Parallel and Gradual Genome Erosion in the Blattabacterium Endosymbionts of Mastotermes darwiniensis and Cryptocercus Wood Roaches

Parallel and Gradual Genome Erosion in the *Blattabacterium* Endosymbionts of *Mastotermes darwiniensis* and *Cryptocercus* Wood Roaches

Yukihiro Kinjo[1,2,3,4], Thomas Bourguignon[3,5], Kwei Jun Tong[6], Hirokazu Kuwahara[2], Sang Jin Lim[7], Kwang Bae Yoon[7], Shuji Shigenobu[8], Yung Chul Park[7], Christine A. Nalepa[9], Yuichi Hongoh[1,2], Moriya Ohkuma[1], Nathan Lo[6,*], and Gaku Tokuda[4,*]

[1]Japan Collection of Microorganisms, RIKEN BioResource Research Center, Tsukuba, Japan

[2]Department of Life Science and Technology, Tokyo Institute of Technology, Tokyo, Japan

[3]Okinawa Institute of Science and Technology, Graduate University, Okinawa, Japan

[4]Tropical Biosphere Research Center, University of the Ryukyus, Okinawa, Japan

[5]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic

[6]School of Life and Environmental Sciences, University of Sydney, NSW, Australia

[7]Division of Forest Science, Kangwon National University, Chuncheon, Republic of Korea

[8]National Institute for Basic Biology, NIBB Core Research Facilities, Okazaki, Japan

[9]Department of Entomology, North Carolina State University, Raleigh, North Carolina, USA

*Corresponding authors: E-mails: nathan.lo@sydney.edu.au; tokuda@comb.u-ryukyu.ac.jp.

Accepted: May 29, 2018

Data deposition: This project has been deposited in the International Nucleotide Sequence Database (GenBank/ENA/DDBJ) under the accession numbers given in Table 1.

Abstract

Almost all examined cockroaches harbor an obligate intracellular endosymbiont, *Blattabacterium cuenoti*. On the basis of genome content, *Blattabacterium* has been inferred to recycle nitrogen wastes and provide amino acids and cofactors for its hosts. Most *Blattabacterium* strains sequenced to date harbor a genome of ~630 kbp, with the exception of the termite *Mastotermes darwiniensis* (~590 kbp) and *Cryptocercus punctulatus* (~614 kbp), a representative of the sister group of termites. Such genome reduction may have led to the ultimate loss of *Blattabacterium* in all termites other than *Mastotermes*. In this study, we sequenced 11 new *Blattabacterium* genomes from three species of *Cryptocercus* in order to shed light on the genomic evolution of *Blattabacterium* in termites and *Cryptocercus*. All genomes of *Cryptocercus*-derived *Blattabacterium* genomes were reduced (~614 kbp), except for that associated with *Cryptocercus kyebangensis*, which comprised 637 kbp. Phylogenetic analysis of these genomes and their content indicates that *Blattabacterium* experienced parallel genome reduction in *Mastotermes* and *Cryptocercus*, possibly due to similar selective forces. We found evidence of ongoing genome reduction in *Blattabacterium* from three lineages of the *C. punctulatus* species complex, which independently lost one cysteine biosynthetic gene. We also sequenced the genome of the *Blattabacterium* associated with *Salganea taiwanensis*, a subsocial xylophagous cockroach that does not vertically transmit gut symbionts via proctodeal trophallaxis. This genome was 632 kbp, typical of that of nonsubsocial cockroaches. Overall, our results show that genome reduction occurred on multiple occasions in *Blattabacterium*, and is still ongoing, possibly because of new associations with gut symbionts in some lineages.

Key words: *Blattabacterium*, genome reduction, *Cryptocercus*, proctodeal trophallaxis.

Introduction

Cockroaches and *Mastotermes darwiniensis*, the most primitive termite, harbor the endosymbiotic bacteria *Blattabacterium cuenoti* (hereafter *Blattabacterium*) in their fat bodies. *Blattabacterium* is transovarially transmitted between host generations, and is essential to host growth and reproduction (Brooks and Richards 1955; Brooks 1970). On the basis of whole genome sequencing of various strains, *Blattabacterium* has been inferred to provide essential amino acids (hereafter EAA) and vitamins for its host, and participate in the recycling of nitrogen wastes (López-Sánchez et al. 2009; Sabree et al. 2009). The symbiotic relationship between *Blattabacterium* and cockroaches is believed to have been established >235 Ma (Bourguignon et al. 2018), and has been maintained via strict vertical transmission since then (Lo et al. 2003).

Although essential to most cockroaches, *Blattabacterium* symbionts were lost in all termites (which are a form of derived social cockroach; Lo et al. 2000) except *Mastotermes darwiniensis* (Bandi et al. 1995; Lo et al. 2003). The sister group of termites is the rotten wood-feeding and subsocial cockroach genus *Cryptocercus*. By studying the *Blattabacterium* strains of *M. darwiniensis* and *Cryptocercus*, insights into the factors leading to the loss of this symbiont from all other termites can be obtained. *Blattabacterium* may also have been lost in the enigmatic cave roach genus *Nocticola* (Lo et al. 2007), although the uncertain phylogenetic position of this taxon among other cockroaches (Bourguignon et al. 2018) has made it difficult to test this hypothesis. If the ancestors of *Nocticola* did indeed lose *Blattabacterium*, the reasons for this loss cannot easily be investigated because *Nocticola* has no known close relatives.

Most *Blattabacterium* strains sequenced to date harbor a genome of ~630 kbp, with the exception of MADAR from *M. darwiniensis* (~590 kbp; Sabree et al. 2012) and CPU from *Cryptocercus punctulatus* (~614 kbp; Neef et al. 2011). The causes of the increased levels of genome degradation in these strains are not clear. One common cause of genome reduction in endosymbionts is the association of the host with a new endosymbiotic partner (e.g., Husnik et al. 2013). Although *Wolbachia* has been found in some species (Vaishampayan et al. 2007), obligate intracellular nutritional mutualists other than *Blattabacterium* are not known from cockroaches, and do not appear to be responsible for the genome reduction found in CPU and MADAR.

Many of the genes missing in CPU and MADAR, but present in other *Blattabacterium* strains, are associated with EAA synthesis. An increase in available EAAs in the diets of their ancestral hosts may have led to the loss of these genes in CPU and MADAR, due to relaxed selection. One source of such EAAs could have been microbes associated with rotting wood, which could then be digested and taken up by the host in the midgut (Neef et al. 2011). To test this hypothesis, Tokuda et al., (2013) sequenced the *Blattabacterium* genome of *Panesthia angustipennis*, another rotten-wood feeding cockroach. However, this genome was found to encode all the EAA biosynthesis genes found in most *Blattabacterium* strains. Alternative hypotheses to explain the loss of EAA pathways in MADAR and CPU include: 1) increased EAA levels in the diets of ancestral hosts due to behaviors associated with subsociality and eusociality, such as proctodeal trophallaxis (Fujita et al. 2001; Tokuda et al. 2014); 2) the presence of symbionts in the guts of the ancestors of *M. darwiniensis* and *C. punctulatus* which were able to provision the host with EAAs. For example, cellulolytic protists in the guts of *Cryptocercus* and *M. darwiniensis* host bacterial endosymbionts that fix nitrogen, produce essential amino acids and participate in the nitrogen metabolism of their protist hosts (Hongoh 2010; Ohkuma et al. 2015). These two hypotheses are not mutually exclusive.

Many genes, particularly those associated with EAA biosynthesis, are absent in both CPU and MADAR, suggesting they were lost in the common ancestor of these species (Patiño-Navarrete et al. 2013). Nonetheless, several genes are absent in only one of the strains, indicating independent gene loss. For example, MADAR retains *cysE* and *cysK* genes, but CPU has lost both these genes. On the contrary, *metB* and *lysA* are pseudogenised in CPU but functional in MADAR. Because many genes were lost independently by MADAR and CPU, the possibility remains that some genes in EAA pathways have been lost independently by both lineages, and not by their common ancestor as currently believed.

To test between the hypotheses of gene loss in a common ancestor of CPU and MADAR versus independent gene loss in each of these lineages, we sequenced additional *Blattabacterium* genomes from one host sample of *Cryptocercus kyebangensis*, one specimen of *C. clevelandi* and nine specimens of *C. punctulatus*. We also investigated the influence of subsocial behavior and wood-feeding on *Blattabacterium* genome evolution by sequencing the strain from one specimen of *Salganea taiwanensis*, a wood-feeding subsocial cockroach which does not exhibit proctodeal trophallaxis.

Materials and Methods

Sample Collection and DNA Extraction

Locations of the sample collection for *Cryptocercus* cockroaches are indicated in figure 1 (details are indicated in supplementary table S1, Supplementary Material online). Although four species of *Cryptocercus* cockroaches have been described from the Appalachian Mountains, (Burnside et al. 1999), we refer to all of them as *Cryptocercus punctulatus* species complex, and distinguish among them by their chromosome numbers. *Salganea taiwanensis* were collected in Iriomote-island, Okinawa, Japan. We extracted DNA from

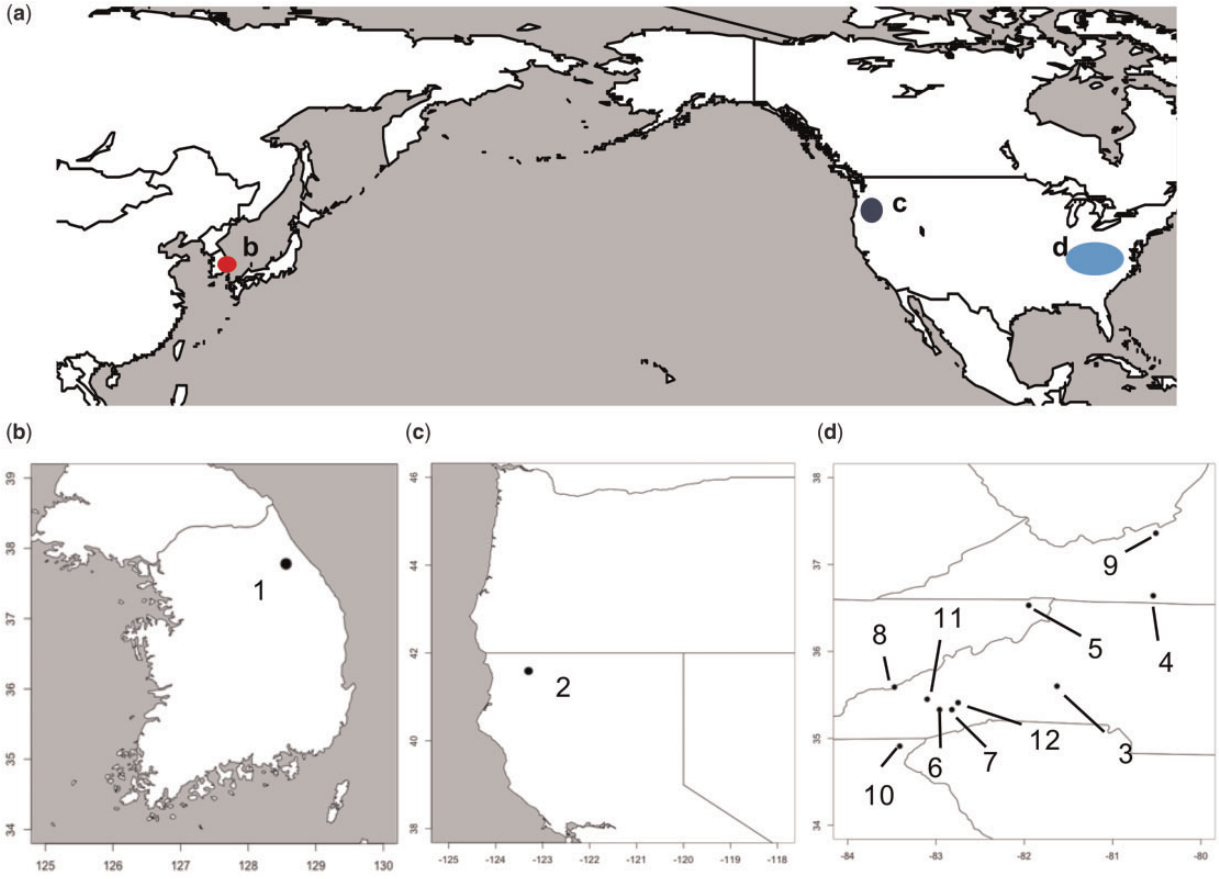

Fig. 1.—Sampling locations of *Cryptocercus* cockroaches. Detailed information for these locations are provided in supplementary table S1, Supplementary Material online.

the fat bodies of a single individual. DNA extraction was done by using ISOPLANT II DNA extraction Kit (NIPPON GENE, Tokyo) under the manufacturer's instructions.

Genome Sequencing and Assembly

Libraries were prepared from fat body total DNA using the TruSeq DNA Sample Preparation Kit according to the manufacturer's protocol. Libraries were then mixed in equimolar concentration and sequenced with Illumina MiSeq or HiSeq2000 sequencing system. Reads were quality checked, trimmed and filtered using FaQCs (Lo and Chain 2014). High quality reads were assembled with the "TCSF and IMRA" pipeline (Kinjo et al. 2015). Gaps between contigs were filled using GapFiller (Nadalin et al. 2012). All final assemblies consisted of a single circular chromosome. The quality of the final assembly was evaluated using REAPER (Hunt et al. 2013), and no assembly errors were detected. Screening for potential secondary symbionts in the sequence libraries for each species was performed using blastx, implemented in the DIAMOND software package (Buchfink et al. 2015). All assembled contigs were subject to blastx searches against a UniRef90 database (Suzek et al. 2015) with a 1e-30 *E*-value threshold. All assembled genome data were deposited in the International Nucleotide Sequence Database (GenBank/ENA/DDBJ) under the accession numbers given in table 1.

Genome Annotation

Prediction of protein coding regions was carried out using Prodigal (Hyatt et al. 2010) with a 0.6 score cutoff. In addition to the Prodigal prediction, we also carried out homology-based ORF prediction using blastp search, implemented in the BLAST+ package (Camacho et al. 2009), against the Swiss-prot database (released in March 2017). Predictions for rRNA, tRNA, and other noncoding RNAs were done by using RNAmmer (Lagesen et al. 2007), tRNAscan-SE (Lowe and Eddy 1997), and Infernal (Nawrocki and Eddy 2013), respectively. Functional annotation of predicted coding sequences was done by blastp search against the COG database (Galperin et al. 2015) with curation using CD-search (Marchler-Bauer and Bryant 2004).

Phylogenetic Tree Inference

We determined a set of orthologous genes shared by all genomes used in this study using Proteinortho ver. 5

Table 1
Genome Characteristics of All Sequenced *Blattabacterium* Strains

Organism (host scientific name) Strain	Plsmd.	Size (Kb)	G + C%	CDS	rRNA	tRNA	ncRNA	Pseudogene	Accession Number	Site No.[b]
Blattabacterium sp. (*Panesthia angustipennis spadica*) str. BPAA	0[a]	632	26.4	578	3	34	3	0	NC_020510.1	—
Blattabacterium sp. (*Pa. angustipennis yaeyamensis*) str. BPAY	0[a]	632	26.3	577	3	34	3	0	NZ_AP014609.1	—
Blattabacterium sp. (*Salganea taiwanensis taiwanensis*) str. **STAT**	0[a]	632	24.8	575	3	33	2	0	AP014608	—
Blattabacterium sp. (*Nauphoeta cinerea)* str. BNCIN	1	627	26.1	568	3	34	2	0	NC_022550.1-NC_022551.1	—
Blattabacterium sp. (*Blaberus giganteus*) str. BGIGA	1	633	25.7	577	3	34	2	1	NC_017924.1-NC_017925.1	—
Blattabacterium sp. (*Blattella germanica*) str. BGE	1	641	27.1	591	3	34	3	0	NC_013454.1-NC_015679.1	—
Blattabacterium sp. (*Periplaneta americana*) str. BPLAN	1	640	28.2	589	3	33	3	5	NC_013418.2-NC_013419.1	—
Blattabacterium sp. (*Blatta orientalis*) str. BOR	1	638	28.2	576	3	34	3	13	NC_020195.1-NC_020196.1	—
Blattabacterium sp. (*Cryptocercus kyebangensis*) str. **CKYod**	1	637	25.7	571	3	32	2	4	CP029820-CP029821	1
Blattabacterium sp. (*C. cleveladi*) str. **CCLhc**	1	621	24.5	551	3	32	2	6	CP029844-CP029845	2
Blattabacterium sp. (C. punctulatus species complex: *2n* = 43*)* str. **CPUsm**	0[a]	614	23.8	548	3	32	2	7	CP029810	3
Blattabacterium sp. (C. punctulatus species complex: *2n* = 43) str. **CPUpc**	0[a]	614	23.8	544	3	32	2	5	CP029811	4
Blattabacterium sp. (C. punctulatus species complex: *2n* = 43*)* str. **CPUsv**	0[a]	614	23.8	544	3	32	2	5	CP029812	5
Blattabacterium sp. (*C. punctulatus* species complex: *2n* = 45) str. **CPUbt**	0[a]	613	23.9	546	3	32	2	3	CP029813	6
Blattabacterium sp. (C. punctulatus species complex: *2n* = 45*)* str. **CPUmp**	0[a]	613	23.8	545	3	32	2	3	CP029814	7
Blattabacterium sp. (*C. punctulatus* species complex: *2n* = 45) str. **CPUmc**	0[a]	613	23.9	546	3	32	2	4	CP029815	8
Blattabacterium sp. (*C. punctulatus* species complex: *2n* = 43) str. **CPUml**	1	616	24.1	548	3	32	2	2	AP014610	9
Blattabacterium sp. (C. punctulatus species complex: *2n* = 39*)* str. **CPUbr**	1	609	23.8	544	3	32	2	5	CP029816-CP029817	10

(continued)

Table 1 Continued

Organism (host scientific name) Strain	Plsmd.	Size (Kb)	G + C%	CDS	rRNA	tRNA	ncRNA	Pseudogene	Accession Number	Site No.[b]
Blattabacterium sp. (*C. punctulatus* species complex: *2n* = 39) str. **CPUwf**	1	611	23.8	546	3	32	2	4	CP029818-CP029819	11
Blattabacterium sp. (*C. punctulatus* species complex: *2n* = 37) str. CPU	1	610	23.9	547	3	32	3	2	NC_016621.1-NC_016598.1	12
Blattabacterium sp. (*Mastotermes darwiniensis*) str. MADAR	1	590	27.5	547	3	34	3	1	NC_016146.1-NC_016150.1	—

[a]A plasmid is integrated into the chromosome
[b]Numbers correspond to the sampling locations shown in figure 1.

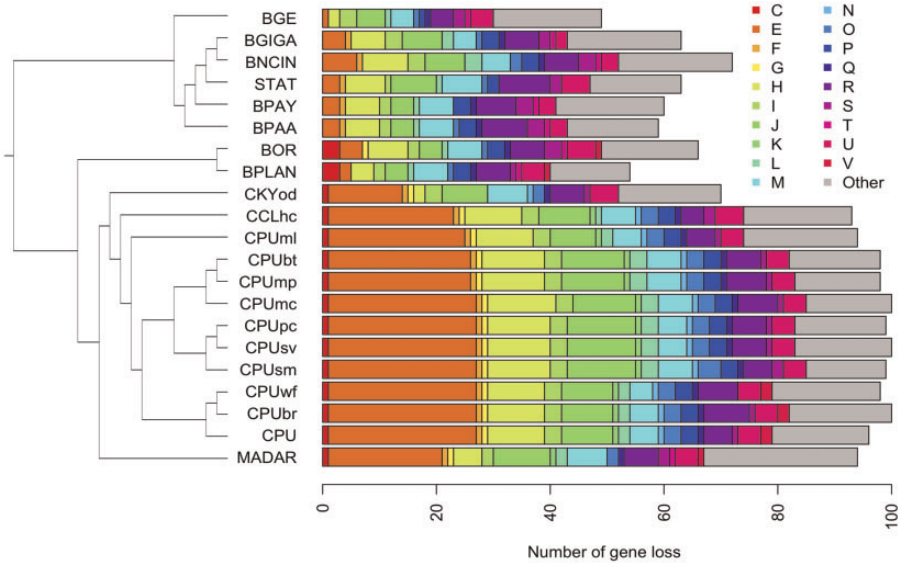

Fig. 2.—Number of gene losses from the pan-genome of *Blattabacterium* strains. Number of gene losses in each strain was calculated by comparing with constructed pan-genome of all *Blattabacterium* strains used in this study. Singletons in each genome which did not match any reference sequences in the COG database by Reverse PSI-BLAST search were removed from the pan-genome data set.

(Lechner et al. 2011) and AMPHORA2 (Wu and Scott 2012). All orthologous genes were aligned with MAFFT (Katoh and Standley 2013). Aligned amino acid sequences were trimmed with Gblocks (Castresana 2000) run under default parameters, and back translated into codon sequence by using Pal2nal (Suyama et al. 2006). All codon alignments were then concatenated with FAST (Lawrence et al. 2015). A phylogenetic tree was reconstructed using the maximum likelihood algorithm implemented in RAxML v8.2 (Stamatakis 2014), with the GTRCAT model of nucleotide substitution.

Results and Discussion

We obtained the complete genome sequences of *Blattabacterium* strains associated with one specimen of *C. kyebangensis*, one specimen of *C. clevelandi* and nine specimens of *C. punctulatus*. The content of each *Blattabacterium* genome sequenced to date is summarized in table 1. The genome sizes of the *Blattabacterium* strains associated with *Cryptocercus* spp. ranges from 609 to 637 kbp. These differences are largely due to the loss of EAA biosynthesis genes in some genomes (figs. 2 and 3). The topology recovered in our phylogenetic analysis of *Blattabacterium* concurs with previous phylogenetic inferences of the endosymbiont derived from the *C. punctulatus* species complex (Che et al. 2016). CPU *Blattabacterium* strains are divided into 2 clades (A and B in supplementary fig. S1, Supplementary Material online), except for CPUml. The plasmid of the endosymbiont was integrated in the chromosome in all strains belonging to clade B.

Although MADAR and CPU strains generally have similar amino acid biosynthesis gene repertoires, some genes are

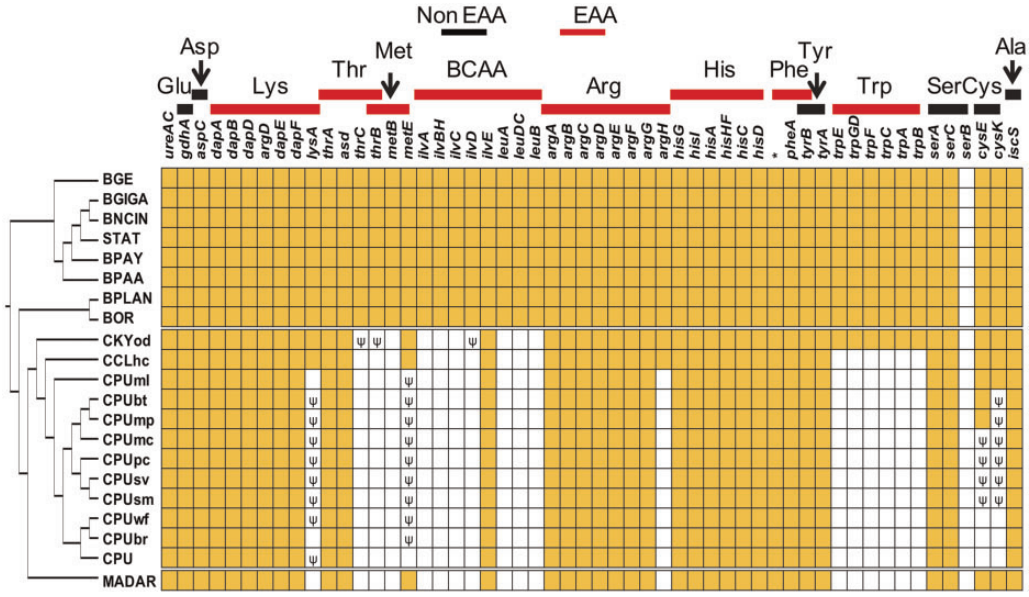

Fig. 3.—Comparison of genes involved in amino acid biosynthesis in genomes of *Blattabacterium* strains. Yellow and white boxes represent presence or absence of genes, respectively. A pseudogene is denoted with a psi (ψ) symbol. Red horizontal bars indicate biosynthetic pathways of essential amino acids (EAAs), whereas black bars represent non-essential amino acid (non-EAA) pathways. An asterisk in the phenylalanine biosynthesis pathway corresponds to chorismate mutase (EC 5.4.99.5). Amino acids that are not shown here implicate that the entire synthesis pathway is absent in *Blattabacterium*. Gly and Pro are abundant in the hemolymph of cockroaches and thus not considered to provision the host (Patiño-Navarrete et al. 2014). BCAA, blanched chain amino acids (i.e., Ile, Leu, and Val). See table 1 for host species of each strain.

present only in some of the *Cryptocercus*-derived strains (fig. 3), namely those of *C. kyebangensis* (CKYod) and *C. clevelandi* (CCLhc). For example, genes involved in tryptophan synthesis are found in CKYod only, and fully intact copies of *lysA* and *argH* are only found in CKYod and CCLhc. This shows that these genes have been lost independently (or have become putative pseudogenes) in MADAR and in strains of CPU.

Only one gene involved in methionine synthesis, *metB*, and a few genes involved in BCAA synthesis (*ilvA, ilvBH*, *ilvC*, and *ilvD*), are absent in all *Blattabacterium* strains associated with *Mastotermes* and *Cryptocercus* (fig. 3). However, some of the genes involved in the methionine and BCAA synthesis pathways, such as *metE* or *ilvE*, are still present in the genomes of MADAR and/or *Cryptocercus*-derived strains, either intact or as putative pseudogenes. Independent losses of methionine and BCAA synthesis in the *Blattabacterium* of *Mastotermes* and *Cryptocercus* cannot, therefore, be ruled out.

The *Blattabacterium* genome content varies between CPU strains of the *C. punctulatus* species complex (Nalepa et al. 2002; Everaerts et al. 2008). For example, one strain (CPUml) retains the entire gene set involved in cysteine synthesis (i.e., *cysE* and *cysK*), but one or both these have been lost or occur as putative pseudogenes in the other nine strains (fig. 3). The phylogeny of *C. punctulatus* suggests that the *cysE* gene was lost or pseudogenized independently three times (fig. 4).

We sequenced the endosymbiont genome of another wood-feeding cockroach, *Salganea taiwanensis*, a subsocial insect unrelated to termites and *Cryptocercus*. Although subsocial, *Salganea taiwanensis* does not exhibit proctodeal trophallaxis (Maekawa et al. 2008). We found that its genome is 632 kbp with a GC content of 24.8%. The genome comprises 575 coding sequences (CDSs), a single rRNA operon, 32 tRNAs, and three other noncoding RNAs. The plasmid is integrated into the genome, similar to the case of related *Panesthia* cockroaches. Overall, genomic characteristics differ from the strains derived from *Mastotermes* and *Cryptocercus*, and are highly similar to the strains of other cockroaches, including the strains of the related *Panesthia* species (supplementary fig. S2*A*, Supplementary Material online). This includes the presence of all amino acid biosynthetic genes typically present in *Blattabacterium* genomes (fig. 3).

Our results show that the genomes of *Blattabacterium* independently underwent erosion in the lineages leading to extant *Cryptocercus* and *Mastotermes*. Among *Cryptocercus* strains, *Blattabacterium* genome erosion was gradual, and comparison of gene loss events with previous chronograms estimated for *Cryptocercus* (Che et al. 2016) indicates that these events took place sequentially over the last 60 My (fig. 4). We also found evidence of further gene loss during the last 5 My in the *Blattabacterium* genome of the *C. punctulatus* species complex (CPU strains). In the earliest branching *C. punctulatus* lineage, *Blattabacterium* CPUml possesses functional *cysE* and *cysK*, and is presumably able to synthesize cysteine. In contrast, three other lineages of the *C. punctulatus* species complex harbor endosymbionts with almost

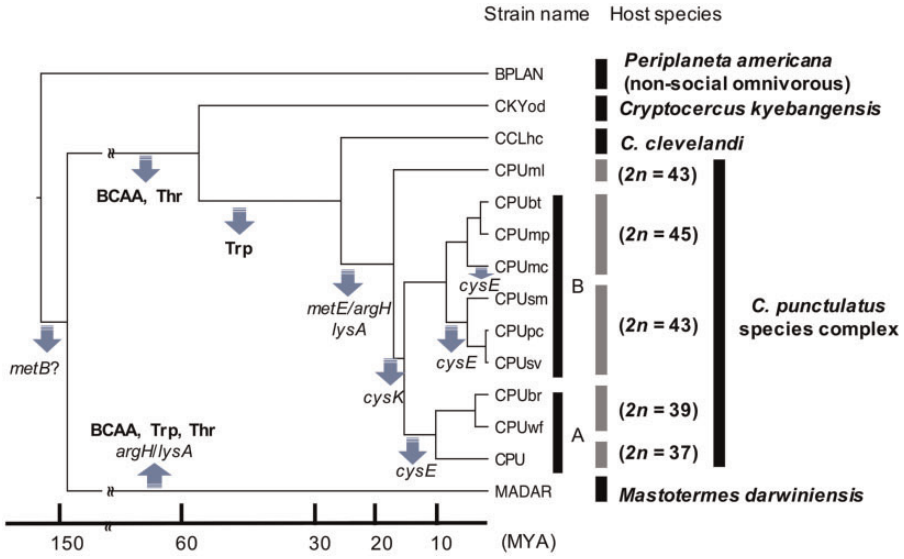

Fig. 4.—Estimated evolutionary history of gene losses in *Blattabacterium* strains. Blue arrows represent gene loss events related to amino acid biosynthesis. Deficient amino acids or genes caused by each gene loss event are shown with arrow. Time scale is based on a previous estimate by Che et al. (2016).

identical gene repertoires, all of which lack functional *cysK* and/or *cysE*. The phylogenetic tree of the *C. punctulatus* species complex suggests that the loss of *cysE* occurred independently in three lineages. *Blattabacterium* genome reduction is therefore ongoing among members of the *C. punctulatus* species complex, and involves the same set of genes.

Primary endosymbionts often experience genome instability subsequent to the establishment of new symbiotic associations between their host and a secondary obligate (or co-obligate) symbionts. Genes with redundant functions in the primary and secondary symbiont genomes are easily lost because of relaxed selective pressures. As a result, the two genomes evolve to complement each other, and become mutually dependent. For example, in the aphid lineages associated with the secondary obligate symbiont *Serratia symbiotica*, the aphid primary endosymbiont *Buchnera aphidicola* underwent massive gene losses, having a genome size of only 425–453 kbp (Manzano-Marín et al. 2016). Secondarily acquired endosymbionts can trigger even more extreme genome reduction, such as in *Candidatus* Sulcia muelleri, *Candidatus* Carsonella rudii, and *Candidatus* Portiera aleyrodidarum, primary endosymbionts of cicada, psyllid and whitefly, respectively, which have genomes varying in size between 114 and 245 kbp (McCutcheon et al. 2009; Sloan and Moran 2012; Rao et al. 2015). Although these genomes lack many genes involved in various functions, they retain most genes involved in EAAs biosynthesis.

The genomes of MADAR and CPU are between 30 and 50 kbp smaller than that of other strains of *Blattabacterium*, and most of the genes they lost were involved in EAA biosynthesis. This markedly contrasts with the 200 kbp of genes lost by the *Buchnera* strains whose host aphids secondarily acquired *Serratia* endosymbionts. The reason why *Blattabacterium* mostly lost EAA biosynthesis genes, but retained genes with other functions, is unclear. One possible explanation is the presence of secondary intracellular symbionts in *Cryptocercus* and termites. Such secondary symbionts typically comprise a single, or a few, microbial species, all localized in bacteriocytes, allowing metabolic collaborations with primary symbionts (McCutcheon et al. 2009; Sloan and Moran 2012; Rao et al. 2015). We searched for secondary symbionts in *Cryptocercus* but found no evidence of their presence. No assembled contig from *Cryptocercus* fat body sequence libraries in this study was found to have a reliable blastx matches with bacterial protein sequences from the Uniref90 database. Although a few small (<2 Knt) contigs were found to have matches with bacterial protein sequence, each of these had low depth (>50 fold lower than those of *Blattabacterium* contigs), and were not shared among libraries. We therefore conclude that these contigs are most likely derived from environmental or intestinal contaminants present during cockroach dissection and DNA extraction.

An alternative explanation for gene loss in *Blattabacterium* is the gradual development of novel associations among intestinal symbiotic microbes, which in some way enhanced the availability of essential amino acids to their hosts. One such association known in *Cryptocercus* and lower termites, including *M. darwiniensis*, is the presence of oxymonad and hypermastigid flagellates in the guts of these insects. These flagellates themselves have bacterial ecto- and endosymbionts which contribute to nitrogen metabolism (Ohkuma 2008; Ohkuma et al. 2015; Hongoh 2010). Some symbiotic bacteria of gut flagellates are diazotrophs, having the capability of fixing atmospheric nitrogen into protein, potentially allowing their hosts access to new nitrogen sources (Tai et al. 2016). How

the production of EAAs by symbiotic microbes could be accessed by hosts is unclear. One possibility is an increase in EAA concentration in the gut lumen, followed by uptake in the rectum (Phillips et al. 1986). Alternatively, microbes could be consumed by nestmates via proctodeal trophallaxis or coprophagy (Fujita et al. 2001; Machida et al. 2001; Nalepa et al. 2001; Tokuda et al. 2014), and EAAs released and taken up in their midguts.

Unlike the genomes of MADAR and *Cryptocercus*-derived *Blattabacterium* strains, the *Blattabacterium* genome of *Salganea* has not undergone significant further reduction. *Salganea* spp. are known to engage in social behaviors, including stomodeal trophallaxis and filial coprophagy, which are expected to provide larvae with nutritional stability during early development. However, our results indicate that such behaviors do not necessarily promote genome reduction. Unlike termites and *Cryptocercus, Salganea* does not exhibit proctodeal trophallaxis. Proctodeal trophallaxis, along with the acquisition of novel gut symbionts, may therefore be a key factor in the genome reduction of *Blattabacterium*.

Conclusion

In this study, we found that the losses of the same set of EAAs biosynthesis genes by MADAR and CPU were primarily the result of parallel evolution. We also found that additional genome reduction occurred during the last 5 My in several strains present in *Cryptocercus*, highlighting the phenomenon of ongoing gene loss in *Blattabacterium*. Future studies investigating the contribution of proctodeal trophallaxis and gut microbes to cockroach and Mastotermes metabolism are required to test the hypothesis of a link between genome reduction in *Blattabacterium* on the one hand, and social behavior and associations with novel gut symbiotic partners on the other.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The insects were collected under permits GRSM-2014-SCI-0015, GRSM-2015-SCI-1243, BLRI-2009-SCI-0001, and BLRI-2015-SCI-0021 from the U.S. National Park Service, 2015_0059 from North Carolina Department of Environment and Natural Resources, 132015 from Georgia Department of Natural Resources, SC-013299 from California Department of Fish and Wildlife, and 2017-01 from Wonju Regional Environment Office in South Korea. Tatsuya Kitazume is acknowledged for Illumina sequencing at Functional Genomics Facility, NIBB Core Research Facilities. This work was supported by Japan Society for the Promotion of Science KAKENHI 17H01510, 26292177, 26117722, 18K14767, and NIBB Cooperative Research Program no. 11-723. Y.K. was supported by RIKEN Junior Research Associate Program. N.L. was supported by the Australian Research Council (FT160100463).

Literature Cited

Bandi C, et al. 1995. The establishment of intracellular symbiosis in an ancestor of cockroaches and termites. Proc R Soc B Biol Sci. 259(1356):293–299.

Bourguignon T, et al. 2018. Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. Mol Biol Evol. 35(4):970–983.

Brooks MA. 1970. Comments on the classification of intracellular symbiotes of cockroaches and a description of the species. J Invert Pathol. 16(2):249–258.

Brooks MA, Richards GA. 1955. Intracellular symbiosis in cockroaches. I. Production of aposymbiotic cockroaches. Biol Bull. 109(1):22–39.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12(1):59–60.

Burnside CA, Smith PT, Kambhampati S. 1999. Three new species of the wood roach, *Cryptocercus* (Blattodea: cryptocercidae), from the eastern United States. J Kans Entomol Soc 72:361–378.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17(4):540–552.

Che Y, et al. 2016. A global molecular phylogeny and timescale of evolution for *Cryptocercus* woodroaches. Mol Phylogenet Evol. 98:201–209.

Everaerts C, et al. 2008. The *Cryptocercus punctulatus* species complex (Dictyoptera: cryptocercidae) in the eastern United States: comparison of cuticular hydrocarbons, chromosome number, and DNA sequences. Mol Phylogenet Evol. 47(3):950–959.

Fujita A, Shimizu I, Abe T. 2001. Distribution of lysozyme and protease, and amino acid concentration in the guts of a wood-feeding termite, *Reticulitermes speratus* (Kolbe): possible digestion of symbiont bacteria transferred by trophallaxis. Physiol Entomol. 26(2):116–123.

Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43(D1):D261–D269.

Hongoh Y. 2010. Diversity and genomes of uncultured microbial symbionts in the termite gut. Biosci Biotechnol Biochem. 74(6):1145–1151.

Hunt M, et al. 2013. REAPR: a universal tool for genome assembly evaluation. Genome Biol. 14(5):R47.

Husnik F, et al. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153(7):1567–1578.

Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

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(3):208–220.

Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35(9):3100–3108.

Lawrence TJ, et al. 2015. FAST: FAST analysis of sequences toolbox. Front Genet. 6:172.
Lechner M, et al. 2011. Proteinortho: detection of (co-) orthologs in large-scale analysis. BMC Bioinformatics 12:124.
Lo C-C, Chain PSG. 2014. Rapid evaluation and quality control of next generation sequencing data with FaQCs. BMC Bioinformatics 15:366.
Lo N, et al. 2000. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. Curr Biol. 10(13):801–804.
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(6):907–913.
Lo N, et al. 2007. Cockroaches that lack *Blattabacterium endosymbionts:* the phylogenetically divergent genus *Nocticola*. Biol Lett. 3:327–330.
López-Sánchez MJ, et al. 2009. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. PLoS Genet. 5(11):e1000721.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for inproved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5):955–964.
Machida M, Kitade O, Miura T, Matsumoto T. 2001. Nitrogen recycling through proctodeal trophallaxis in the Japanese damp-wood termite *Hodotermopsis japonica* (Isoptera, Termopsidae). 48:52–56.
Maekawa K, Matsumoto T, Nalepa CA. 2008. Social biology of the wood-feeding cockroach genus *Salganea* (Dictyoptera, Blaberidae, Panesthiinae): did ovoviviparity prevent the evolution of eusociality in the lineage?. Insectes Soc. 55(2):107–114.
Manzano-Marín A, Simon JC, Latorre A. 2016. Reinventing the wheel and making it round again: evolutionary convergence in *Buchnera-Serratia* symbiotic consortia between the distantly related Lachninae aphids *Tuberolachnus salignus* and *Cinara cedri*. Genome Biol Evol. 8(5):1440–1458.
Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32(Web Server):W327–W331.
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(36):15394–15399.
Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics 13(suppl 14):S8.
Nalepa CA, Bignell DE, Bandi C. 2001. Detritivory, coprophagy, and the evolution of digestive mutualisms in Dictyoptera. Insectes Soc. 48(3):194–201.
Nalepa CA, Luykx P, Klass K, Deitz LL. 2002. Distribution of karyotypes of the *Cryptocercus punctulatus* species complex (Dictyoptera: Cryptocercidae) in the Southern Appalachians: relation to habitat and history. Ann Entomol Soc Am. 2: 276–287.
Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933–2935.
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.
Ohkuma M, et al. 2015. Acetogenesis from H_2 plus CO_2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. Proc Natl Acad Sci U S A. 112(33):10224–10230.
Ohkuma M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. Trends Microbiol. 16(7):345–352.
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(2):351–361.
Patiño-Navarrete R, et al. 2014. The cockroach *Blattella germanica* obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. Biol Lett. 10:20140407.
Phillips JE, Hanrahan J, Chamberlin M, Thomson B. 1986. Mechanisms and control of reabsorption in insect hindgut. Adv Insect Physiol. 19:329–422.
Rao Q, et al. 2015. Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*. BMC Genomics 16(1):1–13.
Sabree ZL, 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(1):204–210.
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(46):19521–19526.
Sloan DB, Moran NA. 2012. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Mol Biol Evol. 29(12):3781–3792.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34(Web Server):W609–W612.
Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31(6):926–932.
Tai V, et al. 2016. Genome evolution and nitrogen fixation in bacterial ectosymbionts of a protist inhabiting wood-feeding cockroaches. Appl Environ Microbiol. 82(15):4682–4695.
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(3):20121153.
Tokuda G, et al. 2014. Metabolomic profiling of ^{13}C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. Proc R Soc B Biol Sci. 281(1789):20140990.
Vaishampayan PA, et al. 2007. Molecular evidence and phylogenetic affiliations of *Wolbachia* in cockroaches. Mol Phylogenet Evol. 44(3):1346–1351.
Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28(7):1033–1034.

Associate editor: Esperanza Martinez-Romero