
- Hansen AK, Moran NA. 2011. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci U S A* 108:2849–2854. <https://doi.org/10.1073/pnas.1013465108>.
- Russell CW, Bouvaine S, Newell PD, Douglas AE. 2013. Shared metabolic pathways in a coevolved insect–bacterial symbiosis. *Appl Environ Microbiol* 79:6117–6123. <https://doi.org/10.1128/AEM.01543-13>.
- Wilson ACC, Duncan RP. 2015. Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc Natl Acad Sci U S A* 112:10255–10261. <https://doi.org/10.1073/pnas.1423305112>.
- Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson ACC, von Dohlen CD, Fukatsu T, McCutcheon JP. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153:1567–1578. <https://doi.org/10.1016/j.cell.2013.05.040>.
- Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol Biol Evol* 31:857–871. <https://doi.org/10.1093/molbev/msu004>.
- Hansen AK, Moran NA. 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol Ecol* 23:1473–1496. <https://doi.org/10.1111/mec.12421>.
- Gil R, Silva FJ, Zientz E, Delmotte F, Gonzalez-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Hollidobler B, van Ham RCHJ, Gross R, Moya A. 2003. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc Natl Acad Sci U S A* 100: 9388–9393. <https://doi.org/10.1073/pnas.1533499100>.
- Lo N, Tokuda G, Watanabe H, Rose H, Slaytor M, Maekawa K, Bandi C, Noda H. 2000. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Curr Biol* 10:801–804. [https://doi.org/10.1016/s0960-9822\(00\)00561-3](https://doi.org/10.1016/s0960-9822(00)00561-3).
- Bell WJ, Roth LM, Nalepa CA. 2007. *Cockroaches: Ecology, Behavior, and Natural History*. JHU Press, Charles Village Baltimore.
- Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T. 2003. Evidence for cocladogenesis between diverse Dictyopteran lineages and their intracellular endosymbionts. *Mol Biol Evol* 20:907–913. <https://doi.org/10.1093/molbev/msg097>.

18. Bourguignon T, Tang Q, Ho SYW, Juna F, Wang Z, Arab DA, Cameron SL, Walker J, Rentz D, Evans TA, Lo N. 2018. Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Mol Biol Evol* 35:970–983. <https://doi.org/10.1093/molbev/msy013>.
19. Evangelista DA, Wipfler B, Béthoux O, Donath A, Fujita M, Kohli MK, Legendre F, Liu S, Machida R, Misof B, Peters RS, Podsiadlowski L, Rust J, Schuette K, Tollenaar W, Ware JL, Wappler T, Zhou X, Meusemann K, Simon S. 2019. An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proc Biol Sci* 286: 20182076. <https://doi.org/10.1098/rspb.2018.2076>.
20. Moran NA, Munson MA, Baumann P, Ishikawa H. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc R Soc Lond B Biol Sci* 253:167–171. <https://doi.org/10.1098/rspb.1993.0098>.
21. Cochran DG. 1985. Nitrogen excretion in cockroaches. *Annu Rev Entomol* 30:29–49. <https://doi.org/10.1146/annurev.en.30.010185.000333>.
22. López-Sánchez MJ, Neef A, Peretó J, Patiño-Navarrete R, Pignatelli M, Latorre A, Moya A. 2009. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genet* 5:e1000721. <https://doi.org/10.1371/journal.pgen.1000721>.
23. Sabree ZL, Kambhampati S, Moran NA. 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc Natl Acad Sci U S A* 106:19521–19526. <https://doi.org/10.1073/pnas.0907504106>.
24. de Crécy-Lagard V. 2014. Variations in metabolic pathways create challenges for automated metabolic reconstructions: Examples from the tetrahydrofolate synthesis pathway. *Comput Struct Biotechnol J* 10:41–50. <https://doi.org/10.1016/j.csbj.2014.05.008>.
25. Price MN, Zane GM, Kuehl JV, Melnyk RA, Wall JD, Deutschbauer AM, Arkin AP. 2018. Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. *PLoS Genet* 14:e1007147. <https://doi.org/10.1371/journal.pgen.1007147>.
26. Reddy SRR, Campbell JW. 1977. Invariant basis for the nutritional requirement of arginine in insects. *Experientia* 33:160–161. <https://doi.org/10.1007/BF02124040>.
27. Yip SH-C, Matsumura I. 2013. Substrate ambiguous enzymes within the *Escherichia coli* proteome offer different evolutionary solutions to the same problem. *Mol Biol Evol* 30:2001–2012. <https://doi.org/10.1093/molbev/mst105>.
28. Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA. 2012. Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Appl Environ Microbiol* 78:204–210. <https://doi.org/10.1128/AEM.06540-11>.
29. Kinjo Y, Bourguignon T, Tong KJ, Kuwahara H, Lim SJ, Yoon KB, Shigenobu S, Park YC, Nalepa CA, Hongoh Y, Ohkuma M, Lo N, Tokuda G. 2018. Parallel and gradual genome erosion in the *Blattabacterium* endosymbionts of *Mastotermes darwiniensis* and *Cryptocercus* wood roaches. *Genome Biol Evol* 10:1622–1630. <https://doi.org/10.1093/gbe/evy110>.
30. Neef A, Latorre A, Peretó J, Silva FJ, Pignatelli M, Moya A. 2011. Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities. *Genome Biol Evol* 3:1437–1448. <https://doi.org/10.1093/gbe/evr118>.
31. Kinjo Y, Lo N, Martín PV, Tokuda G, Pigolotti S, Bourguignon T. 2021. Enhanced mutation rate, relaxed selection, and the “domino effect” are associated with gene loss in *Blattabacterium*, a cockroach endosymbiont. *Mol Biol Evol* 38:3820–3831. <https://doi.org/10.1093/molbev/msab159>.
32. Klein CC, Alves JMP, Serrano MG, Buck GA, Vasconcelos ATR, Sagot M-F, Teixeira MMG, Camargo EP, Motta MCM. 2013. Biosynthesis of vitamins and cofactors in bacterium-harboring Trypanosomatids depends on the symbiotic association as revealed by genomic analyses. *PLoS One* 8: e79786. <https://doi.org/10.1371/journal.pone.0079786>.
33. Luan J-B, Chen W, Hasegawa DK, Simmons AM, Wintermantel WM, Ling K-S, Fei Z, Liu S-S, Douglas AE. 2015. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol* 7:2635–2647. <https://doi.org/10.1093/gbe/evv170>.
34. Kinjo Y, Saitoh S, 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. <https://doi.org/10.1264/jsme2.ME14153>.
35. Huang CY, Sabree ZL, Moran NA. 2012. Genome sequence of *Blattabacterium* sp. strain BGIGA, endosymbiont of the *Blaberus giganteus* cockroach. *J Bacteriol* 194:4450–4451. <https://doi.org/10.1128/JB.00789-12>.
36. Patiño-Navarrete R, Moya A, Latorre A, Peretó J. 2013. Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. *Genome Biol Evol* 5:351–361. <https://doi.org/10.1093/gbe/evt011>.
37. Williams LE, Wernegreen JJ. 2015. Genome evolution in an ancient bacteria-ant symbiosis: parallel gene loss among *Blochmannia* spanning the origin of the ant tribe Camponotini. *PeerJ* 3:e881. <https://doi.org/10.7717/peerj.881>.
38. Williams LE, Wernegreen JJ. 2013. Sequence context of indel mutations and their effect on protein evolution in a bacterial endosymbiont. *Genome Biol Evol* 5:599–605. <https://doi.org/10.1093/gbe/evt033>.
39. Degnan PH, Lazarus AB, Wernegreen JJ. 2005. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Res* 15:1023–1033. <https://doi.org/10.1101/gr.3771305>.
40. Williams LE, Wernegreen JJ. 2010. Unprecedented loss of ammonia assimilation capability in a urease-encoding bacterial mutualist. *BMC Genomics* 11:687. <https://doi.org/10.1186/1471-2164-11-687>.
41. Cassone BJ, Wenger JA, Michel AP. 2015. Whole genome sequence of the soybean aphid endosymbiont *Buchnera aphidicola* and genetic differentiation among biotype-specific strains. *J Genomics* 3:85–94. <https://doi.org/10.7150/jgen.12975>.
42. Pérez-Brocail V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A. 2006. A small microbial genome: the end of a long symbiotic relationship? *Science* 314:312–313. <https://doi.org/10.1126/science.1130441>.
43. van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernández JM, Jiménez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Morán F, Moya A. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci U S A* 100:581–586. <https://doi.org/10.1073/pnas.0235981100>.
44. Chong RA, Park H, Moran NA. 2019. Genome evolution of the obligate endosymbiont *Buchnera aphidicola*. *Mol Biol Evol* 36:1481–1489. <https://doi.org/10.1093/molbev/msz082>.
45. Degnan PH, Ochman H, Moran NA. 2011. Sequence conservation and functional constraint on intergenic spacers in reduced genomes of the obligate symbiont *Buchnera*. *PLoS Genet* 7:e1002252. <https://doi.org/10.1371/journal.pgen.1002252>.
46. Sloan DB, Moran NA. 2012. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Mol Biol Evol* 29: 3781–3792. <https://doi.org/10.1093/molbev/mss180>.
47. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. 2006. The 160-kilobase genome of the Bacterial endosymbiont *Carsonella*. *Science* 314:267–267. <https://doi.org/10.1126/science.1134196>.
48. Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, Gurgui M, Oldham NJ, van Echten-Deckert G, Okamura K, Yamamoto K, Inoue H, Ohkuma M, Hongoh Y, Miyagishima S, Hattori M, Piel J, Fukatsu T. 2013. Defensive bacteriome symbiont with a drastically reduced genome. *Curr Biol* 23: 1478–1484. <https://doi.org/10.1016/j.cub.2013.06.027>.
49. Sloan DB, Moran NA. 2012. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biol Lett* 8:986–989. <https://doi.org/10.1098/rsbl.2012.0664>.
50. Santos-García D, Farnier P-A, Beitia F, Zchori-Fein E, Vavre F, Mouton L, Moya A, Latorre A, Silva FJ. 2012. Complete genome sequence of “*Candidatus* Portiera aleyrodidarum” BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *J Bacteriol* 194: 6654–6655. <https://doi.org/10.1128/JB.01793-12>.
51. Chen W, Hasegawa DK, Kaur N, Klot A, Pinheiro PV, Luan J, Stensmyr MC, Zheng Y, Liu W, Sun H, Xu Y, Luo Y, Kruse A, Yang X, Kontsedalov S, Lebedev G, Fisher TW, Nelson DR, Hunter WB, Brown JK, Jander G, Cilia M, Douglas AE, Ghanim M, Simmons AM, Wintermantel WM, Ling K-S, Fei Z. 2016. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol* 14:110. <https://doi.org/10.1186/s12915-016-0321-y>.
52. Sloan DB, Moran NA. 2013. The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biol Evol* 5:783–793. <https://doi.org/10.1093/gbe/evt044>.
53. McCutcheon JP, von Dohlen CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* 21:1366–1372. <https://doi.org/10.1016/j.cub.2011.06.051>.
54. López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. 2011. Complete genome sequence of “*Candidatus* Tremblaya princeps” strain PCVAL, an

- intriguing translational machine below the living-cell status. *J Bacteriol* 193:5587–5588. <https://doi.org/10.1128/JB.05749-11>.
55. Sabree ZL, Huang CY, Okusu A, Moran NA, Normark BB. 2013. The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armoured scale insects: Beneficial endosymbiont of armoured scale insects. *Environ Microbiol* 15:1988–1999. <https://doi.org/10.1111/1462-2920.12058>.
 56. Rosas-Pérez T, Rosenblueth M, Rincón-Rosales R, Mora J, Martínez-Romero E. 2014. Genome sequence of “*Candidatus* Walczuchella monophlebidarum” the flavobacterial endosymbiont of *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). *Genome Biol Evol* 6:714–726. <https://doi.org/10.1093/gbe/evu049>.
 57. Bennett GM, Moran NA. 2013. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol Evol* 5:1675–1688. <https://doi.org/10.1093/gbe/evt118>.
 58. Bennett GM, McCutcheon JP, MacDonald BR, Romanovicz D, Moran NA. 2014. Differential genome evolution between companion symbioses in an insect-bacterial symbiosis. *mBio* 5:e01697-14. <https://doi.org/10.1128/mBio.01697-14>.
 59. Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, Lapidus A, Wu D, McCutcheon JP, McDonald BR, Moran NA, Bristow J, Cheng J-F. 2010. One bacterial cell, one complete genome. *PLoS One* 5:e10314. <https://doi.org/10.1371/journal.pone.0010314>.
 60. McCutcheon JP, Moran NA. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc Natl Acad Sci U S A* 104:19392–19397. <https://doi.org/10.1073/pnas.0708855104>.
 61. Chang H-H, Cho S-T, Canale MC, Mugford ST, Lopes JRS, Hogenhout SA, Kuo C-H. 2015. Complete genome sequence of “*Candidatus* Sulcia muelleri” ML, an obligate nutritional symbiont of maize leafhopper (*Dalbulus maidis*). *Genome Announc* 3:e01483-14. <https://doi.org/10.1128/genomeA.01483-14>.
 62. Koga R, Moran NA. 2014. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *ISME J* 8:1237–1246. <https://doi.org/10.1038/ismej.2013.235>.
 63. Bennett GM, Abbà S, Kube M, Marzachi C. 2016. Complete genome sequences of the obligate symbionts “*Candidatus* Sulcia muelleri” and “*Ca. Nasuia deltocephalinicola*” from the pestiferous leafhopper *Macrostelus quadripunctulatus* (Hemiptera: Cicadellidae). *Genome Announc* 4:e01604-15. <https://doi.org/10.1128/genomeA.01604-15>.
 64. McCutcheon JP, McDonald BR, Moran NA. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci U S A* 106:15394–15399. <https://doi.org/10.1073/pnas.0906424106>.
 65. Bennett GM, Mao M. 2018. Comparative genomics of a quadripartite symbiosis in a planthopper host reveals the origins and rearranged nutritional responsibilities of anciently diverged bacterial lineages. *Environ Microbiol* 20:4461–4472. <https://doi.org/10.1111/1462-2920.14367>.
 66. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* 32:402–407. <https://doi.org/10.1038/ng986>.
 67. Rio RVM, Symula RE, Wang J, Lohs C, Wu Y, Snyder AK, Bjornson RD, Oshima K, Biehler BS, Perna NT, Hattori M, Aksoy S. 2012. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. *mBio* 3:e00240-11. <https://doi.org/10.1128/mBio.00240-11>.
 68. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubio JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li H-M, Meyer JM, Johnson RM, Romero-Severson J, VanZee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci U S A* 107:12168–12173. <https://doi.org/10.1073/pnas.1003379107>.
 69. Price DR, Wilson AC. 2014. A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol* 12:110. <https://doi.org/10.1186/s12915-014-0110-4>.
 70. González-Domenech CM, Belda E, Patiño-Navarrete R, Moya A, Peretó J, Latorre A. 2012. Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two *Blattabacterium cuenoti* strains, primary endosymbionts of cockroaches. *BMC Microbiol* 12:S5. <https://doi.org/10.1186/1471-2180-12-S1-S5>.
 71. Ponce-de-León M, Montero F, Peretó J. 2013. Solving gap metabolites and blocked reactions in genome-scale models: application to the metabolic network of *Blattabacterium cuenoti*. *BMC Syst Biol* 7:114. <https://doi.org/10.1186/1752-0509-7-114>.
 72. Patino-Navarrete R, Piulachs M-D, Belles X, Moya A, Latorre A, Peretó J. 2014. The cockroach *Blattella germanica* obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. *Biol Lett* 10:20140407. <https://doi.org/10.1098/rsbl.2014.0407>.
 73. Zhang H, Yoshizawa S, Sun Y, Huang Y, Chu X, González JM, Pinhasi J, Luo H. 2019. Repeated evolutionary transitions of flavobacteria from marine to non-marine habitats. *Environ Microbiol* 21:648–666. <https://doi.org/10.1111/1462-2920.14509>.
 74. Wakayama EJ, Dillwith JW, Howard RW, Blomquist GJ. 1984. Vitamin B12 levels in selected insects. *Insect Biochem* 14:175–179. [https://doi.org/10.1016/0020-1790\(84\)90027-1](https://doi.org/10.1016/0020-1790(84)90027-1).
 75. Schmidt A, Call L-M, Macheiner L, Mayer HK. 2019. Determination of vitamin B12 in four edible insect species by immunoaffinity and ultra-high performance liquid chromatography. *Food Chem* 281:124–129. <https://doi.org/10.1016/j.foodchem.2018.12.039>.
 76. Brzuszkiewicz E, Waschkowitz T, Wiewer A, Daniel R. 2012. Complete genome sequence of the B12-producing *Shimwellia blattae* Strain DSM 4481, isolated from a cockroach. *J Bacteriol* 194:4436–4436. <https://doi.org/10.1128/JB.00829-12>.
 77. Zhang Y, Rodionov DA, Gelfand MS, Gladyshev VN. 2009. Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization. *BMC Genomics* 10:78. <https://doi.org/10.1186/1471-2164-10-78>.
 78. Buchner P. 1965. Endosymbiosis of animals with plant microorganisms. John Wiley Sons, Hoboken, New Jersey.
 79. Koga R, Bennett GM, Cryan JR, Moran NA. 2013. Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. *Environ Microbiol* 15:2073–2081. <https://doi.org/10.1111/1462-2920.12121>.
 80. Mao M, Yang X, Bennett GM. 2018. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proc Natl Acad Sci U S A* 115:E11691–E11700.
 81. Potrikus CJ, Breznak JA. 1981. Gut bacteria recycle uric acid nitrogen in termites: a strategy for nutrient conservation. *Proc Natl Acad Sci U S A* 78:4601–4605. <https://doi.org/10.1073/pnas.78.7.4601>.
 82. Li S, Zhu S, Jia Q, Yuan D, Ren C, Li K, Liu S, Cui Y, Zhao H, Cao Y, Fang G, Li D, Zhao X, Zhang J, Yue Q, Fan Y, Yu X, Feng Q, Zhan S. 2018. The genomic and functional landscapes of developmental plasticity in the American cockroach. *Nat Commun* 9:1008. <https://doi.org/10.1038/s41467-018-03281-1>.
 83. Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, Chen Z, Childers CP, Glastad KM, Gokhale K, Gowin J, Gronenberg W, Hermansen RA, Hu H, Hunt BG, Huylmans AK, Khalil SM, Mitchell RD, Munoz-Torres MC, Mustard JA, Pan H, Reese JT, Scharf ME, Sun F, Vogel H, Xiao J, Yang W, Yang Z, Yang Z, Zhou J, Zhu J, Brent CS, Elsik CG, Goodisman MA, Liberles DA, Roe RM, Vargo EL, Vilcinskis A, Wang J, Bornberg-Bauer E, Korb J, Zhang G, Liebig J. 2014. Molecular traces of alternative social organization in a termite genome. *Nat Commun* 5:3636. <https://doi.org/10.1038/ncomms4636>.
 84. Hayashi Y, Maekawa K, Nalepa CA, Miura T, Shigenobu S. 2017. Transcriptome sequencing and estimation of DNA methylation level in the subsocial wood-feeding cockroach *Cryptocercus punctulatus* (Blattodea: Cryptocercidae). *Appl Entomol Zool* 52:643–651. <https://doi.org/10.1007/s13355-017-0519-7>.
 85. Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S, Gowin J, Greiner C, Han Y, Hu H, Hughes DST, Huylmans AK, Kemena C, Kremer LPM, Lee SL, Lopez-Ezquerria A, Mallet L, Monroy-Kuhn JM, Moser A, Murali SC, Muzny DM, Otani S, Piulachs MD, Poelchou M, Qu J, Schaub F, Wada-Katsumata A, Worley KC, Xie Q, Ylla G, Poulsen M, Gibbs RA, Schal C, Richards S, Belles X, Korb J, Bornberg-Bauer E. 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol* 2:557–566. <https://doi.org/10.1038/s41559-017-0459-1>.
 86. Seppy M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness, p 227–245. In Kollmar M (ed), *Gene Prediction: Methods and Protocols*. Springer, New York, NY.
 87. Blankenburg S, Balfanz S, Hayashi Y, Shigenobu S, Miura T, Baumann O, Baumann A, Blenau W. 2015. Cockroach GABAB receptor subtypes: molecular characterization, pharmacological properties and tissue distribution. *Neuropharmacology* 88:134–144. <https://doi.org/10.1016/j.neuropharm.2014.08.022>.
 88. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124. <https://doi.org/10.1186/1471-2105-12-124>.

89. Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28:1033–1034. <https://doi.org/10.1093/bioinformatics/bts079>.
90. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33:511–518. <https://doi.org/10.1093/nar/gki198>.
91. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>.
92. Kück P, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool* 11:81. <https://doi.org/10.1186/s12983-014-0081-x>.
93. Hrdy I, Hirt RP, Dolezal P, Bardónová L, Foster PG, Tachezy J, Martin Embley T. 2004. Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* 432:618–622. <https://doi.org/10.1038/nature03149>.
94. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
95. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res* 47:D807–D811. <https://doi.org/10.1093/nar/gky1053>.
96. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. <https://doi.org/10.7717/peerj.2584>.
97. Jones CE, Brown AL, Baumann U. 2007. Estimating the annotation error rate of curated GO database sequence annotations. *BMC Bioinformatics* 8:170. <https://doi.org/10.1186/1471-2105-8-170>.
98. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
99. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:D200–D203. <https://doi.org/10.1093/nar/gkw1129>.
100. Camon EB, Barrell DG, Dimmer EC, Lee V, Magrane M, Maslen J, Binns D, Apweiler R. 2005. An evaluation of GO annotation retrieval for BioCreAtIvE and GOA. *BMC Bioinformatics* 6:S1–S17. <https://doi.org/10.1186/1471-2105-6-S1-S17>.
101. De Ferrari L, Aitken S, van Hemert J, Goryanin I. 2012. EnzML: multi-label prediction of enzyme classes using InterPro signatures. *BMC Bioinformatics* 13:61. <https://doi.org/10.1186/1471-2105-13-61>.
102. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36. <https://doi.org/10.1093/nar/28.1.33>.
103. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res* 48:D445–D453. <https://doi.org/10.1093/nar/gkz862>.
104. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D. 2019. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res* 47:D542–D549. <https://doi.org/10.1093/nar/gky1048>.
105. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. 1999. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96:4285–4288. <https://doi.org/10.1073/pnas.96.8.4285>.
106. Osterman A, Overbeek R. 2003. Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol* 7:238–251. [https://doi.org/10.1016/s1367-5931\(03\)00027-9](https://doi.org/10.1016/s1367-5931(03)00027-9).
107. Chen L, Vitkup D. 2006. Predicting genes for orphan metabolic activities using phylogenetic profiles. *Genome Biol* 7:R17. <https://doi.org/10.1186/gb-2006-7-2-r17>.
108. Hernández-Montes G, Díaz-Mejía JJ, Pérez-Rueda E, Segovia L. 2008. The hidden universal distribution of amino acid biosynthetic networks: a genomic perspective on their origins and evolution. *Genome Biol* 9:R95. <https://doi.org/10.1186/gb-2008-9-6-r95>.
109. Reimer LC, Vetcinova A, Carbasse JS, Söhngen C, Gleim D, Ebeling C, Overmann J. 2019. BacDive in 2019: bacterial phenotypic data for high-throughput biodiversity analysis. *Nucleic Acids Res* 47:D631–D636. <https://doi.org/10.1093/nar/gky879>.