



HAL
open science

Genome reduction and horizontal gene transfer in the evolution of Endomicrobia-rise and fall of an intracellular symbiosis with termite gut flagellates

Undine S Mies, Vincent Hervé, Tom Kropp, Katja Platt, David Sillam-Dussès, Jan Šobotník, Andreas Brune

► To cite this version:

Undine S Mies, Vincent Hervé, Tom Kropp, Katja Platt, David Sillam-Dussès, et al.. Genome reduction and horizontal gene transfer in the evolution of Endomicrobia-rise and fall of an intracellular symbiosis with termite gut flagellates. *mBio*, 2024, 15 (6), <10.1128/mbio.00826-24>. <hal-04760820>

HAL Id: hal-04760820

<https://hal.inrae.fr/hal-04760820v1>

Submitted on 30 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Genome reduction and horizontal gene transfer in the evolution of Endomicrobia—rise and fall of an intracellular symbiosis with termite gut flagellates

Undine S. Mies,¹ Vincent Hervé,¹ Tom Kropp,¹ Katja Platt,¹ David Sillam-Dussès,² Jan Šobotník,^{3,4} Andreas Brune¹

AUTHOR AFFILIATIONS See affiliation list on p. 19.

ABSTRACT Bacterial endosymbionts of eukaryotic hosts typically experience massive genome reduction, but the underlying evolutionary processes are often obscured by the lack of free-living relatives. Endomicrobia, a family-level lineage of host-associated bacteria in the phylum *Elusimicrobiota* that comprises both free-living representatives and endosymbionts of termite gut flagellates, are an excellent model to study evolution of intracellular symbionts. We reconstructed 67 metagenome-assembled genomes (MAGs) of *Endomicrobiaceae* among more than 1,700 MAGs from the gut microbiota of a wide range of termites. Phylogenomic analysis confirmed a sister position of representatives from termites and ruminants, and allowed to propose eight new genera in the radiation of *Endomicrobiaceae*. Comparative genome analysis documented progressive genome erosion in the new genus *Endomicrobiellum*, which comprises all flagellate endosymbionts characterized to date. Massive gene losses were accompanied by the acquisition of new functions by horizontal gene transfer, which led to a shift from a glucose-based energy metabolism to one based on sugar phosphates. The breakdown of glycolysis and many anabolic pathways for amino acids and cofactors in several subgroups was compensated by the independent acquisition of new uptake systems, including an ATP/ADP antiporter, from other gut microbiota. The putative donors are mostly flagellate endosymbionts from other bacterial phyla, including several, hitherto unknown lineages of uncultured *Alphaproteobacteria*, documenting the importance of horizontal gene transfer in the convergent evolution of these intracellular symbioses. The loss of almost all biosynthetic capacities in some lineages of *Endomicrobiellum* suggests that their originally mutualistic relationship with flagellates is on its decline.

IMPORTANCE Unicellular eukaryotes are frequently colonized by bacterial and archaeal symbionts. A prominent example are the cellulolytic gut flagellates of termites, which harbor diverse but host-specific bacterial symbionts that occur exclusively in termite guts. One of these lineages, the so-called Endomicrobia, comprises both free-living and endosymbiotic representatives, which offers the unique opportunity to study the evolutionary processes underpinning the transition from a free-living to an intracellular lifestyle. Our results revealed a progressive gene loss in energy metabolism and biosynthetic pathways, compensated by the acquisition of new functions via horizontal gene transfer from other gut bacteria, and suggest the eventual breakdown of an initially mutualistic symbiosis. Evidence for convergent evolution of unrelated endosymbionts reflects adaptations to the intracellular environment of termite gut flagellates.

KEYWORDS termites, *Endomicrobiaceae*, Parabasalia, endosymbionts, lateral gene transfer, convergent evolution

Editor Margaret J. McFall-Ngai, University of Hawaii at Manoa, Honolulu, Hawaii, USA

Address correspondence to Andreas Brune, brune@mpi-marburg.mpg.de.

The authors declare no conflict of interest.

See the funding table on p. 19.

Received 19 March 2024

Accepted 9 April 2024

Published 14 May 2024

Copyright © 2024 Mies 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/).

Bacterial endosymbionts of eukaryotes typically experience progressive genome erosion (1–4). Apart from several well-characterized endosymbionts in insect tissues, the absence of close but free-living relatives often impedes our understanding of the evolutionary processes and mechanisms underlying the transition from a host-associated to an intracellular lifestyle. A notable exception are the so-called Endomicrobia, which represent a family-level clade in the phylum *Elusimicrobiota* and have been detected almost exclusively in the intestinal tract of insects (5–7) and ruminants (7–9). The family *Endomicrobiaceae* comprises both free-living gut commensals of their respective hosts and closely related lineages of endosymbionts that abundantly colonize the cytoplasm of termite gut flagellates (10, 11). In the context of this study, we use the term “free-living” to distinguish host-associated bacteria that occur freely within the gut from those in an endosymbiotic association with protists.

All evolutionary “lower” termites (all families except Termitidae or “higher” termites) harbor diverse assemblages of flagellate protists that are essential for the digestion of lignocellulose (12, 13). The flagellates themselves are generally associated with bacterial symbionts that colonize the exterior surface, cytoplasm, or nucleus of their hosts (14–18). Although the symbionts belong to different bacterial phyla, they share conspicuous similarities in the metabolic capacities encoded by their respective genomes. Despite a substantial reduction in genome size, the biosynthetic pathways of the symbionts are mostly conserved, which led to the consensus that the symbionts either provide their flagellate host with fixed nitrogen and/or essential amino acids and cofactors (16, 19–21) or contribute metabolic capacities that facilitate cellulose degradation, such as reductive acetogenesis (19, 22). The shared presence of such traits suggests a parallel evolution of unrelated lineages of endosymbionts.

Although the diversity of *Endomicrobiaceae* in termite guts has been investigated in great detail and numerous associations with different termite gut flagellates have been documented (5–7, 11, 23, 24), it is unclear whether all intracellular symbionts were derived from a single endosymbiotic event or were independently acquired from different lineages of free-living ancestors. This is mostly caused by the poorly resolved phylogeny of *Endomicrobiaceae* in 16S rRNA-based studies (10, 11). Moreover, the family comprises only a single isolate, *Endomicrobium proavitum* (25), and a few endosymbionts with sequenced genomes (26–28). While the free-living *Endomicrobium proavitum* and the endosymbiotic *Candidatus Endomicrobium trichonymphae* share a purely fermentative metabolism, they differ in genome size and guanine-cytosine (GC) content and in their metabolic capacities, such as the uptake and activation mechanisms of their major energy substrates, suggesting that the substantial gene loss in the endosymbiont was accompanied by acquisition of novel functions (29). In the absence of genomic information on other representatives of the family, it remains open whether endosymbiosis in *Endomicrobiaceae* arose more than once and how free-living lineage(s) adapted to the intracellular lifestyle.

To address these questions, we reconstructed more than 1,700 bacterial genomes from gut metagenomes of 48 termite species, representing all major families, including all subfamilies of the flagellate-free Termitidae. This effort yielded 67 novel metagenome-assembled genomes (MAGs) of *Endomicrobiaceae*. Together with previously published genomes, including numerous MAGs from termites (30), ruminants (31–34), and anaerobic bioreactors (35) that had not been analyzed to date, this provided a total of 106 genomes. We used this comprehensive data set to analyze the phylogeny of *Endomicrobiaceae* and the metabolic potential of the individual lineages, with the aims of identifying the distribution of putative endosymbionts and tracing the loss and gain of relevant functions during the evolutionary history of *Endomicrobiaceae*.

RESULTS

Metagenome-assembled genomes

Termite gut metagenomes yielded 1,723 MAGs of high and medium quality, ranging between 13 and 64 MAGs per host species (Table S1). Together with the 589 MAGs from our previous study (30), they comprise 2,228 MAGs from the bacterial domain and 84 MAGs from the archaeal domain. The archaeal MAGs comprised members of the euryarchaeal phyla *Thermoplasmatota* (33 MAGs), *Methanobacteriota* (15 MAGs), and *Halobacterota* (10 MAGs), and the crenarchaeal phylum *Thermoproteota* (26 MAGs); these have been covered in a separate study (36). The bacterial MAGs were classified into 30 phyla. The majority is represented by *Bacillota* (24.1%), *Bacteroidota* (15.6%), *Spirochaetota* (14.2%), and *Pseudomonadota* (13.3%). Most of the remaining MAGs fell into (in descending order) *Actinomycetota*, *Desulfobacterota*, *Verrucomicrobiota*, *Elusimicrobiota*, *Planctomycetota*, *Patescibacteria*, *Synergistota*, *Fibrobacterota*, *Campylobacterota*, *Acidobacteriota*, and *Myxococcota* (Fig. 1). The remaining 11 phyla were represented by less than 10 MAGs each (for details, see Table S2). Of particular interest for the present

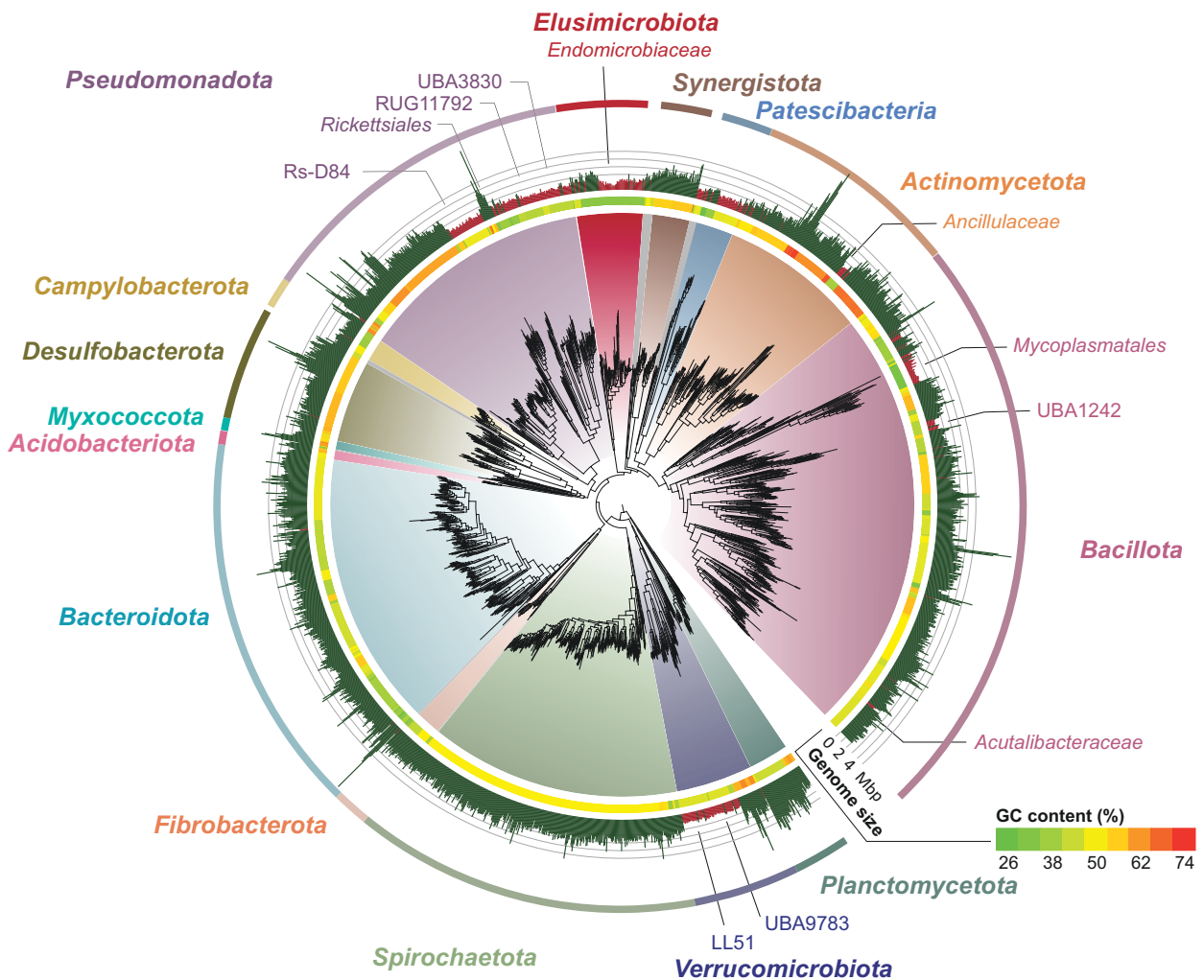


FIG 1 Phylogeny of 2,228 bacterial MAGs reconstructed from termite gut metagenomes. Phyla represented by more than 10 MAGs and subclades harboring putative endosymbionts are labeled. Bars illustrate genome size, MAGs < 1.5 Mbp are marked in red. GC content is indicated at family level. The maximum-likelihood tree was inferred from the concatenated amino acid alignment of 120 protein-coding genes generated by GTDB-Tk using the LG+F+I+G4 model of evolution. For a detailed classification of all MAGs and additional information, see Table S2.

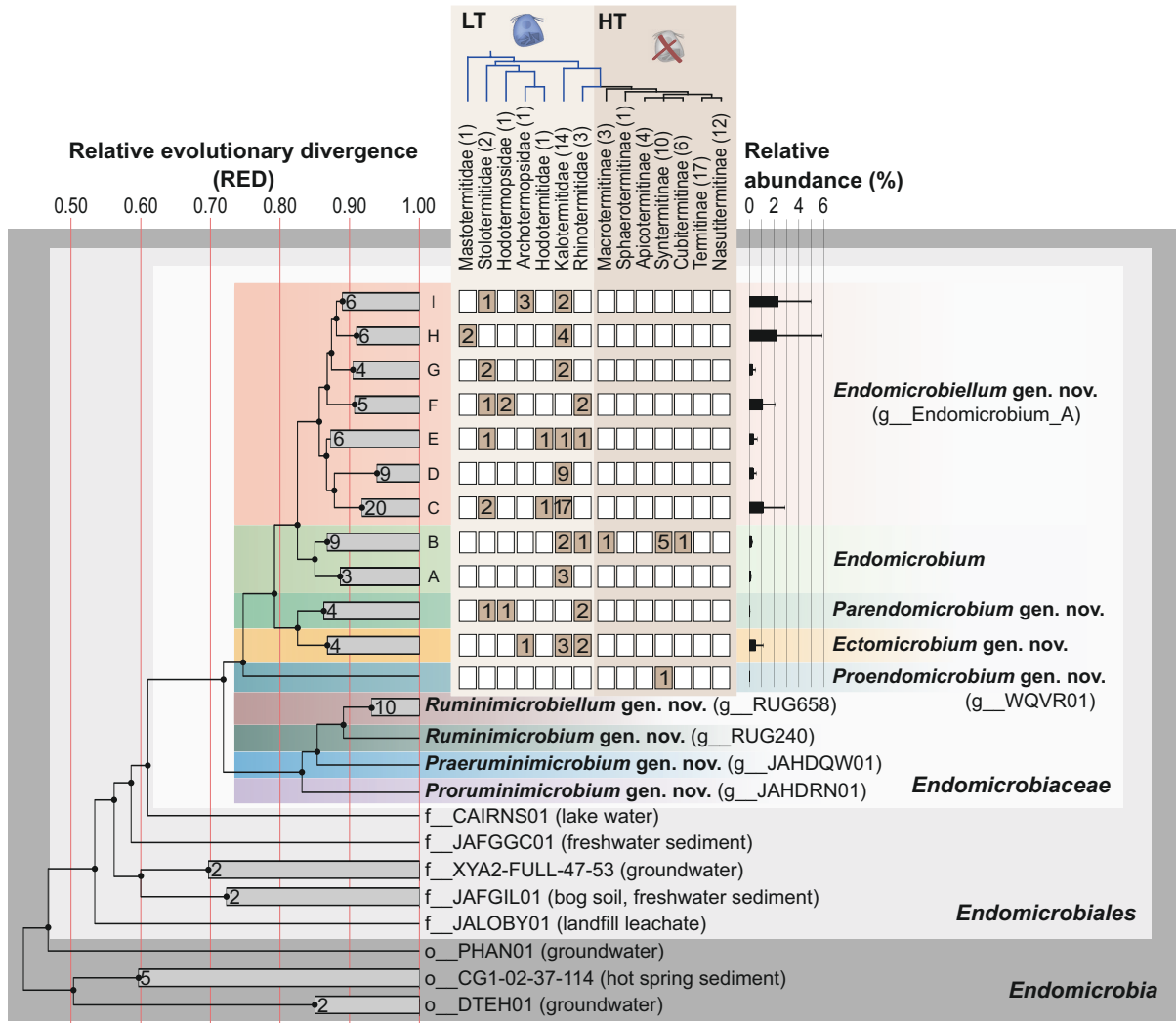


FIG 2 Phylogeny of the class *Endomicrobia*, illustrating the newly proposed taxonomy of *Endomicrobiaceae*. Provisional names from the GTDB taxonomy and environmental origin of ancestral lineages are indicated. The number of MAGs recovered from different families of lower termites (LT) and subfamilies of higher termites (HT), and the relative abundance (average) of the MAGs among the total sequencing reads of the respective metagenomes are shown. The maximum-likelihood tree was inferred using the same parameters as in Fig. 1 and was normalized using relative evolutionary divergence (RED) values determined with PhyloRank. Bullets on internal nodes indicate SH-aLRT/UFBoot support (●, both $\geq 95/99\%$; 1,000 replicates each). The number of genomes in the collapsed clades is indicated (for details, see Fig. 3; Fig. S1).

study were bacterial lineages comprising MAGs with small genome sizes (<1.5 Mbp) (Fig. 1). They include *Endomicrobiaceae* (*Elusimicrobiota*) and many lineages from other phyla that are already known to harbor endosymbionts of termite gut flagellates.

Diversity of *Endomicrobiaceae*

Almost all MAGs from the phylum *Elusimicrobiota* were members of the family *Endomicrobiaceae*. A detailed phylogenomic analysis that included all genomes of this family available to date (Table S3) revealed two highly supported monophyletic clades, one consisting exclusively of representatives from termites, the other of MAGs recovered from various ruminants or an anaerobic digester (Fig. 2).

Based on tree topology and relative evolutionary divergence (RED), the termite-specific clade comprises the genus *Endomicrobium* and four so-far unclassified genus-level lineages without cultured representatives. We propose to classify the MAGs in these lineages in the new genera *Proendomicrobium*, *Ectomicrobium*, *Parendomicrobium*,



FIG 3 Expanded phylogeny of *Endomicrobiaceae*, illustrating the key genes/pathways encoded by representative genomes of each lineage. The tree is the same as in Fig. 2; type strains are indicated in bold (type species in blue). More details, including the accession numbers of all genomes and the position of additional, low-quality MAGs, are given in Fig. S2. Genome size was estimated from assembly size and completeness (indicated by shaded and filled bars). Glycolysis was considered absent if fructose-bisphosphate aldolase or pyruvate kinase is missing; the absence of other enzymes of the Embden-Meyerhoff pathway (EMP) is discussed in the text. HexUA, hexuronate; Ldh, lactate dehydrogenase; AdhE, bifunctional aldehyde–alcohol dehydrogenase; PFOR, pyruvate:ferredoxin/flavodoxin oxidoreductase; Pta, phosphate acetyltransferase; Ack, acetate kinase; Hyd, hydrogenase; NifHDK, nitrogenase; AmtB, ammonium transporter; AroP, aromatic amino acid transporter; SdaC, serine transporter; ProT, proline transporter; GltS, sodium–glutamate symporter. For information on the individual genes of each pathway, including xylose and hexuronic acid metabolism and components of the cell envelope, see Fig. S4, S10, S12 and S13.

and *Endomicrobiellum* (see “Taxonomy” below). Almost all genera, including also the well-supported subgroups (A–I) of *Endomicrobium* and *Endomicrobiellum*, are represented by at least one high-quality MAG (completeness $\geq 90\%$, contamination $\leq 5\%$). All previously characterized endosymbionts, including all candidate species described to date (*Ca. Endomicrobium trichonymphae*, *Ca. Endomicrobium pyrsonymphae*, *Ca. Endomicrobium agile*, and *Ca. Endomicrobium dinenymphae*) fall into the new genus *Endomicrobiellum*. The second clade comprises four so-far unclassified genus-level lineages, two of which are represented by MAGs from the rumen. We propose to classify the MAGs in these lineages in the new genera *Proruminimicrobium*, *Praeruminimicrobium*, *Ruminimicrobium*, and *Ruminimicrobiellum* (see “Taxonomy” below).

The vast majority of the *Endomicrobiaceae* MAGs recovered from termite guts were from lower termites (Fig. 2). Most of them fell into the genus *Endomicrobiellum*. Except for *Prorhinotermes canalifrons*, each termite sample yielded one or more medium- or high-quality MAGs. Only a few MAGs were recovered from higher termites. One fell into the genus *Proendomicrobium*, all others into the genus *Endomicrobium*, which are of low relative abundance in both lower and higher termites (Fig. 2). Each MAG represents a phylotype unique to a particular termite (Table S3). Low-quality MAGs were recovered from many samples, indicating the presence of additional, albeit smaller populations. The inclusion of low-quality MAGs in the phylogenomic tree corroborated the status of *Proendomicrobium* (represented by only a single medium-quality MAG) as a separate lineage (Fig. S1). A 16S rRNA-based analysis of *Endomicrobiaceae*, including all homologs recovered from the MAGs and the corresponding metagenomes, allowed us to identify all genus-level lineages defined in the phylogenomic tree except *Proendomicrobium* (Fig. S3). This includes the genus *Parendomicrobium*, which comprises MAGs that are linked to ectosymbionts of a *Spirotrichonympha*-like flagellate in *Porotermes adamsoni* (24) and to the gut microbiota of *Reticulitermes flavipes* by several unbinned sequences from the corresponding metagenomes.

Although most genera of *Endomicrobiaceae* comprise representatives with relatively large genomes (average size estimates 2.11 Mbp \pm 0.26 Mbp) and a relatively high GC content (average 38.0% \pm 3.2%), the genomes of *Proendomicrobium*, *Ectomicrobium*, and *Endomicrobiellum* are considerably smaller (average 1.16 Mbp \pm 0.30 Mbp; Table S4). Members of the genus *Endomicrobiellum*, which comprises all established endosymbionts and is absent from higher termites, have the smallest genomes (average 1.13 Mbp \pm 0.29 Mbp) and lowest GC contents (average 32.3% \pm 3.0%) of the entire family, corroborating that they are endosymbionts of termite gut flagellates.

Energy substrates

We compared the pathways involved in energy metabolism and assimilatory functions encoded in the genomes of *Endomicrobiaceae* (Fig. 3). The metabolic capacities shared by all family members and major differences between the putatively free-living lineages and endosymbiotic genus *Endomicrobiellum* are summarized in Fig. 4.

All members of *Endomicrobiaceae* are obligately anaerobic sugar fermenters. They generally possess two glucose uptake systems: a putative ABC transporter of the CUT1 family and a phosphotransferase system (PTS) of the mannose type. Although the ABC transporter was lost in all members of *Endomicrobiellum*, the PTS system is preserved in most lineages. Instead, almost all *Parendomicrobium* and *Endomicrobiellum* genomes encode one or two transporters of the organophosphate:inorganic phosphate antiporters (OPA) family, which are absent from all other genera. They fall into the radiation of sugar phosphate transporters from a wide range of bacterial lineages, including the biochemically characterized glucose-6-phosphate (Glc6P) transporter in *Chlamydia pneumoniae* and Glc6P-transporting sensor of *Escherichia coli* (UhpC) (37) (Fig. 5). One clade (OPA1) clusters among homologs from *Elusimicrobiaceae*, the other (OPA2) clusters among homologs of *Bacteroidales* and *Opiritaceae*, which are abundantly represented among the MAGs of this study. Homologs from this clade are frequently annotated as glycerol-3-phosphate (Gly3P) transporters, but because none of the transporters in

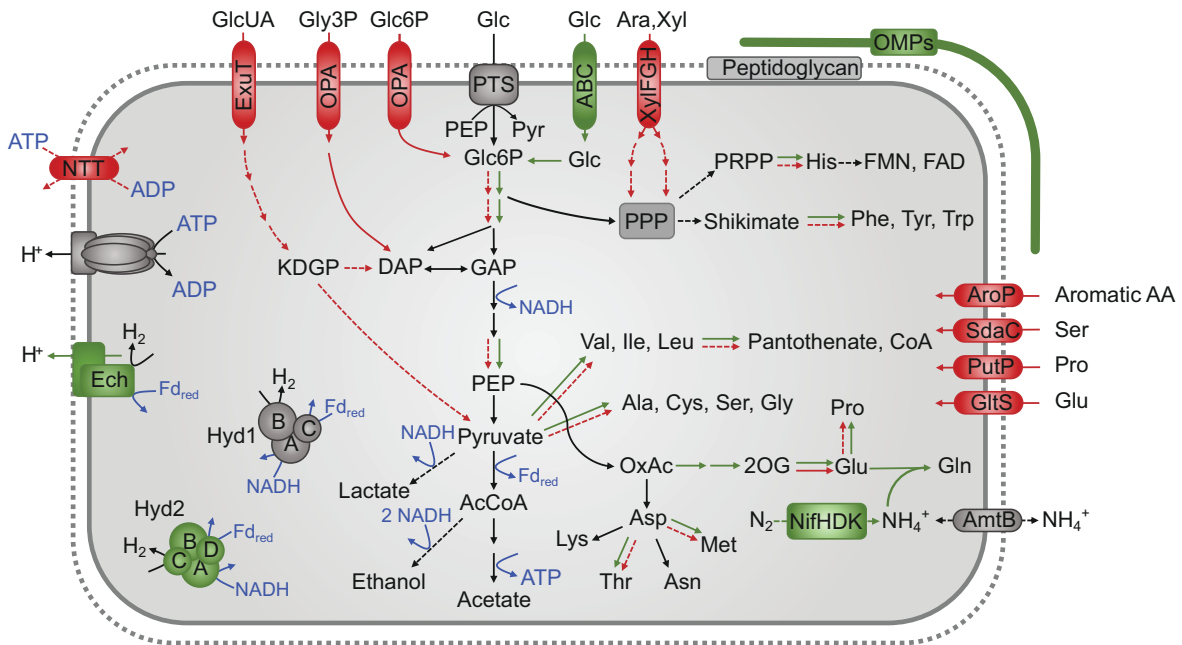


FIG 4 Metabolic capacities shared by members of *Endomicrobiaceae*. Pathways present only in putatively free-living lineages (green) or in endosymbionts of the genus *Endomicrobiellum* (red) are highlighted. Dashed lines indicate that a pathway is not present in all members of each group. GlcUA, glucuronic acid; Gly3P, glycerol 3-phosphate; PTS, phosphotransferase system; ABC, ABC transporter; XylFGH, arabinose/xylose transporter; OMPs, outer membrane proteins; NTT, nucleotide triphosphate transporter; PPP, pentose phosphate pathway, PRPP, phosphoribosyl pyrophosphate; KDGP, 2-keto-3-deoxygluconate 6-phosphate; for more abbreviations, see the legend of Fig. 3.

the respective clades have been biochemically characterized, their substrate specificity remains tentative. Nevertheless, the results suggest that members of the genera *Parendomicrobium* and *Endomicrobiellum* have acquired the capacity to take up sugar phosphates.

Intriguingly, members of *Endomicrobiellum* lineage D encode a nucleotide triphosphate transporter (NTT) that is embedded among homologs from the order UBA3830, a lineage of uncultured alphaproteobacteria from termites and ruminants abundantly represented among the MAGs of this study, and that is closely related to biochemically characterized ATP/ADP antiporters of *Rickettsiales* (38) (Fig. 6). The residues critical for function and substrate specificity in the antiporter of *Arabidopsis thaliana* (39) are conserved in both the homologs of *Endomicrobiellum* lineage D and the order UBA3830 (Fig. S4). A few MAGs from other *Endomicrobiellum* lineages encode NTTs that fall into the radiation of homologs present in other MAGs from termite guts (*Acutalibacteraceae* and the alphaproteobacterial order RUG11792).

Fermentative metabolism

All *Endomicrobiaceae* are strict anaerobes that convert glucose to pyruvate via the canonical Embden-Meyerhoff pathway (EMP; Fig. 3). There is no evidence for respiratory metabolism, and a catalase gene is absent from all genomes. However, genes encoding key enzymes of the glycolytic pathway are absent in many members of *Endomicrobiellum* and several other lineages. The loss of glucokinase (GlcK) is compensated by the PTS and/or the acquisition of UhpC, both of which provide the EMP with glucose 6-phosphate (see "Energy substrates"). While the formation of fructose 1,6-bisphosphate via pyrophosphate-fructose 6-phosphate 1-phosphotransferase (Pfp) instead of the canonical phosphofructokinase (Pfk) would even save ATP (40), the loss of enolase is difficult to explain. Notably, an enolase gene is consistently absent not only in two apical clades of *Endomicrobiellum* (lineages H-I) but also from members of other genera

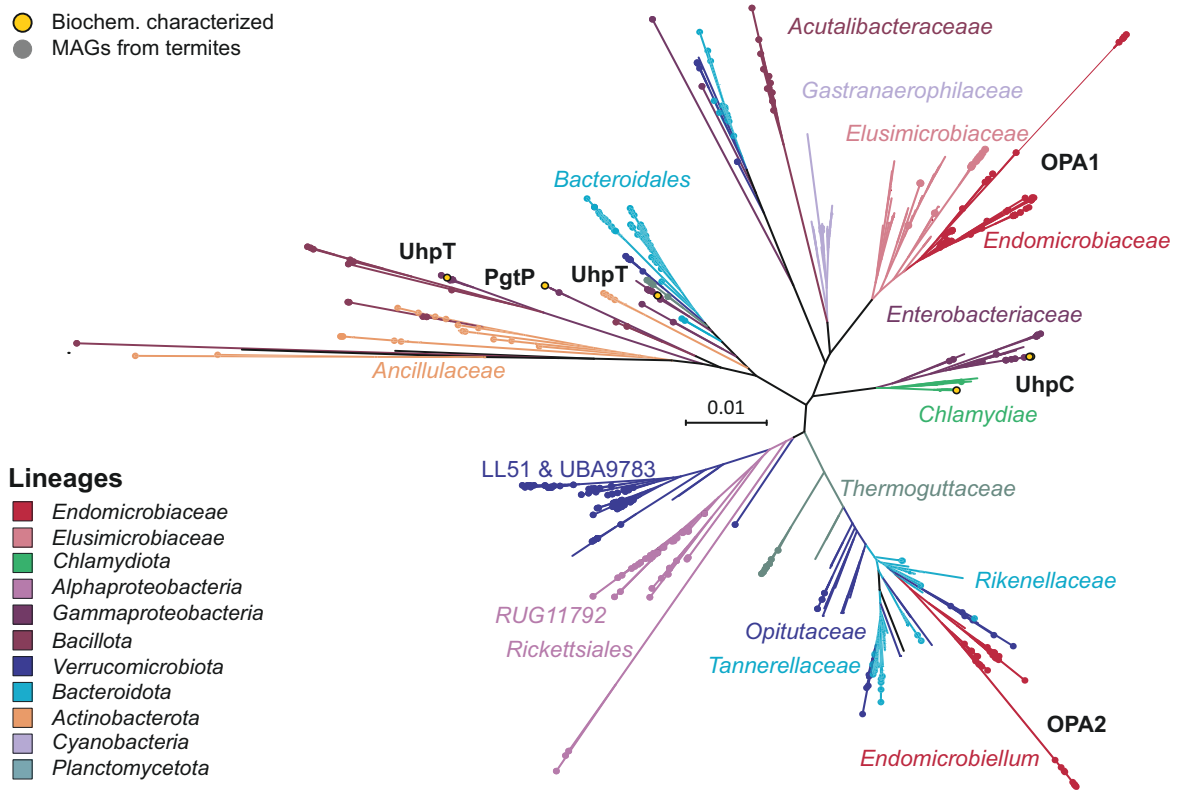


FIG 5 Phylogeny of the OPA family transporters in *Endomicrobiaceae* and their homologs in public databases. Organismal lineages are color coded; filled circles indicate homologs in MAGs obtained from termite guts. Biochemically characterized homologs (SWISS-PROT) are highlighted. The maximum-likelihood tree is based on a curated alignment of 428 amino acid positions and was inferred under the LG+F+R9 model of evolution. Bootstrap values were omitted for clarity.

(e.g., *Proendomicrobium*, *Parendomicrobium*, and *Ruminimicrobiellum*). Considering that a methylglyoxal synthase gene is absent in *Endomicrobiaceae*, the enolase reaction cannot be circumvented via the methylglyoxal pathway, i.e., the conversion of dihydroxyacetone phosphate (DAP) to pyruvate (41).

Several subclades of *Endomicrobiellum* and members of the genera *Ectomicrobium* and *Parendomicrobium* possess gene sets for the uptake and metabolism of hexuronic acids via the 2-keto-3-deoxy-phosphogluconate (KDPG) pathway (Fig. 3). Genes for the metabolism of xylose and arabinose, which had been detected already in single-cell amplified genomes of *Endomicrobiellum dinenymphae* and *Endomicrobiellum pyrsonymphae* (28), were found also in a few MAGs from lineages C and G (Fig. S5). Genes for glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydratase, key enzymes of the Entner-Doudoroff pathway, are absent from the entire family.

Endomicrobiaceae generally oxidize pyruvate to acetyl-CoA via the canonical four-subunit pyruvate:ferredoxin oxidoreductase (PFOR) present in many anaerobes. Only *Endomicrobium proavitum* and members of *Endomicrobium* subcluster A possess a flavodoxin-dependent homolog with fused subunits, which was apparently acquired from *Elusimicrobiaceae* ([27]; Fig. S5 and S6). In all members of the termite clade, acetyl-CoA is converted to acetate by phosphotransacetylase (Pta) and acetate kinase (Ack). Pta is absent in the entire ruminant clade, whereas an acetyl-CoA synthetase (AMP-forming) is present in most *Endomicrobiaceae* (Fig. S5).

Reduced coenzymes (NADH and ferredoxin) are reoxidized either by the reduction of pyruvate to lactate and/or ethanol, or by the formation of hydrogen. While the lactate dehydrogenase (Ldh) of *Endomicrobiaceae* falls into the radiation of homologs from other *Elusimicrobiota* (Fig. S7), the homologs of the bifunctional aldehyde/alcohol

Lineages

- *Endomicrobiaceae*
- *Chlamydiota*
- *Alphaproteobacteria*
- *Gammaproteobacteria*
- *Bacillota*
- *Verrucomicrobiota*
- *Bacteroidota*
- *Desulfobacterota*
- *Dependentiae*
- *Cyanobacteria*
- *Acidobacteriota*
- *Planctomycetota*
- *Photosynth. Eukarya*
- *Fungi*

- Biochem. characterized
- MAGs from termites

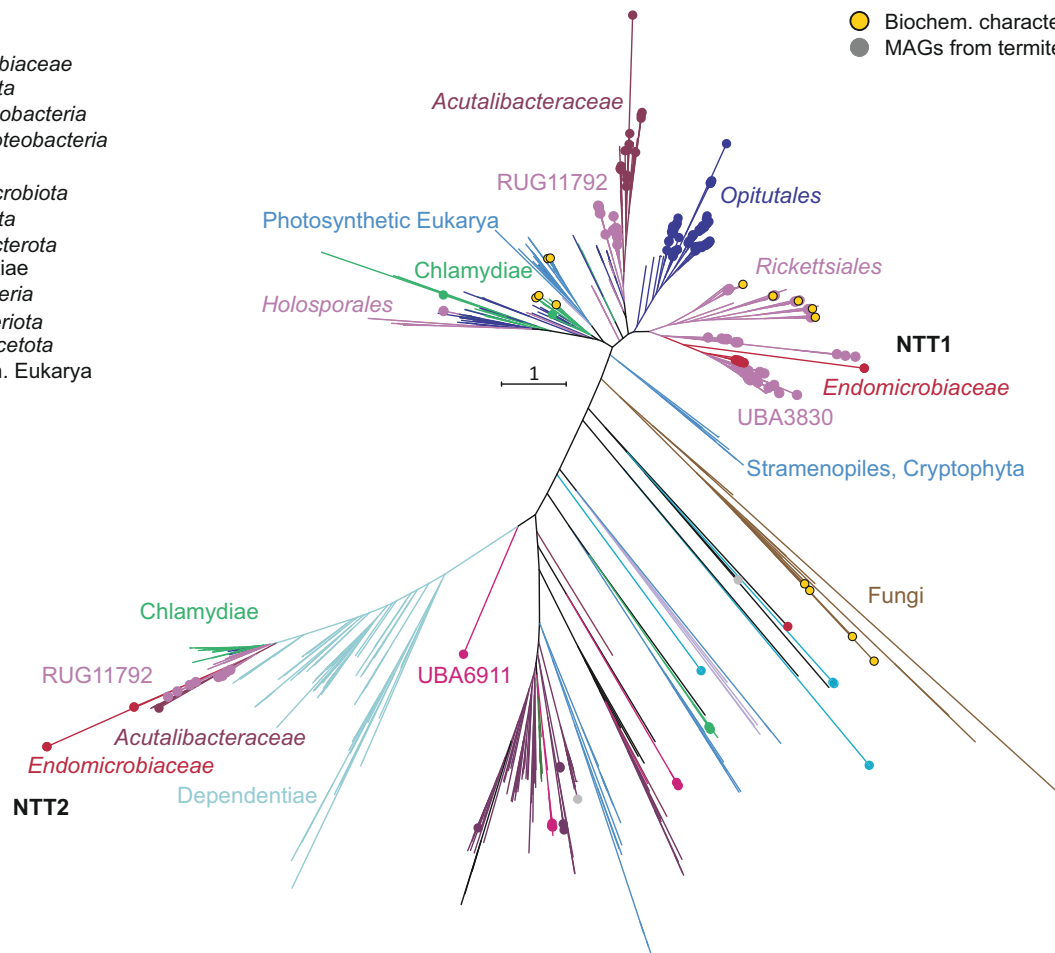


FIG 6 Phylogeny of the NTT in *Endomicrobiaceae* and their homologs in public databases. Organismal lineages are color coded; filled circles indicate homologs in MAGs obtained from termite guts. Biochemically characterized homologs (SWISS-PROT) are highlighted. The maximum-likelihood tree is based on a curated alignment of 982 amino acid positions and was inferred under the LG+F+R5 model of evolution. Bootstrap values were omitted for clarity.

dehydrogenase (AdhE) in the genera *Ectomicrobium*, *Parendomicrobium*, and *Endomicrobiellum* are not directly related and were apparently acquired independently via horizontal gene transfer from other phyla (Fig. S8).

Hydrogenase genes of *Endomicrobiaceae* were previously identified in *Endomicrobium proavitum* and *Endomicrobiellum trichonymphae* (29). Members of most genera encode a trimeric, electron-bifurcating [FeFe] hydrogenase (HydABC) of group A3 that is absent from other representatives of *Elusimicrobiota*. Several free-living lineages, including those from ruminants, possess a second, tetrameric homolog of the same subclass but of independent phylogenetic origin (Fig. S9). A membrane-bound [NiFe] hydrogenase of group 4g, which comprises uncharacterized, presumably ferredoxin-dependent, hydrogen-evolving enzymes (42), is restricted to members of the genus *Endomicrobium* and only distantly related to homologs from the order *Elusimicrobiales* (see Fig. S10).

Biosynthetic pathways

Endomicrobiaceae generally lack genes encoding fructose 1,6-bisphosphatase (Fbp) and PEP carboxykinase (PckA), the key enzymes of gluconeogenesis from pyruvate. The pathway for the synthesis of oxaloacetate, which provides the carbon skeleton for amino acids of the aspartate family, including the corresponding aminotransferases, is present in all *Endomicrobiaceae*. The pathway for the synthesis of 2-oxoglutarate,

the precursor for amino acids of the glutamate family, however, is absent from most members of *Endomicrobiellum* (Fig. S11). Most members of the free-living genera encode an ammonia transporter (AmtB), glutamine:oxoglutarate aminotransferase (GltA), and glutamine synthetase (GlnA), the enzymes required for ammonia assimilation into glutamine. *Endomicrobiaceae* lack a canonical citrate synthase, but all genera, including several lineages of the endosymbiotic *Endomicrobiellum*, encode homologs of a putative *Re*-type citrate synthase, which fall into the same clade as the biochemically characterized enzymes from other phyla (29).

All putatively free-living lineages possess pathways for the biosynthesis of almost all proteinogenic amino acids (Fig. 4; Fig. S11). Glutamate is synthesized by a tetrameric variant of GltA that is present also in *Pyrococcus* sp. (43). An asparagine synthetase (AsnAB) is absent from all *Endomicrobiaceae*, but the majority of the MAGs encode an unusual aspartyl-tRNA synthetase (AspS2), which loads an aspartyl residue onto tRNA^{Asn}, where it is subsequently converted to an asparaginy residue by aspartyl-tRNA amidotransferase (GatCAB) (44). All *Endomicrobiaceae* synthesize glycine from serine via serine hydroxymethyltransferase (GlyA); the alternative threonine utilization pathway is absent. These findings explain the prototrophy of *Endomicrobium proavitum* for all amino acids (29) except serine. A complete phosphoserine pathway (SerABC) is present only in members of the ruminant clade. Members of the termite clade encode only the 3-phosphoglycerate dehydrogenase (SerA) of the phosphoserine pathway and the serine:pyruvate aminotransferase (Spa) but not the hydroxypyruvate reductase (Hpr) of the glycerate pathway.

Despite massive genome erosion in the genus *Endomicrobiellum*, several lineages retained pathways for the biosynthesis of up to 15 amino acids (Fig. 3). Although pathways for the biosynthesis of aspartate, asparagine, lysine, glycine, and alanine are conserved in almost all endosymbionts, the capacity to synthesize serine, glutamine, and aromatic amino acids was lost in several subclades. GltA is present in some lineages, but the oxidative branch of the tricarboxylic acid (TCA) cycle forming 2-oxoglutarate is missing in all members of the genus. The presence of several amino acid transporters, such as an aromatic amino acid transporter (AroP), a serine transporter (SdaC), a proline symporter (ProT), and a glutamate symporter (GltS), exclusively in members of the genus *Endomicrobiellum* suggests that the loss of the respective biosynthetic capacities is compensated by the uptake of amino acids provided by their flagellate hosts.

A few members of the genera *Ectomicrobium* and *Endomicrobium*, including the diazotrophic *Endomicrobium proavitum* (25), possess a gene cluster encoding a group IV nitrogenase (NifDKH) (Fig. 3), which is represented also in *nifH* gene inventories of termite guts (e.g., references 45, 46; Fig. S12). The conservation of AmtB in many *Endomicrobiellum* lineages may indicate the necessity to export NH₄⁺ resulting from a utilization of amino acids as alternative carbon sources (see Discussion).

Most representatives of *Endomicrobiaceae* possess the biosynthetic pathways for most cofactors (Fig. 3). Although the capacity to synthesize biotin from pimelyl-CoA is present in most genera, the synthesis of the pimelate moiety is unclear and (as in many other bacteria) differs from the established pathways (47). Several subclades of *Endomicrobiellum* (lineages D, G–I) lost the ability to synthesize biotin, coenzyme A, and flavin nucleotides (FMN, FAD) (Fig. S11).

The pathways for the biosynthesis of peptidoglycan are conserved in most members of *Endomicrobiaceae*, suggesting that also the endosymbiotic lineages have retained the capacity to produce a murein sacculus (Fig. S13). Notably, the bifunctional murein transglycosylase/transpeptidase (MrcB) is absent from almost the entire termite clade, and a D-alanine:D-alanine ligase (Ddl) is lacking in several subclades of *Endomicrobiellum*. Also, key genes for lipopolysaccharide (LPS) synthesis and several outer membrane proteins, including OmpA, TolB, and the vitamin B₁₂ transporter BtuB, are present in most genera but absent in many members of *Endomicrobiellum*. Components characteristic for the outer membrane of other bacteria (e.g., the outer membrane complex assembly

proteins BamBCD, the long-chain fatty acid protein FadL, and the LPS assembly protein LptD) are absent from all *Endomicrobiaceae*.

The only cell appendages encoded by *Endomicrobiaceae* are pilins. The absence of a flagellar machinery from all genomes of the phylum *Elusimicrobiota* is consistent with previous studies (26, 29, 48) and agrees with the observation that all isolates from the phylum *Elusimicrobiota* are non-motile (49).

Genes for informational processing, genome modification, and repair

We found no notable difference between putatively free-living and endosymbiotic lineages in gene functions that are typically lost during genome erosion in bacterial endosymbionts of insect tissues, such as DNA replication or repair, protein folding and stability, and competence and transformation (50) (Fig. S14). Transposons were detected in more than half of the genomes but seem to be randomly distributed among free-living and endosymbiotic lineages (Table S5). CRISPR-Cas systems were present only in a few representatives (15 MAGs) of diverse lineages, including 11 systems of type II-C and 3 systems of type I-C (Table S6). Their random distribution between free-living and endosymbiotic lineages makes it unlikely that these subgroups are involved in genome reduction and rearrangement in *Endomicrobiellum* (27). Also, restriction-modification systems, which are abundant in *Endomicrobiellum trichonymphae* and have been implicated in genome rearrangement (51), are rare and randomly distributed among *Endomicrobiaceae* (Table S6).

Potential endosymbionts in other phyla

Additionally, numerous lineages of MAGs from other bacterial phyla that occur exclusively in lower termites showed genome sizes that were in the same range as those of the genus *Endomicrobiellum* (<1.5 Mbp) (Fig. 1). Several lineages represented established endosymbionts of termite gut flagellates ("*Ancillulaceae*" and *Acutalibacteraceae*) and other protists (e.g., *Rickettsiales*, *Acholeplasmatales*, *Mycoplasmatales*, and the *Opitutales* families LL51 and UBA9783). Others may represent so-far unknown lineages of endosymbionts (e.g., the alphaproteobacterial orders RUG11792 and UBA3830). Additional lineages with small genomes were represented by MAGs from both lower termites and (flagellate-free) higher termites (e.g., *Patescibacteria*, *Dehalococcoidales*, the clostridial order UBA1242, and the alphaproteobacterial order Rs-D84).

Genome size and GC content of "*Ancillulaceae*," three clades of *Alphaproteobacteria* (*Rickettsiales*, UBA3830, and RUG11792), and a large clade of *Verrucomicrobiota* (families LL51 and UBA9783) were consistently lower than those of MAGs from the corresponding parent taxa (Fig. 7). Other lineages comprised only a few MAGs with significantly smaller genomes (e.g., "*Adiutricaceae*," *Acutalibacteraceae*, *Azobacteroidaceae*, and *Desulfovibrionaceae*) and/or consisted of MAGs from both lower and higher termites (e.g., UBA1242, Rs-D84, "*Adiutricaceae*"), corroborating the notion that not all MAGs of these lineages represent endosymbionts of termite gut flagellates (19, 52, 53).

An inspection of all MAGs representing putative endosymbionts for symbiosis-related traits detected in *Endomicrobiellum* revealed that members of almost all lineages have organophosphate transporters (OPAs) (Table S7). KDPG aldolase and the corresponding transporter for hexuronic acids were present only in the already established endosymbionts among *Azobacteroidaceae*, "*Ancillulaceae*," and *Acutalibacteraceae*. An NTT is present in the endosymbiotic (or putatively endosymbiotic) *Acutalibacteraceae*, *Opitutales*, and the alphaproteobacterial orders UBA3830 and RUG11792. Phylogenetic analyses of the key genes for the respective functions revealed that they are typically absent from related taxa from other environments and most likely acquired by horizontal gene transfer within the termite gut (Fig. 5 and 6; Fig. S16 and S17). Only a few lineages of flagellate endosymbionts ("*Adiutricaceae*," "*Ancillulaceae*") encode homologs of aromatic amino acid transporters (AroP), which were most likely derived from distantly related organisms (typically from different phyla) (Fig. S17). The capacity for nitrogen

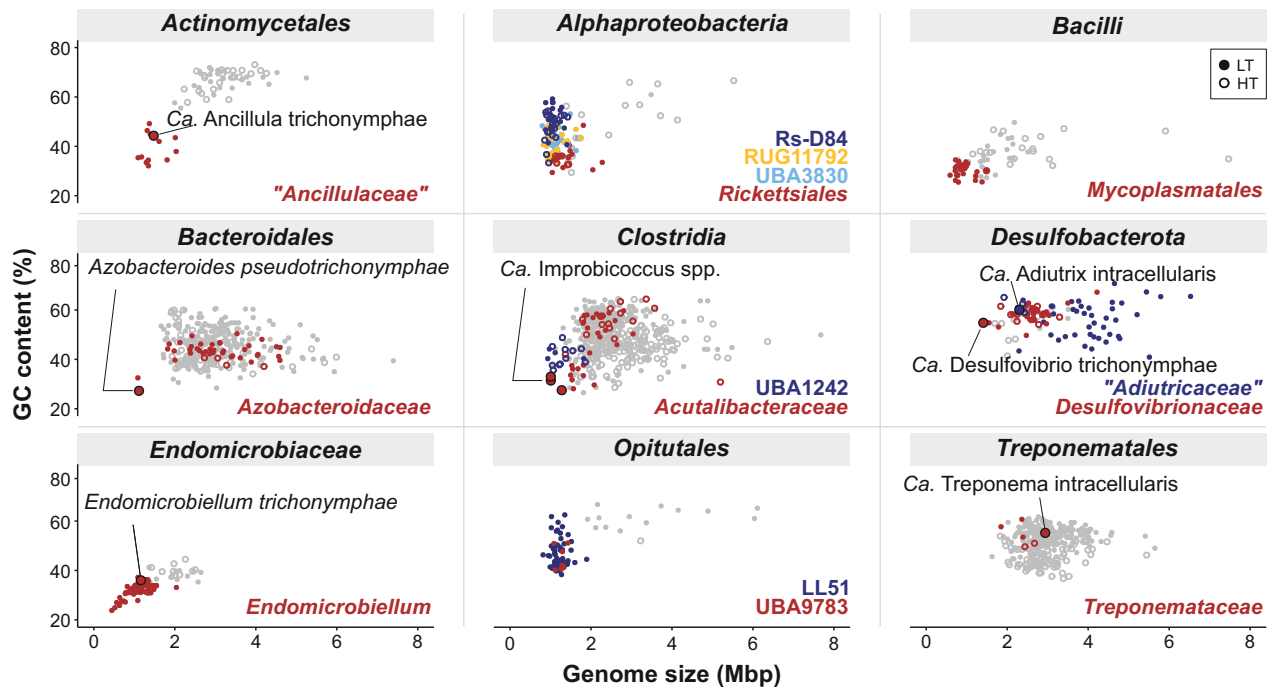


FIG 7 Genome size and GC content of the MAGs from termite guts, comparing lineages that comprise established and putative endosymbionts (in color) with their closest relatives (in gray). The origin of the MAGs from lower termites (●) and higher termites (○) is indicated; the genomes of well-characterized endosymbionts are marked. For an interactive map that shows the identity of each MAG, see Fig. S15.

fixation is present only among “Aduitricaceae,” Azobacteroidaceae, and Desulfovibrionaceae (Table S7).

DISCUSSION

The discovery of the first sequences of the “Termite group 1” (54, 55) and their association with termite gut flagellates (5, 7), the phylogenetic origin of Endomicrobia, and the relationship between free-living and endosymbiotic lineages remained unresolved. Our study revealed that host-associated *Endomicrobiaceae* originated among aquatic ancestors that subsequently colonized the intestinal tract of insects and ruminants. Although most lineages represent free-living gut commensals, a common ancestor of the genus *Endomicrobiellum* engaged in an intracellular symbiosis with termite gut flagellates. Progressive genome reduction led to a breakdown of glycolysis and many biosynthetic pathways, which was compensated by the acquisition of new functions via horizontal gene transfer. In some subgroups, the apparent loss of all biosynthetic capacities that could potentially benefit the flagellate host indicates the decline of an originally mutualistic relationship.

Phylogenetic diversity of *Endomicrobiaceae*

The presence of two lineages from anaerobic digestors (*Proruminimicrobium* and *Praeruminimicrobium*) in a position ancestral to the host-associated members of the ruminant clade (*Ruminimicrobium* and *Ruminimicrobiellum*) corroborates the origin of the family *Endomicrobiaceae* among lineages of aquatic bacteria in the order *Endomicrobiales* (49). Although the environmental origin of the ruminant clade is supported also by 16S rRNA-based diversity data, the origin of the termite-associated clade remains unresolved (Fig. S3). Previous 16S-rRNA-based studies detected members of *Endomicrobiaceae* also in other insects, such as cockroaches (7, 10, 11) and a scarab beetle larva (56), but, with the exception of lower termites, always in low abundance (11, 57, 58). This explains

the absence of *Endomicrobiaceae* among MAGs from cockroaches (59, 60) and the low recovery from higher termites (Fig. 2). Notably, the sister position of the 16S rRNA genes of *Endomicrobiaceae* from cockroaches to the genus *Endomicrobium* is only poorly supported (Fig. S3) and remains to be verified with phylogenomic data. The 16S rRNA gene tree documents an affinity of *Ca. Endomicrobium superficiale*, the ectosymbionts of *Trichonympha magna* from *Porotermes adamsoni*, and spirotrichosomid flagellates from the related *Stolotermes victoriensis* (24), to the genus *Ectomicrobium*, but in the absence of a genome sequence, this assignment must remain tentative.

The genus *Endomicrobiellum* comprises all endosymbiotic representatives that were localized in previous studies (5–7, 23, 61). Compared to other genera of *Endomicrobiaceae*, their genomes are strongly reduced, which is in agreement with the endosymbiotic nature of the entire genus. Notably, also members of the genus *Ectomicrobium*, which are associated with gut flagellates (24), have slightly smaller genomes than most other family members, including their sister genus, *Parendomicrobium* (Table S4).

The genera *Ruminimicrobium* and *Ruminimicrobiellum* comprise MAGs recovered from large-scale studies of ruminal microbiota of sheep, cows, deer, and goats (31–34, 62). It has been speculated that *Endomicrobiaceae* present in the rumen are associated with ciliates (9). This is supported by a significantly smaller genome size in members of the genus *Ruminimicrobiellum* compared to most other lineages of the family (Table S4), and by our observation that 16S rRNA genes of *Endomicrobiaceae* recovered from capillary-picked suspension of rumen ciliates (*Isotricha* spp.) fall into the radiation of this genus (U. S. Mies and A. Brune, unpublished data; see Fig. S3). Therefore, it is possible that an association with protists has evolved independently in three lineages of *Endomicrobiaceae*.

Genome erosion and acquisition of new functions in *Endomicrobiellum*

The fermentative energy metabolism of *Endomicrobium proavitum* (29) is shared by all members of *Endomicrobiaceae*. Unlike several other lineages of the class *Elusimicrobia* that encode respiratory chains for oxygen and other inorganic electron acceptors (63), all members of *Endomicrobiaceae* are strict anaerobes that activate glucose to Glc6P with glucokinase and/or a PTS, and ferment Glc6P via the EMP pathway to ethanol, lactate, acetate, hydrogen, and CO₂. All members of the endosymbiotic genus *Endomicrobiellum*, including the previously characterized *Em. trichonymphae* (26, 27, 29), lack the ABC transporter and often also the PTS for the uptake of glucose. These gene losses are compensated by two organophosphate:phosphate antiporters (OPA1 and OPA2) that were independently acquired by different subclades and allow the direct uptake of sugar phosphates from the cytoplasm of their respective hosts.

Although annotated as Glc6P and Gly3P transporters, both OPA1 and OPA2 fall into phylogenetic clades without biochemically characterized representatives (Fig. 5); therefore, their exact substrates remain uncertain. Nevertheless, either of these substrates would serve to link the energy metabolism of endosymbiont and host, as observed in other bacteria with an intracellular lifestyle (e.g., *Listeria* and *Chlamydia* [64, 65]). Here, the switch to sugar phosphates is considered an adaptation to the low concentrations of free glucose in eukaryotic cells that provides an additional, energetic advantage because no ATP is needed for substrate activation. It is not clear whether the acquisition of OPAs by members of the genus *Endomicrobiellum* was a predisposition for an intracellular lifestyle, as in the ancestors of *Chlamydia* (65), or a response to the conditions in their intracellular habitat. In any case, these transporters may be the only means to obtain an energy substrate for those members of the genus *Endomicrobiellum* that lack both glucokinase and PTS (lineages H and I).

The absence of enolase in several basal lineages with an otherwise complete glycolytic pathway is rather perplexing. The same situation has been reported for a variety of host-associated bacteria, including *Treponematales* and *Bacteroidales* symbionts of termite gut flagellates (22, 66) and several *Saccharibacteria* (67, 68) and *Clostridiales* from mammals (69). Because a methylglyoxylate shunt, which allows

circumvention of the enolase function in the rumen bacterium *Butyrivibrio proteoclasticus* (70), is absent from all *Endomicrobiaceae*, it remains unclear how members of these lineages generate the phosphoenolpyruvate required, e.g., by PTS.

The fermentative metabolism of glucose and other energy substrates requires regeneration of the reduced cofactors formed in the oxidative part of the pathway. Although most members of *Endomicrobiaceae* encode homologs of Ldh and/or AdhE and at least one hydrogenase, the lineages of *Endomicrobiellum* without a functional glycolysis (lineages H and I) also lack the enzymes that reoxidize NADH and reduced ferredoxin, indicating that they lost the capacity for sugar fermentation. It is possible that they conserve energy by oxidative decarboxylation of amino-acid-derived 2-oxoacids to the corresponding acyl-CoA esters, as in *Elusimicrobium minutum* (71) and in hyperthermophilic archaea (72, 73).

An interesting case is presented by the MAGs that possess an ATP/ADP antiporter, which enables them to import ATP directly from the host. This NTT was independently acquired from uncultured alphaproteobacteria (orders UBA3830 and RUG11792) by *Endomicrobiellum* lineage D and several MAGs from other lineages. ATP/ADP antiporters and other NTTs are regarded as a hallmark of parasitism and are common in intracellular parasites of amoebae, such as *Chlamydia* and *Rickettsia*, where they are used by the parasites to obtain energy and nucleotide triphosphates for their growth and replication from their respective hosts (65, 74). However, NTTs are found also in free-living, non-parasitic bacteria (75); therefore, it remains open to speculation whether members of *Endomicrobiellum* have evolved into energy parasites or still provide a function that is beneficial for their flagellate hosts.

Host–symbiont interactions

The intracellular location provides endosymbionts with a protective barrier against environmental forces. At the same time, an intact cell envelope may also hinder horizontal gene transfer and the exchange of metabolites between symbiont and host, which explains why the pathways for the synthesis of peptidoglycan and an outer membrane are frequently lost during genome erosion (76). Members of the genera *Endomicrobiellum* retained the capacity to synthesize a murein sacculus, LPS, and several outer membrane proteins (OMPs), which suggests that their cell envelope is still intact. However, the absence of the bifunctional transglycosylase/transpeptidase (MrcB) and D-alanine:D-alanine ligase (Ddl) in several subclades of *Endomicrobiellum* may affect the integrity of their cell wall (77), which may facilitate metabolite exchange between endosymbionts and host.

By default, an intracellular bacterium must synthesize all amino acids and vitamins that it cannot acquire from its host. Therefore, any deleterious mutations in its biosynthetic machinery must be compensated by an uptake mechanism that was either present already in the free-living ancestors of the endosymbiont or was later acquired by horizontal gene transfer. These two scenarios are exemplified by the ancestral presence of an uptake system for a proline transporter (ProT) and the acquisition from other bacteria of transporters for serine (SdaC), glutamate (GltS), and aromatic amino acid (AroP). These latter transporters are present in almost all lineages of *Endomicrobiellum* and are most closely related to homologs from unrelated bacteria present in termite guts.

The acquisition of amino acid transporters and conservation of many biosynthetic pathways in the genus *Endomicrobiellum* are in agreement with the long-standing hypothesis that flagellate endosymbionts provide essential amino acids to their respective hosts. Hongoh et al. (26) found that despite the severe genome erosion of *Em. trichonymphae*, pathways for the biosynthesis of aromatic amino acids are mostly conserved and several key genes are duplicated, which suggests that the endosymbionts provide the flagellate with aromatic amino acids. However, the capacity to synthesize tryptophan, tyrosine, and phenylalanine is not conserved in all lineages of the genus *Endomicrobiellum*. The same is true for biosynthetic pathways for proline and histidine.

All members of the genus lack glutamine synthetase (GlnA), indicating a general dependence of the endosymbionts on the provision of glutamine by the host cell. It is unclear how this is accomplished because all members of *Endomicrobiellum* lack a glutamine transporter (GlnT). Lineage I lost the capacity to synthesize almost all amino acids, suggesting that they have lost their original function and may no longer play an essential role in the symbiosis. This applies also to a putative role of the endosymbionts in the provision of specific vitamins because the distribution of the corresponding biosynthetic pathways among different lineages is not conclusive (Fig. 3).

A function that is frequently encountered among bacterial symbionts of eukaryotic hosts is dinitrogen fixation. Examples among termite gut flagellates are the endosymbiotic *Azobacteroides pseudotriconymphae* (78) and flagellate-associated bacteria in the gut of dry-wood termites (45, 79). Unlike the homolog of nitrogenase group VI, which occurs in other lineages of the phylum *Elusimicrobiota* and is potentially involved in tetrapyrrole modification (63), the group IV homologs encountered in members of the genera *Ectomicrobium* and *Endomicrobium* were shown to encode a functional nitrogenase (25). All other lineages of *Endomicrobiaceae*, including the genus *Endomicrobiellum*, lack nitrogenase and cannot contribute to dinitrogen fixation.

Horizontal gene transfer among flagellate symbionts drives convergent evolution

There is abundant evidence supporting a convergent evolution among the endosymbionts of termite gut flagellates. Significant genome reduction has been reported for endosymbiotic representatives of *Acutalibacteraceae* (53), "*Ancillulaceae*" (21), *Azobacteroidaceae* (78), "*Adiutricaceae*" (19), and *Treponemataceae* (22). The same trend seems to be present also among the putatively endosymbiotic members of *Rickettsiales* and related *Alphaproteobacteria* (orders RUG11792 and UBA3830), *Mycoplasmatales*, and *Opitutales* (families LL51 and UBA9783) (Fig. 7). Notably, endosymbiotic representatives of these lineages not only co-occur in the same termites but also often co-localize in the same flagellate species, which provides ample opportunity for horizontal gene transfer.

Horizontal gene transfer among flagellate symbionts is most evident in the case of sugar phosphate transporters (OPAs), which are present in almost all lineages of established and putative flagellate endosymbionts except intracellular treponemes (Table S7). One homolog (OPA2) of *Endomicrobiellum* clusters among a clade of OPAs from termite-associated MAGs; the other homolog (OPA1) has close relatives among distantly related *Elusimicrobiaceae* associated with arthropods and ruminants but is absent from all other lineages of the *Elusimicrobiota* phylum. Their placement among the radiation of homologs from termite gut MAGs of other phyla suggests that both *Endomicrobiaceae* and *Elusimicrobiaceae* acquired their OPAs independently of each other from the same donor.

Another case of convergent evolution involves the nucleotide antiporters (NTTs), which were independently acquired by *Endomicrobiellum* from putative flagellate endosymbionts in the alphaproteobacterial orders UBA3830 and RUG11792 (see "Genome erosion and acquisition of new functions in *Endomicrobiellum*"). The transporter of RUG11792 clusters also with a homolog from endosymbionts of the family *Acutalibacteraceae* (*Oscillospirales*) (53). All these lineages of endosymbionts have been localized in flagellates of the genus *Trichonympha*, suggesting that the genes encoding NTTs may be frequently transferred among the endosymbionts of termite gut flagellates. By contrast, the NTTs of the putatively endosymbiotic *Opitutales*, which comprise the endonuclear *Candidatus* Nucleococcus from a termite gut flagellate (17) and several ciliate endosymbionts (80), are embedded among lineages from other environments, suggesting an ancestral presence in *Opitutales* (Fig. 6).

A hexuronate pathway occurs in the endosymbiotic genera *Endomicrobiellum*, *Ca. Ancillula*, and *Azobacteroides* (19, 21, 78) but also in other members of the family *Azobacteroidaceae*. Although *Endomicrobiaceae* apparently acquired the pathway from a clade of *Lachnospiraceae* associated with termites and ruminants, "*Ancillulaceae*"

acquired it from *Lactobacillaceae* (Fig. S16). The origin of the hexuronate pathway in several putatively free-living *Endomicrobiaceae* differs from that in the genus *Endomicrobiellum* (Fig. S16). The capacity of these endosymbionts to degrade hexuronates may represent a convergent adaptation to their intracellular niche, because the hydrolysis of hemicelluloses in the digestive vacuoles of their flagellate hosts should yield large amounts of hexuronates (81).

The loss of pathways for the biosynthesis of amino acids in the genus *Endomicrobiellum* was compensated by the acquisition of several amino acid transporters absent from the ancestral, free-living lineages (Fig. 4). In the case of AroP, the homolog from *Endomicrobiellum* originates among homologs of “*Adiutricaceae*” (Fig. S17), whereas those of other *Endomicrobiaceae* cluster with homologs from putative endosymbionts in the order UBA3830. Again, the respective lineages co-colonize the same flagellates (*Trichonympha* spp.; [17]), providing further evidence that the convergent evolution of endosymbionts of termite gut flagellates is driven by horizontal gene transfer.

Conclusions

Our comparative genomic analysis of *Endomicrobiaceae* provided new insights into the evolutionary processes underlying symbiogenesis in intracellular symbionts of termite gut flagellates. The transition from a free-living to an intracellular lifestyle not only restricts gene flow but also affects the ability to take up substrates and other nutrients from the environment. Members of the genus *Endomicrobiellum* possess functions that either represent adaptations to their intracellular niche or serve to compensate for gene losses during ongoing genome erosion. Some functions represent predispositions to an intracellular lifestyle that were present already in the common ancestor of *Endomicrobiaceae*, whereas others were apparently acquired by horizontal gene transfer. The acquisition of similar traits also by other established or putative flagellate symbionts underscores the importance of horizontal gene transfer from other gut microbiota in the convergent evolution of endosymbiotic lineages. The lack of biosynthetic functions common to all members of the genus *Endomicrobiellum* suggests that their original role in the symbiosis became compromised during the progressive genome reduction. It is possible that certain lineages now just represent a metabolic burden to their host. The acquisition of an NTT may indicate the transition from a mutualistic to a parasitic relationship, but it may also be the last straw to ensure survival for a beneficial endosymbiont whose energy metabolism became compromised owing to genome erosion. It is likely that dysfunctional members of *Endomicrobiellum* will be eventually replaced by secondary symbionts that co-occur in the same flagellates, as shown for other primary endosymbionts of eukaryotic hosts that have experienced severe gene losses (1, 82).

Taxonomy

Most members of the family *Endomicrobiaceae* belong to genus-level lineages that are either unclassified or require reclassification. The presence of both high-quality genomes and 16S rRNA gene sequences for most lineages allow the proposal of new taxa under the Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode) (83). The new names and new combinations, together with the designated type material, are listed in Table 1. The authors of previously proposed *Candidatus* names were assigned as descriptors of the corresponding new taxa. The full protocols including etymology, description, and other information are given in the supplemental material (Text S1).

MATERIALS AND METHODS

Metagenome sequencing

Termite colonies collected in the field from various locations were sampled within a week of collection. Samples from termite colonies maintained in other laboratories were processed within a few days after arrival. Termite identification was confirmed by

TABLE 1 New genera in *Endomicrobiaceae* proposed under SeqCode, their designated type species, and other species assigned to the respective genus^a

Genus	Type species	Other species in the genus
<i>Endomicrobium</i> (25)	<i>Endomicrobium proavitum</i> (25)	<i>Endomicrobium embiratermitis</i> sp. nov., <i>E. labiatermitis</i> sp. nov., <i>E. macrotermitis</i> sp. nov., <i>E. neocapritermitis</i> sp. nov., <i>E. procryptotermitis</i> sp. nov.
<i>Endomicrobiellum</i> gen. nov.	<i>Endomicrobiellum trichonymphae</i> sp. nov.	<i>Endomicrobiellum africanum</i> sp. nov., <i>Em. agilis</i> sp. nov., <i>Em. basalitermitum</i> sp. nov., <i>Em. calcaritermitis</i> sp. nov., <i>Em. calonymphae</i> sp. nov., <i>Em. cryptotermitis</i> sp. nov., <i>Em. cubanum</i> sp. nov., <i>Em. dinonymphae</i> sp. nov., <i>Em. glyptotermitis</i> sp. nov., <i>Em. guadaloupense</i> sp. nov., <i>Em. incisermitis</i> sp. nov., <i>Em. mastotermitis</i> sp. nov., <i>Em. meruensis</i> sp. nov., <i>Em. neotermitis</i> sp. nov., <i>Em. porotermitis</i> sp. nov., <i>Em. pyrsonymphae</i> sp. nov., <i>Em. roisinermitis</i> sp. nov., <i>Em. siamense</i> sp. nov.
<i>Ectomicrobium</i> gen. nov.	<i>Ectomicrobium neotermitis</i> sp. nov.	
<i>Parendomicrobium</i> gen. nov.	<i>Parendomicrobium reticulitermitis</i> sp. nov.	<i>Parendomicrobium porotermitis</i> sp. nov.
<i>Proendomicrobium</i> gen. nov.	<i>Proendomicrobium guianensium</i> sp. nov.	
<i>Ruminimicrobium</i> gen. nov.	<i>Ruminimicrobium bovinum</i> sp. nov.	
<i>Ruminimicrobiellum</i> gen. nov.	<i>Ruminimicrobiellum bubulum</i> sp. nov.	<i>Ruminimicrobiellum caprinum</i> sp. nov., <i>R. ovillum</i> sp. nov., <i>R. tauri</i> sp. nov.
<i>Praeruminimicrobium</i> gen. nov.	<i>Praeruminimicrobium purgamenti</i> sp. nov.	
<i>Proruminimicrobium</i> gen. nov.	<i>Proruminimicrobium quisquiliarum</i> sp. nov.	

^aThe nomenclatural types (genomes) of all new species and their full protocols can be found in the supplemental material (Text S1).

sequence analysis of the cytochrome oxidase II genes (84); sequences that were not represented in public databases have been submitted to NCBI GenBank. Details of each sample can be found in Table S1.

Worker termites were dissected with fine-tipped forceps. Gut suspensions of 30–50 individuals were either extracted directly or stored in 2-mL polypropylene tubes containing RNAlater (Invitrogen) at –20°C. DNA was extracted with a bead-beating protocol using the NucleoSpin Soil kit (Macherey-Nagel) as described previously (85). The extracted DNA was quantified with a Qubit fluorometer (Invitrogen), checked for purity with a NanoDrop ND-1000 spectrophotometer (PiqLab), and stored at –20°C.

Metagenomic libraries were sequenced using an Illumina NovaSeq 6000 platform in paired-end mode (2 × 150 bp) (Eurofins, Germany). Sequencing depth was 50 million read pairs for higher termites and 100 million read pairs for lower termites, accounting for a higher proportion of eukaryotic DNA contributed by their gut flagellates. The metagenomic data sets were submitted to the NCBI Sequence Read Archive (SRA) under the BioProject [PRJNA732531](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA732531) (for accession numbers, see Table S1).

Sequence processing and genome recovery

Each of the 48 metagenomes was processed independently. First, we used the illumina-utils library v1.4.1 (86) to remove noise from raw reads with the program iu-filter-quality-minoche (87). Then, we identified host sequences by mapping our reads with Bowtie 2 v2.2.5 (88) against a database composed of the three termite genomes available at the time of the analysis: *Zootermopsis nevadensis* (89), *Cryptotermes secundus* (90), and *Macrotermes natalensis* (91). Subsequently, we removed the host reads by retaining only the unmapped reads using SAMtools v1.7 (92). Then, we assembled the reads of each metagenomic sample with MEGAHIT v1.2.8 (93) using a minimum contig length of 1,000 bp and a list of various *k*-mer sizes (21, 31, 41, 51, 61, 71, 81, 91, 99, 109, and 119). The coverage profile of each contig was computed with the bwa-mem algorithm from BWA v0.7.17 (94). These abundance scores combined with normalized tetranucleotide frequency score were used to bin the assembled reads with MetaBAT v2.14 (95).

The quality of the resulting MAGs was evaluated using CheckM v1.1.1 (96). The MAGs were classified using the Genome Taxonomy Database (GTDB) release 214 (97, 98) and the associated GTDB-Tk v2.3.0 (99, 100). The MAGs have been deposited at the NCBI Assembly Database under the accessions [JAIRJQ0000000000–JAITXW0000000000](#). Detailed information on each MAG is listed in Table S2.

Phylogenomic analysis

To infer phylogenetic relationships between the genomes, we collected a set of *Endomicrobiaceae* genomes and used *Endomicrobiales* genomes as outgroup. After checking the quality of all genomes with CheckM, we used GTDB-Tk to extract the concatenated and filtered alignment of 120 bacterial single-copy marker proteins (99). A maximum-likelihood tree was inferred with IQ-TREE v1.6.11 (101) employing the best-fitting model determined by ModelFinder based on the Bayesian information criterion (102). Branch support was assessed using ultrafast bootstrap approximation (UFBoot) (103, 104) and Shimodaira–Hasegawa approximate likelihood-ratio test (SH-aLRT) (105, 106).

16S rRNA gene phylogeny

A curated alignment of 16S rRNA genes of *Endomicrobiaceae* (11) was expanded with all 16S rRNA gene sequences of *Endomicrobiaceae* (>400 bp) extracted from the MAGs and termite gut metagenomes using Barrnap v0.9 (107). After manual curation, the alignment was exported and used to infer a maximum-likelihood tree using IQ-TREE with the best-fit evolutionary model suggested by ModelFinder; branch support was assessed with SH-aLRT and UFBoot analysis.

Functional annotation

Open reading frames were predicted and annotated for the *Endomicrobia* MAGs and all previously published genomes and genomic scaffolds using Prokka version 1.14.5 (108), with the following settings: -kingdom Bacteria, -increment 1, -mincontiglen 200, -evaluate 1e-06.

For functional annotation, the generated protein files were investigated by Hidden Markov Model (HMM) searching in HMMER version 3.3 using hmmsearch (109) and the HMMs from the Pfam database (releases 31.0, 33.1) (110) and the TIGRFAM database (release 15.0) (111). Transporters and hydrogenases were annotated using the Transporter Classification Database (TCDB) (112) and the hydrogenase database (HydDB) (113). The results were verified using BLASTp version 2.6.0 (114) against a custom database that consisted of the annotated *Endomicrobiaceae* genomes and the SWISS-PROT database (115), which provides manually annotated, non-redundant protein sequences of high quality.

Transposase genes and insertion sequences (IS) were predicted using Issaga (116; http://issaga.biotoul.fr/issaga_index.php). The clustered regularly interspaced palindromic repeats (CRISPR) and intergenic spacer sequences were identified using CRISPR recognition tool (117); the related *cas* genes in the genomes were identified by BLASTp searches against the NCBI non-redundant protein sequence database.

Protein phylogenies

Amino acid sequences of genes were recovered from the *Endomicrobiaceae* MAGs and were expanded with SWISS-PROT entries. The deduced sequences were aligned with MAFFT version 7.427 and were used to calculate a first gene tree in IQTREE using the best-fitting model of evolution suggested by ModelFinder. Node support was assessed with UFBoot (103, 104) and SH-aLRT (105, 106). The alignment and resulting consensus tree were imported into ARB version 6.1 (118), where the alignment was curated manually and exported again to calculate a new tree using the same settings as above.

ACKNOWLEDGMENTS

This study was funded by the Max Planck Society and a grant of the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 987. U.S.M. and T.K. received scholarships from the International Max Planck Research School “Principles of Microbial Life: From molecules to cells, from cells to interactions.”

We are grateful to Rudy Plarre, Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany, for providing termite samples and to Ivan Hrdý, Nathan Lo, James O. Nonoh, Valeria Palma Onetto, Yves Roisin, Rudolf H. Scheffrahn, and Petr Stiblík for sharing material collected in the field. We thank Karen A. Brune for correcting the manuscript.

AUTHOR AFFILIATIONS

¹Research Group Insect Gut Microbiology and Symbiosis, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

²Laboratory of Experimental and Comparative Ethology LEEC, UR 4443, University Sorbonne Paris Nord, Villetaneuse, France

³Faculty of Tropical AgriSciences, Czech University of Life Sciences, Prague, Czechia

⁴Biology Centre, Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czechia

AUTHOR ORCIDs

Undine S. Mies  <http://orcid.org/0000-0002-4879-7325>

Vincent Hervé  <http://orcid.org/0000-0002-3495-561X>

Katja Platt  <http://orcid.org/0000-0001-5093-6155>

David Sillam-Dussès  <http://orcid.org/0000-0001-5027-8703>

Jan Šobotník  <http://orcid.org/0000-0002-8581-637X>

Andreas Brune  <http://orcid.org/0000-0002-2667-4391>

FUNDING

Funder	Grant(s)	Author(s)
Max-Planck-Institut für Terrestrische Mikrobiologie (MPI for Terrestrial Microbiology)		Andreas Brune
Deutsche Forschungsgemeinschaft (DFG)	SFB 987	Andreas Brune

DIRECT CONTRIBUTION

This article is a direct contribution from Andreas Brune, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Yuichi Hongoh, Tokyo Institute of Technology, and Jared Leadbetter, California Institute of Technology.

DATA AVAILABILITY

The metagenomic data sets were submitted to the NCBI Sequence Read Archive (SRA) under the BioProject [PRJNA732531](#) (for accession numbers, see Table S1). The MAGs have been deposited at the NCBI Assembly Database under the accessions [JAIRJQ000000000–JAITXW000000000](#).

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Text S1 (mBio00826-24-s0001.pdf). Protocols for new prokaryotic taxa. Etymology and description of the new taxa proposed under SeqCode.

Supplemental Figures (mBio00826-24-s0002.pdf). Fig. S1–S17.

Fig. S15 (mBio00826-24-s0003.html). Interactive HTML file.

Supplemental Tables (mBio00826-24-s0004.xlsx). Tables S1–S7.

REFERENCES

- McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26. <https://doi.org/10.1038/nrmicro2670>
- Wernegreen JJ. 2015. Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann N Y Acad Sci* 1360:16–35. <https://doi.org/10.1111/nyas.12740>
- George EE, Husnik F, Tashyreva D, Prokopchuk G, Horák A, Kwong WK, Lukeš J, Keeling PJ. 2020. Highly reduced genomes of protist endosymbionts show evolutionary convergence. *Curr Biol* 30:925–933. <https://doi.org/10.1016/j.cub.2019.12.070>
- Husnik F, Tashyreva D, Boscaro V, George EE, Lukeš J, Keeling PJ. 2021. Bacterial and archaeal symbioses with protists. *Curr Biol* 31:R862–R877. <https://doi.org/10.1016/j.cub.2021.05.049>
- Stingl U, Radek R, Yang H, Brune A. 2005. “*Endomicrobia*”: cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol* 71:1473–1479. <https://doi.org/10.1128/AEM.71.3.1473-1479.2005>
- Ikeda-Ohtsubo W, Desai M, Stingl U, Brune A. 2007. Phylogenetic diversity of “*Endomicrobia*” and their specific affiliation with termite gut flagellates. *Microbiology (Reading)* 153:3458–3465. <https://doi.org/10.1099/mic.0.2007/009217-0>
- Ohkuma M, Sato T, Noda S, Ui S, Kudo T, Hongoh Y. 2007. The candidate phylum ‘Termite group 1’ of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol Ecol* 60:467–476. <https://doi.org/10.1111/j.1574-6941.2007.00311.x>
- Wong MT, Wang W, Lacourt M, Couturier M, Edwards EA, Master ER. 2016. Substrate-driven convergence of the microbial community in lignocellulose-amended enrichments of gut microflora from the Canadian beaver (*Castor canadensis*) and North American moose (*Alces americanus*). *Front Microbiol* 7:961. <https://doi.org/10.3389/fmicb.2016.00961>
- Levy B, Jami E. 2018. Exploring the prokaryotic community associated with the rumen ciliate protozoa population. *Front Microbiol* 9:2526. <https://doi.org/10.3389/fmicb.2018.02526>
- Ikeda-Ohtsubo W, Faivre N, Brune A. 2010. Putatively free-living ‘*Endomicrobia*’-ancestors of the intracellular symbionts of termite gut flagellates? *Environ Microbiol Rep* 2:554–559. <https://doi.org/10.1111/j.1758-2229.2009.00124.x>
- Mikaelyan A, Thompson CL, Meuser K, Zheng H, Rani P, Plarre R, Brune A. 2017. High-resolution phylogenetic analysis of *Endomicrobia* reveals multiple acquisitions of endosymbiotic lineages by termite gut flagellates. *Environ Microbiol Rep* 9:477–483. <https://doi.org/10.1111/1758-2229.12565>
- Ohkuma M, Brune A. 2011. Diversity, structure, and evolution of the termite gut microbial community, p 413–438. In *Biology of termites: a modern synthesis*. Springer, Netherlands.
- Brune A. 2014. Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol* 12:168–180. <https://doi.org/10.1038/nrmicro3182>
- Brune A, Dietrich C. 2015. The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. *Annu Rev Microbiol* 69:145–166. <https://doi.org/10.1146/annurev-micro-092412-155715>
- Hongoh Y. 2011. Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell Mol Life Sci* 68:1311–1325. <https://doi.org/10.1007/s00018-011-0648-z>
- Hongoh Y, Sato T, Dolan MF, Noda S, Ui S, Kudo T, Ohkuma M. 2007. The motility symbiont of the termite gut flagellate *Caduceia versatilis* is a member of the “*Synergistes*” group. *Appl Environ Microbiol* 73:6270–6276. <https://doi.org/10.1128/AEM.00750-07>
- Sato T, Kuwahara H, Fujita K, Noda S, Kihara K, Yamada A, Ohkuma M, Hongoh Y. 2014. Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. *ISME J* 8:1008–1019. <https://doi.org/10.1038/ismej.2013.222>
- Hongoh Y, Ohkuma M. 2010. Termite gut flagellates and their methanogenic and eubacterial symbionts, p 55–79. In Hackstein JHP (ed), (Endo)symbiotic methanogenic archaea. Springer International Publishing, Berlin, Heidelberg.
- Ikeda-Ohtsubo W, Strassert JFH, Köhler T, Mikaelyan A, Gregor I, McHardy AC, Tringe SG, Hugenholtz P, Radek R, Brune A. 2016. “*Candidatus* *Adiutrix intracellularis*”, an endosymbiont of termite gut flagellates, is the first representative of a deep-branching clade of *Deltaproteobacteria* and a putative homoacetogen. *Environ Microbiol* 18:2548–2564. <https://doi.org/10.1111/1462-2920.13234>
- Kuwahara H, Yuki M, Izawa K, Ohkuma M, Hongoh Y. 2017. Genome of “*Ca. Desulfovibrio trichonymphae*”, an H₂-oxidizing bacterium in a tripartite symbiotic system within a protist cell in the termite gut. *ISME J* 11:766–776. <https://doi.org/10.1038/ismej.2016.143>
- Strassert JFH, Mikaelyan A, Woyke T, Brune A. 2016. Genome analysis of ‘*Candidatus* *Ancillula trichonymphae*’, first representative of a deep-branching clade of *Bifidobacteriales*, strengthens evidence for convergent evolution in flagellate endosymbionts. *Environ Microbiol Rep* 8:865–873. <https://doi.org/10.1111/1758-2229.12451>
- Ohkuma M, Noda S, Hattori S, Iida T, Yuki M, Starns D, Inoue J, Darby AC, Hongoh Y. 2015. Acetogenesis from H₂ plus CO₂ and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *Proc Natl Acad Sci U S A* 112:10224–10230. <https://doi.org/10.1073/pnas.1423979112>
- Desai MS, Strassert JFH, Meuser K, Hertel H, Ikeda-Ohtsubo W, Radek R, Brune A. 2010. Strict cospeciation of devescoviniid flagellates and *Bacteroidales* ectosymbionts in the gut of dry-wood termites (Kalotermitidae). *Environ Microbiol* 12:2120–2132. <https://doi.org/10.1111/j.1462-2920.2009.02080.x>
- Izawa K, Kuwahara H, Sugaya K, Lo N, Ohkuma M, Hongoh Y. 2017. Discovery of ectosymbiotic *Endomicrobium* lineages associated with protists in the gut of stolonitermitid termites. *Environ Microbiol Rep* 9:411–418. <https://doi.org/10.1111/1758-2229.12549>
- Zheng H, Dietrich C, Radek R, Brune A. 2016. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a group IV nitrogenase. *Environ Microbiol* 18:191–204. <https://doi.org/10.1111/1462-2920.12960>
- Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Complete genome of the uncultured termite group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A* 105:5555–5560. <https://doi.org/10.1073/pnas.0801389105>
- Izawa K, Kuwahara H, Kihara K, Yuki M, Lo N, Itoh T, Ohkuma M, Hongoh Y. 2016. Comparison of intracellular “*Ca. Endomicrobium trichonymphae*” genomovars illuminates the requirement and decay of defense systems against foreign DNA. *Genome Biol Evol* 8:3099–3107. <https://doi.org/10.1093/gbe/evw227>
- Stephens ME, Benjamino J, Graf J, Gage DJ. 2022. Simultaneous single-cell genome and transcriptome sequencing of termite hindgut protists reveals metabolic and evolutionary traits of their endosymbionts. *mSphere* 7:e0002122. <https://doi.org/10.1128/msphere.00021-22>
- Zheng H, Dietrich C, Brune A. 2017. Genome analysis of *Endomicrobium proavitum* suggests loss and gain of relevant functions during the evolution of intracellular symbionts. *Appl Environ Microbiol* 83:e00656-17. <https://doi.org/10.1128/AEM.00656-17>

30. Hervé V, Liu P, Dietrich C, Sillam-Dussès D, Stiblik P, Šobotník J, Brune A. 2020. Phylogenomic analysis of 589 metagenome-assembled genomes encompassing all major prokaryotic lineages from the gut of higher termites. *PeerJ* 8:e8614. <https://doi.org/10.7717/peerj.8614>
31. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>
32. Stewart RD, Auffret MD, Warr A, Wisner AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat Commun* 9:870. <https://doi.org/10.1038/s41467-018-03317-6>
33. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. 2019. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol* 37:953–961. <https://doi.org/10.1038/s41587-019-0202-3>
34. Gharechahi J, Vahidi MF, Bahram M, Han JL, Ding XZ, Salekdeh GH. 2021. Metagenomic analysis reveals a dynamic microbiome with diversified adaptive functions to utilize high lignocellulosic forages in the cattle rumen. *ISME J* 15:1108–1120. <https://doi.org/10.1038/s41396-020-00837-2>
35. Ewens SD, Gomberg AFS, Barnum TP, Borton MA, Carlson HK, Wrighton KC, Coates JD. 2021. The diversity and evolution of microbial dissimilatory phosphite oxidation. *Proc Natl Acad Sci U S A* 118:e2020024118. <https://doi.org/10.1073/pnas.2020024118>
36. Protasov E, Nonoh JO, Kästle Silva JM, Mies US, Hervé V, Dietrich C, Lang K, Mikulski L, Platt K, Poehlein A, Köhler-Ramm T, Miambi E, Boga HI, Feldewert C, Ngugi DK, Plarre R, Sillam-Dussès D, Šobotník J, Daniel R, Brune A. 2023. Diversity and taxonomic revision of methanogens and other archaea in the intestinal tract of terrestrial arthropods. *Front Microbiol* 14:1281628. <https://doi.org/10.3389/fmicb.2023.1281628>
37. Schwöppe C, Winkler HH, Neuhaus HE. 2002. Properties of the glucose-6-phosphate transporter from *Chlamydia pneumoniae* (HPTcp) and the glucose-6-phosphate sensor from *Escherichia coli* (UhpC). *J Bacteriol* 184:2108–2115. <https://doi.org/10.1128/JB.184.8.2108-2115.2002>
38. Audia JP, Winkler HH. 2006. Study of the five *Rickettsia prowazekii* proteins annotated as ATP/ADP translocases (Tlc): only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. *J Bacteriol* 188:6261–6268. <https://doi.org/10.1128/JB.00371-06>
39. Trentmann O, Decker C, Winkler HH, Neuhaus HE. 2000. Charged amino-acid residues in transmembrane domains of the plastidic ATP/ADP transporter from *Arabidopsis* are important for transport efficiency, substrate specificity, and counter exchange properties. *Eur J Biochem* 267:4098–4105. <https://doi.org/10.1046/j.1432-1033.2000.01468.x>
40. Mertens E. 1991. Pyrophosphate-dependent phosphofructokinase, an anaerobic glycolytic enzyme? *FEBS Lett* 285:1–5. [https://doi.org/10.1016/0014-5793\(91\)80711-b](https://doi.org/10.1016/0014-5793(91)80711-b)
41. Cooper RA. 1984. Metabolism of methylglyoxal in microorganisms. *Annu Rev Microbiol* 38:49–68. <https://doi.org/10.1146/annurev.mi.38.100184.000405>
42. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM, Morales SE. 2016. Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. *ISME J* 10:761–777. <https://doi.org/10.1038/ismej.2015.153>
43. Jongsareejit B, Rahman RN, Fujiwara S, Imanaka T. 1997. Gene cloning, sequencing and enzymatic properties of glutamate synthase from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1. *Mol Gen Genet* 254:635–642. <https://doi.org/10.1007/s004380050461>
44. Min B, Pelaschier JT, Graham DE, Tumbula-Hansen D, Söll D. 2002. Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation. *Proc Natl Acad Sci U S A* 99:2678–2683. <https://doi.org/10.1073/pnas.012027399>
45. Desai MS, Brune A. 2012. Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. *ISME J* 6:1302–1313. <https://doi.org/10.1038/ismej.2011.194>
46. Yamada A, Inoue T, Noda S, Hongoh Y, Ohkuma M. 2007. Evolutionary trend of phylogenetic diversity of nitrogen fixation genes in the gut community of wood-feeding termites. *Mol Ecol* 16:3768–3777. <https://doi.org/10.1111/j.1365-294X.2007.03326.x>
47. Lin S, Cronan JE. 2011. Closing in on complete pathways of biotin biosynthesis. *Mol Biosyst* 7:1811–1821. <https://doi.org/10.1039/c1mb05022b>
48. Geissinger O, Herlemann DPR, Mörschel E, Maier UG, Brune A. 2009. The ultramicrobacterium “*Elusimicrobium minutum*” gen. nov., sp. nov., the first cultivated representative of the termite group 1 phylum. *Appl Environ Microbiol* 75:2831–2840. <https://doi.org/10.1128/AEM.02697-08>
49. Brune A, Mies US. 2023. *Elusimicrobiota*, p 1–5. In Whitman W (ed), *Bergey’s manual of systematics of archaea and bacteria*, 1st ed. Wiley.
50. Moran NA, Bennett GM. 2014. The tiniest tiny genomes. *Annu Rev Microbiol* 68:195–215. <https://doi.org/10.1146/annurev-micro-091213-112901>
51. Zheng H, Dietrich C, Hongoh Y, Brune A. 2016. Restriction-modification systems as mobile genetic elements in the evolution of an intracellular symbiont. *Mol Biol Evol* 33:721–725. <https://doi.org/10.1093/molbev/msv264>
52. Arora J, Kinjo Y, Šobotník J, Buček A, Clitheroe C, Stiblik P, Roisin Y, Žifčáková L, Park YC, Kim KY, Sillam-Dussès D, Hervé V, Lo N, Tokuda G, Brune A, Bourguignon T. 2022. The functional evolution of termite gut microbiota. *Microbiome* 10:78. <https://doi.org/10.1186/s40168-022-01258-3>
53. Takahashi K, Kuwahara H, Horikawa Y, Izawa K, Kato D, Inagaki T, Yuki M, Ohkuma M, Hongoh Y. 2023. Emergence of putative energy parasites within *Clostridia* revealed by genome analysis of a novel endosymbiotic clade. *ISME J* 17:1895–1906. <https://doi.org/10.1038/s41396-023-01502-0>
54. Ohkuma M, Kudo T. 1996. Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Appl Environ Microbiol* 62:461–468. <https://doi.org/10.1128/aem.62.2.461-468.1996>
55. Hongoh Y, Ohkuma M, Kudo T. 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiol Ecol* 44:231–242. [https://doi.org/10.1016/S0168-6496\(03\)00026-6](https://doi.org/10.1016/S0168-6496(03)00026-6)
56. Shelomi M, Lin S-S, Liu L-Y. 2019. Transcriptome and microbiome of coconut rhinoceros beetle (*Oryctes rhinoceros*) larvae. *BMC Genomics* 20:957. <https://doi.org/10.1186/s12864-019-6352-3>
57. Dietrich C, Köhler T, Brune A. 2014. The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl Environ Microbiol* 80:2261–2269. <https://doi.org/10.1128/AEM.04206-13>
58. Bourguignon T, Lo N, Dietrich C, Šobotník J, Sidek S, Roisin Y, Brune A, Evans TA. 2018. Rampant host switching shaped the termite gut microbiome. *Curr Biol* 28:649–654. <https://doi.org/10.1016/j.cub.2018.01.035>
59. Domínguez-Santos R, Pérez-Cobas AE, Cuti P, Pérez-Brocal V, García-Ferris C, Moya A, Latorre A, Gil R. 2021. Interkingdom gut microbiome and resistome of the cockroach *Blattella germanica*. *mSystems* 6:e01213-20. <https://doi.org/10.1128/mSystems.01213-20>
60. Dukes HE, Tinker KA, Ottesen EA. 2023. Disentangling hindgut metabolism in the American cockroach through single-cell genomics and metatranscriptomics. *Front Microbiol* 14:1156809. <https://doi.org/10.3389/fmicb.2023.1156809>
61. Strassert JFH, Desai MS, Radek R, Brune A. 2010. Identification and localization of the multiple bacterial symbionts of the termite gut flagellate *Joenia annectens*. *Microbiology (Reading)* 156:2068–2079. <https://doi.org/10.1099/mic.0.037267-0>
62. Xie F, Jin W, Si H, Yuan Y, Tao Y, Liu J, Wang X, Yang C, Li Q, Yan X, Lin L, Jiang Q, Zhang L, Guo C, Greening C, Heller R, Guan LL, Pope PB, Tan Z, Zhu W, Wang M, Qiu Q, Li Z, Mao S. 2021. An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome* 9:137. <https://doi.org/10.1186/s40168-021-01078-x>
63. Méheust R, Castelle CJ, Matheus Carnevali PB, Farag IF, He C, Chen L-X, Amano Y, Hug LA, Banfield JF. 2020. Groundwater *Elusimicrobia* are metabolically diverse compared to gut microbiome *Elusimicrobia* and some have a novel nitrogenase paralog. *ISME J* 14:2907–2922. <https://doi.org/10.1038/s41396-020-0716-1>

64. Passalacqua KD, Charbonneau M-E, O’Riordan MXD. 2016. Bacterial metabolism shapes the host-pathogen interface. *Microbiol Spectr* 4:4. <https://doi.org/10.1128/microbiolspec.VMBF-0027-2015>
65. Dharamshi JE, Köstlbacher S, Schön ME, Collingro A, Ettema TJG, Horn M. 2023. Gene gain facilitated endosymbiotic evolution of Chlamydiae. *Nat Microbiol* 8:40–54. <https://doi.org/10.1038/s41564-022-01284-9>
66. Treitli SC, Kolisko M, Husnik F, Keeling PJ, Hampl V. 2019. Revealing the metabolic capacity of *Streblomastix strix* and its bacterial symbionts using single-cell metagenomics. *Proc Natl Acad Sci U S A* 116:19675–19684. <https://doi.org/10.1073/pnas.1910793116>
67. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, Thomas BC, Banfield JF. 2013. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *mBio* 4:e00708-13. <https://doi.org/10.1128/mBio.00708-13>
68. McLean JS, Bor B, Kerns KA, Liu Q, To TT, Solden L, Hendrickson EL, Wrighton K, Shi W, He X. 2020. Acquisition and adaptation of ultra-small parasitic reduced genome bacteria to mammalian hosts. *Cell Rep* 32:107939. <https://doi.org/10.1016/j.celrep.2020.107939>
69. Hackmann TJ, Ngugi DK, Firkins JL, Tao J. 2017. Genomes of rumen bacteria encode atypical pathways for fermenting hexoses to short-chain fatty acids. *Environ Microbiol* 19:4670–4683. <https://doi.org/10.1111/1462-2920.13929>
70. Kelly WJ, Leahy SC, Altermann E, Yeoman CJ, Dunne JC, Kong Z, Pacheco DM, Li D, Noel SJ, Moon CD, Cookson AL, Attwood GT. 2010. The glycobiome of the rumen bacterium *Butyrivibrio proteoclasticus* B316T highlights adaptation to a polysaccharide-rich environment. *PLoS One* 5:e11942. <https://doi.org/10.1371/journal.pone.0011942>
71. Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A. 2009. Genomic analysis of “*Elusimicrobium minutum*,” the first cultivated representative of the phylum “*Elusimicrobia*” (formerly termite group 1). *Appl Environ Microbiol* 75:2841–2849. <https://doi.org/10.1128/AEM.02698-08>
72. Kengen SWM, Stams AJM. 1994. Formation of L-alanine as a reduced end product in carbohydrate fermentation by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Arch Microbiol* 161:168–175. <https://doi.org/10.1007/BF00276479>
73. Heider J, Mai X, Adams MWW. 1996. Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. *J Bacteriol* 178:780–787. <https://doi.org/10.1128/jb.178.3.780-787.1996>
74. Schmitz-Esser S, Linka N, Collingro A, Beier CL, Neuhaus HE, Wagner M, Horn M. 2004. ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. *J Bacteriol* 186:683–691. <https://doi.org/10.1128/JB.186.3.683-691.2004>
75. Major P, Embley TM, Williams TA. 2017. Phylogenetic diversity of NTT nucleotide transport proteins in free-living and parasitic bacteria and eukaryotes. *Genome Biol Evol* 9:480–487. <https://doi.org/10.1093/gbe/evx015>
76. Claessen D, Errington J. 2019. Cell wall deficiency as a coping strategy for stress. *Trends Microbiol* 27:1025–1033. <https://doi.org/10.1016/j.tim.2019.07.008>
77. Sugawara K, Toyoda H, Kimura M, Hayasaka S, Saito H, Kobayashi H, Ihara K, Ida T, Akaike T, Ando E, Hyodo M, Hayakawa Y, Hamamoto S, Uozumi N. 2021. Loss of cell wall integrity genes *cpxA* and *mrcB* causes flocculation in *Escherichia coli*. *Biochem J* 478:41–59. <https://doi.org/10.1042/BCJ20200723>
78. Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Genome of an endosymbiotic coupling N₂ fixation to cellulolysis within protist cells in termite gut. *Science* 322:1108–1109. <https://doi.org/10.1126/science.1165578>
79. Noda S, Ohkuma M, Kudo T. 2002. Nitrogen fixation genes expressed in the symbiotic microbial community in the gut of the termite *Coptotermes formosanus*. *Microb Environ* 17:139–143. <https://doi.org/10.1264/jisme.2.17.139>
80. Serra V, Gammuto L, Nitla V, Castelli M, Lanzoni O, Sasseria D, Bandi C, Sandeep BV, Verni F, Modeo L, Petroni G. 2020. Morphology, ultrastructure, genomics, and phylogeny of *Euplotes vanleeuwenhoekii* sp. nov. and its ultra-reduced endosymbiont “*Candidatus Pinguicoccus supinus*” sp. nov. *Sci Rep* 10:20311. <https://doi.org/10.1038/s41598-020-76348-z>
81. de Vries RP, Visser J. 2001. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiol Mol Biol Rev* 65:497–522. <https://doi.org/10.1128/MMBR.65.4.497-522.2001>
82. Perreau J, Moran NA. 2022. Genetic innovations in animal–microbe symbioses. *Nat Rev Genet* 23:23–39. <https://doi.org/10.1038/s41576-021-00395-z>
83. Hedlund BP, Chuvochina M, Hugenholtz P, Konstantinidis KT, Murray AE, Palmer M, Parks DH, Probst AJ, Reysenbach A-L, Rodriguez-R LM, Rossello-Mora R, Sutcliffe IC, Venter SN, Whitman WB. 2022. SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat Microbiol* 7:1702–1708. <https://doi.org/10.1038/s41564-022-01214-9>
84. Pester M, Brune A. 2006. Expression profiles of *fhs* (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ Microbiol* 8:1261–1270. <https://doi.org/10.1111/j.1462-2920.2006.01020.x>
85. Rossmassler K, Dietrich C, Thompson C, Mikaelyan A, Nonoh JO, Scheffrahn RH, Sillam-Dussès D, Brune A. 2015. Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites. *Microbiome* 3:56. <https://doi.org/10.1186/s40168-015-0118-1>
86. Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high quality short reads using Illumina paired-end technology. *PLoS One* 8:e66643. <https://doi.org/10.1371/journal.pone.0066643>
87. Minoche AE, Dohm JC, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* 12:1–15. <https://doi.org/10.1186/gb-2011-12-11-r112>
88. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
89. Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, Chen Z, Childers CP, Glastad KM, Gokhale K, et al. 2014. Molecular traces of alternative social organization in a termite genome. *Nat Commun* 5:1–12. <https://doi.org/10.1038/ncomms4636>
90. Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feidel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S, et al. 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol* 2:557–566. <https://doi.org/10.1038/s41559-017-0459-1>
91. Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, Nygaard S, Nobre T, Klaubauf S, Schindler PM, Hauser F, Pan H, Yang Z, Sonnenberg ASM, de Beer ZW, Zhang Y, Wingfield MJ, Gimmelikhuijzen CJP, de Vries RP, Korb J, Aanen DK, Wang J, Boomsma JJ, Zhang G. 2014. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc Natl Acad Sci U S A* 111:14500–14505. <https://doi.org/10.1073/pnas.1319718111>
92. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
93. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
94. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
95. Kang DD, Li F, Kirtan E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
96. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
97. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>
98. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete

- genome-based taxonomy. *Nucleic Acids Res* 50:D785–D794. <https://doi.org/10.1093/nar/gkab776>
99. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
 100. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38:5315–5316. <https://doi.org/10.1093/bioinformatics/btac672>
 101. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
 102. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
 103. Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30:1188–1195. <https://doi.org/10.1093/molbev/mst024>
 104. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>
 105. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>
 106. Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* 55:539–552. <https://doi.org/10.1080/10635150600755453>
 107. Seemann T. 2013. barrnap 0.9: rapid ribosomal RNA prediction. GitHub repository. <https://github.com/tseemann/barrnap>.
 108. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
 109. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
 110. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>
 111. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res* 31:371–373. <https://doi.org/10.1093/nar/gkg128>
 112. Saier MH, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. 2016. The transporter classification database (TCDB): recent advances. *Nucleic Acids Res* 44:D372–D379. <https://doi.org/10.1093/nar/gkv1103>
 113. Søndergaard D, Pedersen CNS, Greening C. 2016. HydDB: a web tool for hydrogenase classification and analysis. *Sci Rep* 6:34212. <https://doi.org/10.1038/srep34212>
 114. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
 115. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilboud S, Schneider M. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31:365–370. <https://doi.org/10.1093/nar/gkg095>
 116. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* 12:R30. <https://doi.org/10.1186/gb-2011-12-3-r30>
 117. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpidis NC, Hugenholtz P. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. <https://doi.org/10.1186/1471-2105-8-209>
 118. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Buchner A, Lai T, Steppi S, Jobb G, et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res* 32:1363–1371. <https://doi.org/10.1093/nar/gkh293>