AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

EVOLUTIONARY AND GENOMIC MICROBIOLOGY

Genome Analysis of *Endomicrobium proavitum* Suggests Loss and Gain of Relevant Functions during the Evolution of Intracellular Symbionts

Hao Zheng,* Carsten Dietrich,* Andreas Brune

Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

ABSTRACT Bacterial endosymbionts of eukaryotes show progressive genome erosion, but detailed investigations of the evolutionary processes involved in the transition to an intracellular lifestyle are generally hampered by the lack of extant free-living lineages. Here, we characterize the genome of the recently isolated, free-living *Endomicrobium proavitum*, the second member of the *Elusimicrobia* phylum brought into pure culture, and compare it to the closely related "*Candidatus* Endomicrobium trichonymphae" strain Rs-D17, a previously described but uncultured endosymbiont of termite gut flagellates. A reconstruction of the metabolic pathways of *Endomicrobium proavitum* matched the fermentation products formed in pure culture and underscored its restriction to glucose as the substrate. However, several pathways present in the free-living strain, e.g., for the uptake and activation of glucose and its subsequent fermentation, ammonium assimilation, and outer membrane biogenesis, were absent or disrupted in the endosymbiont, probably lost during the massive genome rearrangements that occurred during symbiogenesis. While the majority of the genes in strain Rs-D17 have orthologs in *Endomicrobium proavitum*, the endosymbiont also possesses a number of functions that are absent from the free-living strain and may represent adaptations to the intracellular lifestyle. Phylogenetic analysis revealed that the genes encoding glucose 6-phosphate and amino acid transporters, acetaldehyde/alcohol dehydrogenase, and the pathways of glucuronic acid catabolism and thiamine pyrophosphate biosynthesis were either acquired by horizontal gene transfer or may represent ancestral traits that were lost in the free-living strain. The polyphyletic origin of *Endomicrobia* in different flagellate hosts makes them excellent models for future studies of convergent and parallel evolution during symbiogenesis.

IMPORTANCE The isolation of a free-living relative of intracellular symbionts provides the rare opportunity to identify the evolutionary processes that occur in the course of symbiogenesis. Our study documents that the genome of "*Candidatus* Endomicrobium trichonymphae," which represents a clade of endosymbionts that have coevolved with termite gut flagellates for more than 40 million years, is not simply a subset of the genes present in *Endomicrobium proavitum*, a member of the ancestral, free-living lineage. Rather, comparative genomics revealed that the endosymbionts possess several relevant functions that were either prerequisites for colonization of the intracellular habitat or might have served to compensate for genes losses that occurred during genome erosion. Some gene sets found only in the endosymbiont were apparently acquired by horizontal transfer from other gut bacteria, which suggests that the intracellular bacteria of flagellates are not entirely cut off from gene flow.

KEYWORDS endomicrobia, endosymbionts, genome reduction, termite

Received 17 March 2017 **Accepted** 20 June 2017

Accepted manuscript posted online 23 June 2017

Citation 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.

Editor Harold L. Drake, University of Bayreuth

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

* Present address: Hao Zheng, Department of Integrative Biology, University of Texas at Austin, Austin, Texas, USA; Carsten Dietrich, Strategy and Innovation Technology Center, Siemens Healthcare GmbH, Erlangen, Germany.

Many eukaryotes are engaged in mutualistic symbioses with intracellular bacteria. The transition from a free-living to an intracellular lifestyle is characterized by a progressive genome erosion that increases with the age of the symbiosis (1). While there are numerous studies on the process of genome reduction in obligate intracellular symbionts of insect tissues (2), little is known about the evolutionary events in the endosymbionts of protists.

Cellulolytic flagellates are essential actors in the symbiotic digestion of lignocellulose in lower termites (3). Most flagellate species harbor specific assemblages of intracellular bacteria, which are considered to play an important role in the nutrition of their respective hosts (4, 5). A particularly abundant group of flagellate endosymbionts are *Endomicrobia*. They represent a deep-branching clade in the *Elusimicrobia* phylum that occurs almost exclusively in the guts of animals, particularly termites and cockroaches (6–8).

There are several branches of *Endomicrobia* that colonize different flagellates and were probably derived from ancestral lineages of free-living gut bacteria (7, 9). In the case of "*Candidatus* Endomicrobium trichonymphae," the symbionts were acquired by an ancestral *Trichonympha* species long after the symbiosis between termites and flagellates had been established (10). Homogeneous populations of endosymbionts in each host cell (11) and strict codiversification with their flagellate host (10) are indicative of a strictly vertical mode of transmission.

The genome of "*Ca*. Endomicrobium trichonymphae" strain Rs-D17, the endosymbiont of *Trichonympha agilis*, is relatively small (1.1 Mbp) and has experienced severe genome erosion, which caused the loss or interruption by pseudogenes of numerous biosynthetic pathways (12). However, owing to the lack of genomic information on other *Endomicrobia*, in particular free-living relatives, one could only speculate on the evolutionary events that that occurred during symbiogenesis.

We recently isolated *Endomicrobium proavitum*, the first member of the class *Endomicrobia*, from the gut of the termite *Reticulitermes santonensis* (13). A detailed ultrastructural and physiological characterization identified it as a strictly anaerobic ultramicrobacterium that ferments glucose in a mixed-acid fermentation, resembling in most properties the distantly related *Elusimicrobium minutum*, the only other isolate of the *Elusimicrobia* phylum (14). The free-living *En. proavitum* belongs to the same genus as "*Ca*. Endomicrobium trichonymphae" (>95% 16S rRNA sequence similarity to strain Rs-D17) (13) but is separated from its endosymbiotic relative by at least 40 to 70 million years of evolution (10).

Here, we present a detailed genomic analysis of *En. proavitum*, whose complete genome has been recently sequenced (15), and compare it to the genome of strain Rs-D17, focusing on apparent differences in the metabolism of the endosymbiont. In addition, we identified differences in the coding sequences between the genomes to identify gene functions that involved in the transition to an intracellular lifestyle and characterized the phylogenetic origin of the respective genes to differentiate between ancient traits that were present already in the free-living ancestor or acquired by horizontal gene transfer.

RESULTS

Genome features. With less than 1.6 Mbp, the genome of *En. proavitum* is at the lower end of the size range reported for free-living bacteria. Genome size and G+C content are almost identical to the corresponding values for the distantly related *Elusimicrobium minutum* (16), the only other isolate of the *Elusimicrobia* phylum, but considerably larger than those for the endosymbiotic strain Rs-D17 (Table 1). The density of protein-coding genes in *En. proavitum* (90.0%) is much higher than in the closely related endosymbiont (66.5%), which accumulated an enormous number of pseudogenes in its genome (12). All representatives of *Elusimicrobia* contain only a single rRNA operon, which is in agreement with the low growth rates of the isolates (13, 14).

TABLE 1 Comparison of genomic features of *En. proavitum*, the closely related but endosymbiotic "*Ca*. E. trichonymphae," and their distant relative, *El. minutum*[a]

Organism	GenBank accession no.	Genome size (bp)	G+C content (mol%)	CDS[a]	No. of genes or operons			Avg CDS length (bp)
					Pseudogenes	tRNA genes	rRNA operons	
Class *Endomicrobia*								
En. proavitum	CP009498	1,588,979	39.3	1,344	19	46	1	1,066
"*Ca*. E. trichonymphae" strain Rs-D17	AP009510	1,125,857	35.2	761	121	45	1	984
Class *Elusimicrobia*								
El. minutum	CP001055	1,643,562	39.9	1,529	20	45	1	951

[a]*En. proavitum*, "*Ca*. E. trichonymphae" (12), and *El. minutum* (16) are the only representatives of the *Elusimicrobia* phylum with sequenced genomes. All genomes are on a circular chromosome and have been completely sequenced. CDS, coding sequences, excluding pseudogenes.

Comparative genome analysis revealed that the majority (82%) of the genes in strain Rs-D17 have orthologs in *En. proavitum*, with an average of 61% amino acid identity between the orthologs. However, 110 of the genes in the genome of strain Rs-D17 were not detected in *En. proavitum* (see Data Set S2 in the supplemental material). Besides the restriction-modification and CRISPR-Cas systems already reported in a previous study (17), these genes encode several metabolic functions that may be relevant for the intracellular lifestyle. Conversely, less than half of the coding sequences of *En. proavitum* (668 genes) have orthologs (including pseudogenes) in strain Rs-D17; the remaining genes were apparently lost by the endosymbiont (see Data Set S1 in the supplemental material).

Fermentation pathways. *En. proavitum* possesses a purely fermentative energy metabolism and converts glucose exclusively to lactate, acetate, hydrogen, and CO_2 (13). The metabolic pathways reconstructed from the annotated genome (Fig. 1) are in full agreement with the fermentation products formed in pure culture and underscore the restriction of *En. proavitum* to glucose as the substrate (13).

En. proavitum possesses two possible mechanisms for the uptake and activation of glucose: a putative ABC sugar transporter (CUT1 family) and a phosphotransferase system (PTS) of the Man family. The gene cluster that encodes the PTS components also comprises an ortholog of glycoside hydrolase family 57, which contains both α-amylases and α-mannosidases. However, we found no growth on mannose and fructose when we retested these substrates under previously reported conditions (13). Orthologs of the ABC transporter cassette and glucokinase (*glcK*) are absent from the genome of strain Rs-D17, and the gene encoding the regulatory protein kinase (*hprK*) of the PTS is pseudogenized (12). Conversely, homologs of the *uhpC* gene, which encode a putative glucose 6-phosphate (G6P) transporter in strain Rs-D17, are absent from *En. proavitum*. They are also represented in the composite genome of a closely related albeit phylogenetically distinct lineage of "*Ca*. Endomicrobium trichonymphae" (CET450) recovered from the *Trichonympha* flagellates in *Zootermopsis nevadensis* (18). Phylogenetic analysis places the genes into a clade of putative sugar phosphate transporters in the organophosphate:inorganic phosphate antiporter (OPA) family of major facilitator superfamily (MFS) transporters, which comprises homologs from the distantly related *El. minutum* and various representatives of other bacterial phyla (see Fig. S1 in the supplemental material). They include the hexose phosphate transporter (Hpt) of *Chlamydia* and the G6P sensor protein (UhpC) of *Escherichia coli* and other *Proteobacteria*, whose G6P permease activities have been experimentally documented (19).

Another gene set unique to the endosymbionts encodes the pathway for the uptake and metabolism of glucuronate (20). It comprises a putative hexuronate transporter (ExuT) and all enzymes for the conversion of hexuronic acids to pyruvate and glyceraldehyde 3-phosphate via 2-keto-3-deoxy-phosphogluconate (KDPG) aldolase (21). The genes are present in the genome of strain Rs-D17 and the composite genome CET450 but are absent from both *En. proavitum* and *El. minutum*; they show highest sequence

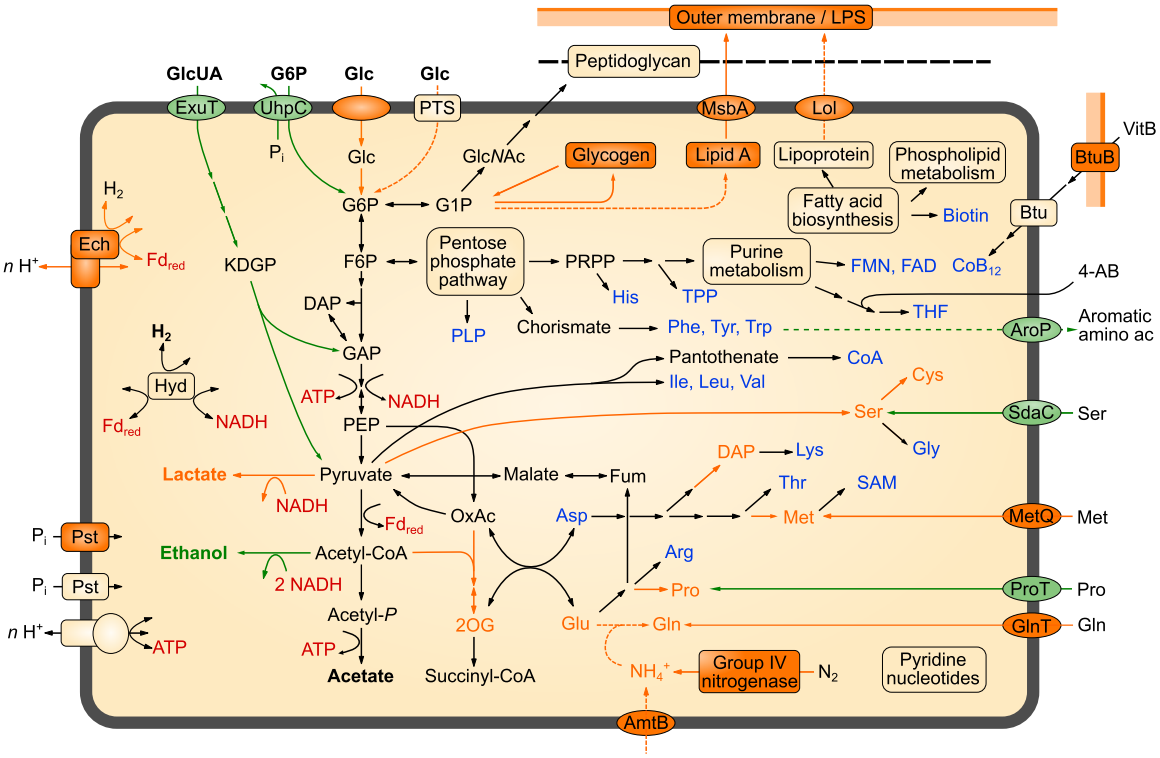

FIG 1 Metabolic pathways in free-living *En. proavitum* (this study) and endosymbiotic strain Rs-D17 (12). Pathways absent or pseudogenized in the endosymbiont and their corresponding products are shown in orange; those that occur only in the endosymbiont are shown in green. Amino acids and cofactors synthesized by both strains are shown in blue; important cosubstrates in energy metabolism are shown in red. Abbreviations (gene names and IUPAC-endorsed abbreviations are not included): 4-AB, 4-amino benzoate; CoB_{12}, coenzyme B_{12}; DAP, dihydroxyacetone phosphate; Fum, fumarate; GAP, glyceraldehyde 3-phosphate; F6P, fructose 6-phosphate; G1P, glucose 1-phosphate; GlcNAc, *N*-acetylglucosamine; GlcUA, glucuronic acid; KDPG, 2-keto-3-deoxy-6-phosphogluconate; 2OG, 2-oxoglutarate; OxAc, oxaloacetate; PEP, phosphoenolpyruvate; PLP, pyridoxal 5-phosphate; PRPP, phosphoribosyl pyrophosphate; PTS, phosphotransferase system; SAM, *S*-adenosylmethionine; THF, tetrahydrofolate; TPP, thiamine pyrophosphate, $VitB_{12}$, vitamin B_{12}.

similarities to homologs in diverse *Firmicutes* and *Bacteroidetes* (see Data Set S2 in the supplemental material).

The genome of *En. proavitum* contains all genes required to oxidize glucose to pyruvate (Embden-Meyerhof pathway) and its subsequent oxidation by a pyruvate:ferredoxin/flavodoxin oxidoreductase (PFOR; with fused subunits), which is closely related to its homolog in *El. minutum* (see Fig. S2 in the supplemental material). Acetyl coenzyme A (acetyl-CoA) is converted to acetate by phosphotransacetylase and acetate kinase. The NADH and reduced ferredoxin formed in the oxidative part of the pathway are apparently regenerated by the reduction of pyruvate to lactate and the formation of H_2.

The lactate dehydrogenase of *En. proavitum* has no homolog in the genome of strain Rs-D17. Conversely, homologs of the bifunctional aldehyde dehydrogenase/ethanol dehydrogenase (AdhE) present in the endosymbionts (both in Rs-D17 and in CET450) are absent from *En. proavitum*. Phylogenetic analysis of the latter revealed a closer relationship to homologs in *Clostridium* and *Fusobacterium* species than to the *adhE* gene of *El. minutum* (Fig. 2), which suggests that the endosymbionts acquired this gene by horizontal gene transfer. Also, the PFOR of the endosymbionts has a different phylogenetic origin than the genes encoding the fused PFOR complex of *En. proavitum* and *El. minutum*; both strains Rs-D17 and CET450 possess the classical four-subunit enzyme complex typical of many strict anaerobes (see Fig. S2 in the supplemental material).

The genome of *En. proavitum* encodes three hydrogenase complexes: two soluble [FeFe]-hydrogenases and one membrane-bound [NiFe]-hydrogenase. Phylogenetic analysis of the large subunit of the [FeFe]-hydrogenases (HydA) revealed that both are

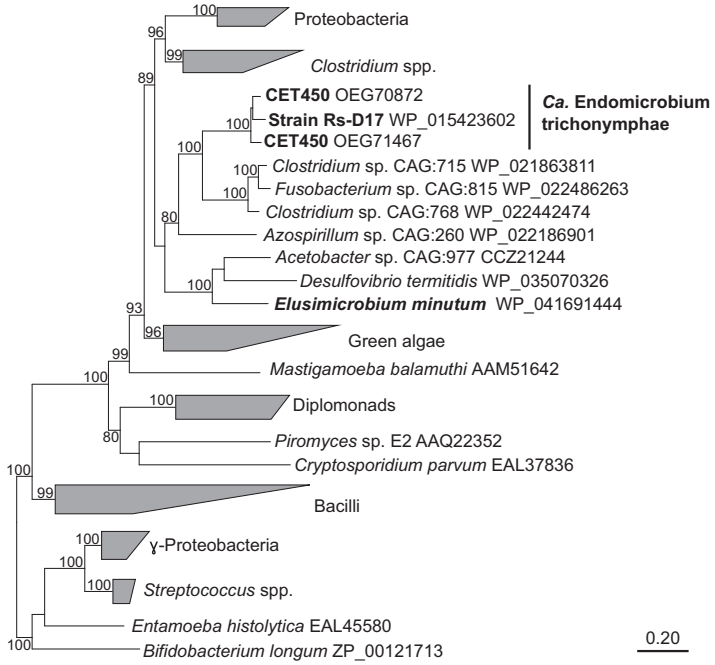

FIG 2 Phylogenetic analysis of the alcohol dehydrogenase AdhE. The maximum-likelihood tree is based on 796 unambiguously aligned amino acid positions and was inferred under the WAG+G+F model. The tree is arbitrarily rooted. Orthologs from the *Elusimicrobia* phylum are marked in boldface.

electron-bifurcating enzymes (group A3) (22). They belong to different, deep-branching clades that comprise orthologs from various bacterial phyla and are only distantly related to the ortholog in *El. minutum* (Fig. 3A). One of the orthologs of *En. proavitum* is closely related to that of strain Rs-D17 and encodes a trimeric enzyme (HydABC), whereas the other has no ortholog in strain Rs-D17 and encodes a tetrameric enzyme (HydABCD) and a putative [FeFe]-hydrogenase maturation protein (HydG; Fig. 3B).

The large subunit of the [NiFe]-hydrogenase falls into the radiation of membrane-bound H_2-evolving enzymes (group 4) (22), with orthologs from *Pelobacter propionicus* and *El. minutum* as closest relatives (see Fig. S3A in the supplemental material). Although they are phylogenetically distinct, their genomic architecture indicates that they are composed of the same number of subunits as the energy-converting hydrogenases (Ech-type) in group 4e (see Fig. S3B in the supplemental material), which suggests that they possess the same function. The genome of *En. proavitum* also contains a set of *hyp* genes, which are essential for the maturation of [NiFe]-hydrogenases (23) and are present also in *El. minutum* (16). None of these genes were detected in the genome of strain Rs-D17 or the composite genome of CET450.

Intermediary metabolism and reserve compounds. Like other anaerobes, *En. proavitum* possesses a putative *Re*-citrate synthase, which falls into a radiation of orthologs that comprise the biochemically characterized enzyme of *Syntrophus aciditrophicus* (24) and are incorrectly annotated as isopropylmalate/homocitrate/citramalate synthase in many genomes (see Fig. S4 in the supplemental material). Succinyl-CoA is produced via a 2-oxoglutarate:flavodoxin oxidoreductase; as in many anaerobic bacteria, a succinate dehydrogenase is absent. Oxaloacetate is produced by PEP carboxykinase. Malate dehydrogenase is absent; fumarate produced during amino acid biosynthesis and purine metabolism is metabolized via malic enzyme (Fig. 1).

The genome of *En. proavitum* encodes the enzymes for glycogen synthesis and degradation, including phosphoglucomutase (*pgcA*), NDP-sugar pyrophosphorylase (*mpg*), glycogen synthase (*glgA*), two α-glucan-branching enzymes (*glgB*), two α-amylases (including the GH of family 57; see above), and maltodextrin phosphorylase (*glgP*). Many of these genes have orthologs in *El. minutum* but are absent from or, in the case

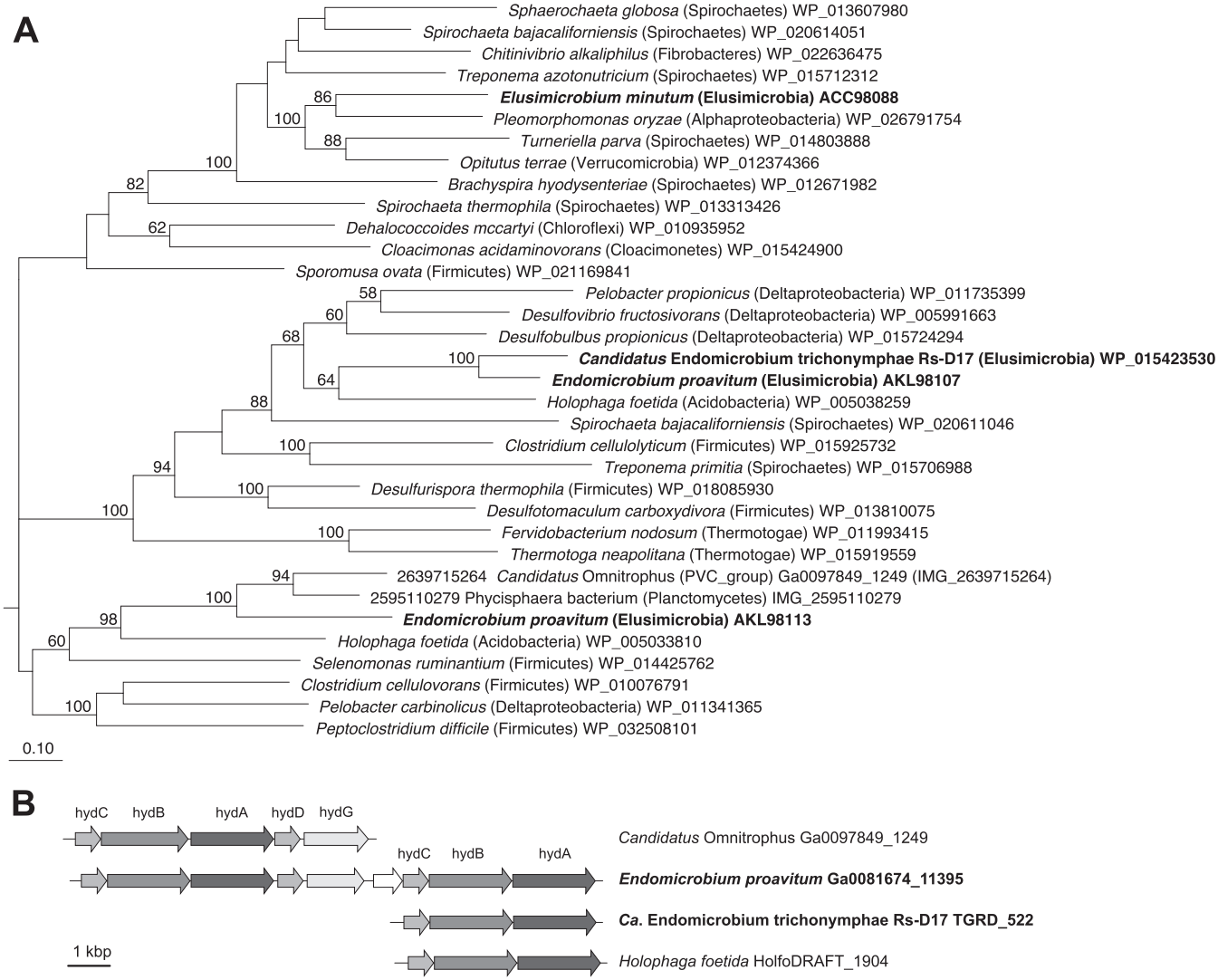

FIG 3 (A) Phylogenetic analysis of the catalytic subunit (HydA) of electron-bifurcating [FeFe]-hydrogenases (group A3) of *Endomicrobia*. The maximum-likelihood tree is based on the comprehensive data set reported by Greening et al. (22) and rooted with representatives of groups A1 and A2. Orthologs from the *Elusimicrobia* phylum are marked in boldface. (B) Organization of the gene sets encoding the [FeFe]-hydrogenases in *En. proavitum* and strain Rs-D17 and in bacteria with the most closely related orthologs identified in the phylogenetic analysis.

of the gene encoding phosphoglucomutase, pseudogenized in the genome of strain Rs-D17.

Nitrogen fixation and assimilation. The genome of *En. proavitum* contains a set of *nif* genes (*nifHDKB*) encoding an unusual group IV nitrogenase. A detailed analysis of the phylogeny of the structural genes and the functionality of their products have been presented already in a previous study (13). The presence of genes encoding an ammonia transporter (*amtB*), two glutamine synthetases (*glnA*; class I and class III), and a PII-like signal-transduction protein (*glnB*) that regulates *nif* gene transcription in diazotrophs (25) is in agreement with both the ability of *En. proavitum* to use ammonia as sole nitrogen source and the repression of *nifH* transcription observed in the presence of ammonia (13).

Neither strain Rs-D17, the metagenome CET450, nor *El. minutum* contain any *nif* genes. Orthologs of *amtB* and class I *glnA* are pseudogenized in strain Rs-D17 (12), and class III *glnA* is entirely absent. Although the lack of genes required for the synthesis of 2-oxoglutarate underscores a dependence of the endosymbiont on exogenous glutamate and/or glutamine, the glutamine transporter (GlnT) of *En. proavitum* is absent from strain Rs-D17 (see Data Set S1 in the supplemental material).

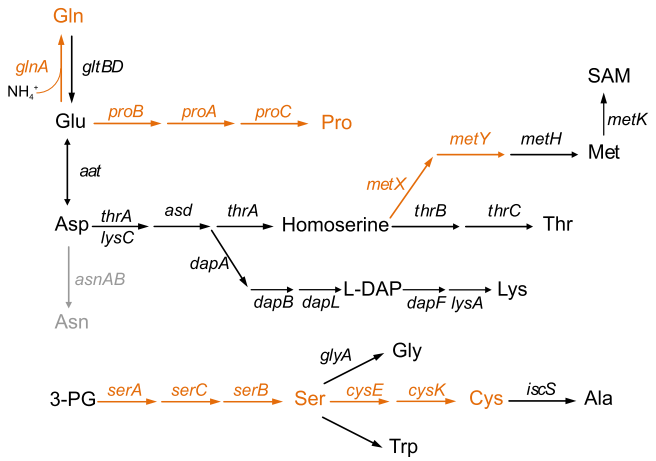

FIG 4 Pathways for amino acid biosynthesis that are present in *En. proavitum* but incomplete in strain Rs-D17. Color code: black, genes present in both genomes; orange, genes absent in strain Rs-D17; gray, genes absent in both strains.

Amino acid and cofactor biosynthesis. The genome of *En. proavitum* encodes the biosynthetic pathways for 19 proteinogenic amino acids, including the key enzymes for glutamine (*glnA*) and methionine production (*metXY*), and for the synthesis of proline, serine, and cysteine; they are either absent or pseudogenized in strain Rs-D17 (Fig. 4). The apparent absence of the genes for asparagine synthetase (*asnAB*) from the genome of *En. proavitum* is puzzling because of its robust growth on mineral medium without organic growth factors other than vitamins (13). A duplication of genes involved in the biosynthesis of aromatic amino acids and coenzyme A (*aroH*, *hisC*, *pheA*, and *coaD*), as in the genome of strain Rs-D17 (12), was not observed.

Although *En. proavitum* does not require amino acids for growth, its genome encodes putative transporters for glutamine and methionine. It is surprising that strain Rs-D17 lost these transporters despite its apparent auxotrophy for glutamine and methionine (see above). Instead, strain Rs-D17 and also the related strain CET450 possess putative transporters for proline (ProT), serine (SdaC), and aromatic amino acids (AroP), which are absent from both *En. proavitum* and *El. minutum*. Although AroP and SdaC cluster with orthologs from *Proteobacteria*, ProT is phylogenetically placed among *Firmicutes* (see Fig. S5 in the supplemental material). Without further orthologs from other members of the *Elusimicrobia* phylum, it is difficult to decide whether the genes of the endosymbionts represent ancestral adaptations to the nutrient-rich gut environment or were acquired more recently, possibly compensating for defects in the corresponding biosynthetic pathways (Fig. 4).

There are no obvious differences between the strains with respect to the ability to synthesize essential cofactors (Fig. 1). Like the endosymbiotic strain Rs-D17 (12), *En. proavitum* possesses almost complete gene sets for the biosynthesis of biotin, coenzyme A, flavin and pyridine nucleotides, lipoate, and *S*-adenosylmethionine. The absence of several genes involved in the biosynthesis of pantothenate (*panE*) and tetrahydrofolate (*folBQ*) from both genomes suggests that the pathways in *Endomicrobia* may slightly differ from those in other bacteria. Pyridoxal phosphate (PLP) seems to be synthesized via the DXP-dependent pathway used by *Enterobacteriaceae*, except that the *pdxH* gene encoding the terminal oxidase reaction is missing. The genes for PLP synthase (*pdxS* and *pdxT*), the key enzyme of an oxygen-independent pathway used by many other bacteria, including strict anaerobes (26), were not found, which suggests that PLP biosynthesis in *Endomicrobia* differs from that in other anaerobes. Both organisms lack the genes required for the entire corrinoid biosynthesis pathway and the conversion of chorismate to 4-aminobenzoate (*pabABC*) in folate biosynthesis, which indicates an auxotrophy for vitamin B_{12} and 4-aminobenzoate; both growth factors are present in the vitamin solution required by *En. proavitum* (13).

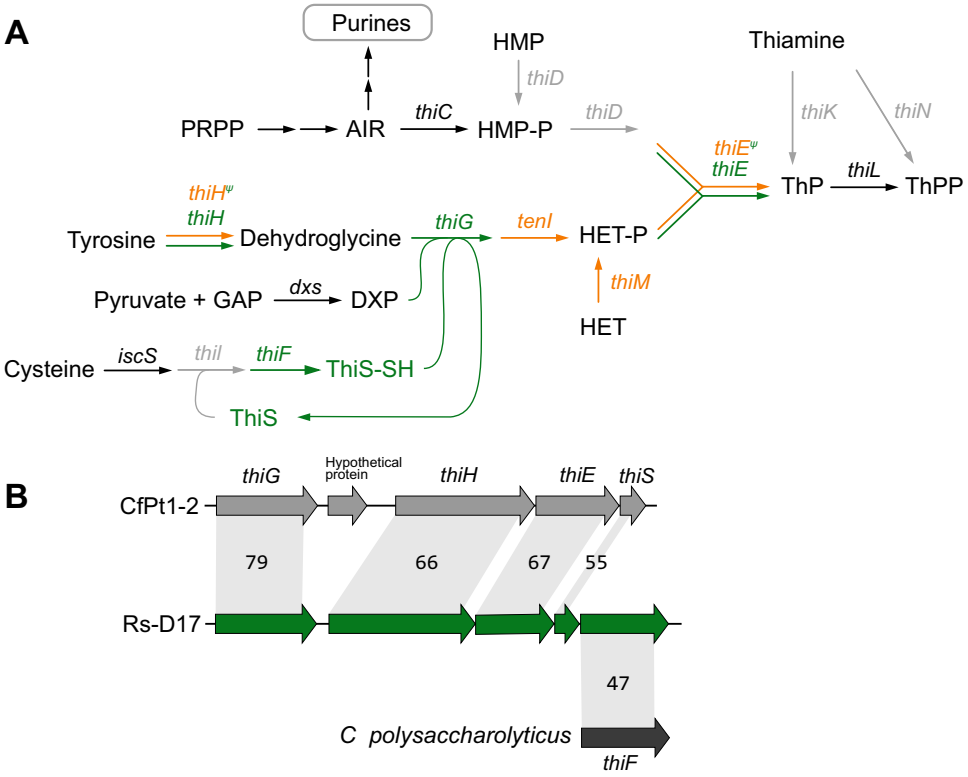

FIG 5 (A) Pathway for the thiamine diphosphate biosynthesis in *En. proavitum* and strain Rs-D17. Color code: black, orthologs present in *En. proavitum* and strain Rs-D17; orange, genes present only in *En. proavitum*; ψ, pseudogenized in strain Rs-D17; green, genes present only in strain Rs-D17; gray, genes absent from both genomes. AIR, aminoimidazole ribotide; HMP, hydroxymethyl pyrimidine; ThP, thiamine phosphate; HET, 4-methyl-5-β-hydroxyethylthiazole; DXP; 1-deoxy-D-xylulose 5-phosphate; ThiS, sulfur carrier protein. (B) *thiGHESF* gene cluster of strain Rs-D17 and orthologs with highest sequence similarity in other bacteria (amino acid sequence similarities [%]). For a detailed phylogenetic analysis of ThiH and ThiE, see Fig. S7 in the supplemental material.

The biosynthesis of thiamine pyrophosphate (TPP) follows the pathway present in anaerobic bacteria (27) but deserves special attention. Although the pathway is not fully resolved, the complete absence of genes involved in thiamine transport and phosphorylation from both *En. proavitum* and strain Rs-D17 strongly suggests that it is operational in both organisms. Several orthologs encoding essential reactions of the biosynthetic pathway (*thiE* and *thiH*) are pseudogenized in strain Rs-D17 (Fig. 5A), but their loss is apparently compensated by a different set of genes (*thiGHESF*) that occur exclusively in the endosymbionts and were most likely acquired by horizontal gene transfer. With the exception of *thiF*, the genes show highest sequence similarity to homologs in "*Candidatus* Azobacteroides pseudotrichonymphae" strain CfPt1-2 (Fig. 5B), an endosymbiont colonizing gut flagellates of other termites. It belongs to a clade of uncultured *Bacteroidetes* (cluster V) that is commonly encountered in termite guts (28), but homologs of the *thi* genes are absent from "*Candidatus* Symbiothrix dinenymphae," the second representative of this clade with a sequenced genome (29). The close relationship of the new gene set to the orthologs in strain CfPt1-2 was confirmed by phylogenetic analysis of the *thiE* and *thiH* gene, whose origin clearly differs from the orthologs in *En. proavitum* (see Fig. S6 in the supplemental material). However, the lack of genomes from other relatives makes it difficult to interpret the direction of the gene flow.

Inorganic nutrients. *En. proavitum* apparently possesses two ABC transport systems for phosphate (Pst), but only one of them has homologs in strain Rs-D17 (Fig. 1). The conservation of a phosphate transporter in a bacterium that uses G6P as energy substrate is not unexpected because UhpC acts as an antiporter that exports one phosphate for each G6P imported into the cell (19).

The presence of *cysE* and *cysK* in the genome of *En. proavitum* explains the ability of the free-living strain to grow in sulfide-reduced mineral medium; their absence from strain Rs-D17 indicates that the endosymbiont requires cysteine not only for protein biosynthesis but also as sulfur source. The genome of *En. proavitum* encodes a putative sulfate permease (SulP) but lacks the pathway for assimilatory sulfate reduction to sulfide.

Both *En. proavitum* and strain Rs-D17 encode transporters for Zn^{2+} and Co^{2+}/Ni^{2+}, but other transporters identified as pseudogenes in the endosymbiont genome seem to be intact in its free-living relative (see Data Set S1 in the supplemental material).

Cell wall and outer membrane biogenesis. The apparent presence of the complete pathways for the biosynthesis of peptidoglycan and lipid A in the genome of *En. proavitum* is in agreement with the ultrastructural analysis of its cell envelope (13), where the trilaminar structure of the outer membrane indicates the presence of a lipopolysaccharide layer (30). While the genes encoding peptidoglycan synthesis have orthologs (and are perfectly syntenic) in strain Rs-D17, the genes required for lipid A synthesis are absent or pseudogenized in the endosymbiont. This includes the genes encoding the lipopolysaccharide translocator (MsbA) and the Tol-Pal proteins, which link the outer membrane with the peptidoglycan and are required for outer membrane integrity in *Escherichia coli* (31) and indicates that the endosymbiont may have retained the capacity to produce a murein sacculus but not an outer membrane (12). Consequently, strain Rs-D17 shares the ABC transporter for $VitB_{12}$ (BtuCDF) with *En. proavitum* but lacks the outer membrane cobalamin receptor protein (BtuB) present in its free-living relative. The general absence from the endosymbiont genome of orthologs encoding characteristic outer membrane proteins in *En. proavitum* (e.g., *ostA* and *lolA*) further corroborates the assumption that the second, outermost membrane observed in electron micrographs of "*Ca*. Endomicrobium trichonymphae" represents a host membrane (6).

DISCUSSION

Comparative genome analysis of the free-living *En. proavitum* and its close but intracellular relative, "*Ca*. Endomicrobium trichonymphae" strain Rs-D17, provided new insights into the evolutionary processes that occurred during symbiogenesis in *Endomicrobia*. The genome of the endosymbiont, which has experienced massive rearrangements during more than 40 million years of coevolution with its flagellate host (17), is not simply a subset of the genes present in the ancestral, free-living lineage. Rather, it possesses a number of functions that are absent from *En. proavitum* and either compensate for genes losses that occurred during genome erosion or represent adaptations to the intracellular habitat. Some of them may represent ancestral traits that facilitated the transition to the endosymbiotic lifestyle but were subsequently lost in the free-living lineage. Others were apparently acquired by horizontal gene transfer from other gut bacteria at some point in evolutionary time. Due to the low sequence similarity to their orthologs, it is not clear whether these events occurred before or after the endosymbiotic event. Nevertheless, the presence of defense systems against foreign DNA in "*Ca*. Endomicrobium trichonymphae" (32) suggests that endosymbionts of termite gut flagellates, unlike the obligately intracellular bacteria colonizing insect tissues (2), are not entirely cut off from gene flow in their intracellular habitat.

Change of energy substrates. The engulfment of an endosymbiont by its host cell not only restricts gene flow but also interferes with the uptake of substrates and nutrients from the environment. Particularly the limited supply of free glucose in the cytosol of most eukaryotic cells forces intracellular bacteria to utilize alternative carbon and energy sources (33). While the free-living *En. proavitum* and *El. minutum* possess two mechanisms to take up and phosphorylate glucose (a putative glucose transporter and a PTS), strain Rs-D17 lacks both the glucose transporter and glucokinase, and the PTS has lost its regulatory function, which suggests that the ability to utilize glucose no longer provides a selective advantage. Instead, the endosymbiont possesses two alternative pathways to supply the cells with a carbon and energy source: a putative

G6P transporter (UhpC), which allows the direct uptake of G6P from the cytoplasm of the host, and a complete pathway for the utilization of glucuronic acid.

Orthologs of UhpC, which are frequently encountered among intracellular pathogens, link the metabolism of the bacterium and its eukaryotic host (34) and are considered an essential factor also in the early stages of plastid evolution (35). It has been argued that the switch from glucose to G6P between the extracellular and intracellular stage of *Listeria monocytogenes* (36) is necessitated by the low affinity of its PTS permeases for glucose (37). Since the *uhpC* genes of the endosymbiotic strains Rs-D17 and CET450 are most closely related to their ortholog of *El. minutum* (Fig. 2), they are most likely an ancestral trait among members of the *Elusimicrobia* phylum that was lost only from *En. proavitum*, possibly representing an important predisposition to an intracellular lifestyle that facilitated the multiple, independent colonization of different flagellate hosts by free-living forms (7, 9).

An interesting case of evolutionary convergence has been observed in "*Candidatus* Ancillula trichonymphae," an obligately intracellular actinobacterium that not only colonizes the same type of microhabitat as "*Ca*. Endomicrobium trichonymphae" but may also occupy a similar ecological niche (38, 39). Also, this endosymbiont has experienced a substantial genome reduction and lacks dedicated transport systems for glucose and hexokinase (39). However, it retained a transporter for glycerol 3-phosphate (GlpT), which belongs to the same family of MFS transporters as UhpC (Fig. 2), which suggests that the switch to sugar phosphates—although not present in all flagellate endosymbionts (28, 40, 41)—provides an advantage in the intracellular habitat. Like "*Ca*. Endomicrobium trichonymphae," also the genome of "*Ca*. Ancillula trichonymphae" encodes the complete pathway for hexuronate utilization. The pathway is present also in "*Candidatus* Azobacteroides pseudotrichonymphae" (28), "*Candidatus* Treponema intracellularis" (41), and "*Candidatus* Adiutrix intracellularis" (40), which colonize other termite gut flagellates and still possess transport systems for glucose, which suggests that hexuronates liberated from the glucuronoxylans in the hemicelluloses of wood (42) are important substrates for these intracellular symbionts.

Differences in energy metabolism. The catabolic pathway of *En. proavitum* reconstructed on the basis of the genome annotation is in agreement with the stoichiometry of glucose fermentation of pure cultures (Eq. 1) (13):

$$\begin{aligned} &\text{glucose} + \text{H}_2\text{O} \rightarrow \text{lactate}^- + \text{acetate}^- + 2\,\text{H}^+ + \text{CO}_2 + 2\text{H}_2 \\ &\Delta G^{\circ\prime} = -207\ \text{kJ/mol} \end{aligned} \quad (1)$$

One of the two NADH formed in glycolysis is regenerated during the reduction of pyruvate to lactate, whereas the other is most likely reoxidized, together with the reduced ferredoxin generated during oxidation of pyruvate to acetate, by the electron-bifurcating [FeFe]-hydrogenase (HydABC). These heterotrimeric enzymes are present in many anaerobic bacteria (22, 43) and have been shown to produce two H_2 from NADH and reduced ferredoxin in *Thermotoga maritima* (44) and *Moorella thermoacetica* (45). The 3 ATP per mol glucose that are formed by substrate-level phosphorylation (SLP) are in agreement with the energetics of the reaction under standard conditions (46). The [NiFe]-hydrogenase of *En. proavitum* is most likely not involved in H_2 production but may have an anabolic function, e.g., the provision of reduced ferredoxin for nitrogen fixation.

The situation is different in strain Rs-D17, which lacks the lactate dehydrogenase (*ldh*) gene present in *En. proavitum*. In principle, the electron-bifurcating [FeFe]-hydrogenase (HydABC) would allow regeneration of the two NADH produced during glycolysis, together with the reduced ferredoxin produced by pyruvate oxidation to acetyl-CoA (Eq. 2):

$$\begin{aligned} &\text{glucose} + 2\,\text{H}_2\text{O} \rightarrow 2\,\text{acetate}^- + 2\,\text{H}^+ + 2\,\text{CO}_2 + 4\,\text{H}_2 \\ &\Delta G^{\circ\prime} = -216\ \text{kJ/mol} \end{aligned} \quad (2)$$

However, it is important to consider that under standard conditions, the free energy of this reaction is insufficient to drive the synthesis of 4 ATP, which are demanded by

the stoichiometry of the pathway and would require a free energy of −280 kJ (~70 kJ per mol ATP) (46). With G6P as the substrate, the reaction becomes slightly more exergonic ($\Delta G^{\circ\prime} = -230$ kJ/mol), but the additional ATP saved in glucose activation exacerbates the thermodynamic dilemma. Unless the number of steps involving SLP is reduced, as in the modified glycolytic pathways of hyperthermophiles (47), the reaction of equation 2 is thermodynamically favorable only at hydrogen partial pressure of <100 Pa (46). However, the hydrogen partial pressures reported for the hindgut paunches of lower termites that harbor *Trichonympha* flagellates with endomicrobial symbionts range from 30 kPa in *Reticulitermes santonensis* to 72 kPa in *Zootermopsis* spp. (48). Considering the proximity of intracellular *Endomicrobia* to the hydrogenosomes of their flagellate hosts (7), it is highly unlikely that hydrogen concentrations in their local microenvironment is substantially lower than that of the hindgut proper.

In that context, the presence of AdhE in strain Rs-D17 may be an important adaptation, because it allows combination of the metabolism of Eq. 2 with an ethanolic fermentation, which yields only 2 ATP per glucose. This reduces the number of ATP conserved and the number of hydrogen produced per glucose to a thermodynamically permissive level and renders the metabolism feasible even under standard conditions (Eq. 3):

$$\begin{aligned} &\text{glucose} + \text{H}_2\text{O} \rightarrow 1\ \text{Acetate}^- + 1\ \text{H}^+ + 1\ \text{ethanol} + 2\ \text{CO}_2 + 2\ \text{H}_2 \\ &\Delta G^{\circ\prime} = -225\ \text{kJ/mol} \end{aligned} \tag{3}$$

This reaction stoichiometry resembles that of *El. minutum* growing on glucose, which presumably regenerates the NADH formed in glycolysis via AdhE and produces H_2 from the ferredoxin reduced during pyruvate oxidation via its [NiFe]-hydrogenase (14); the latter most likely allows conservation of additional energy in form of an electrochemical membrane potential (49), as shown for *Thermoanaerobacter tengcongensis* (50). The [FeFe]-hydrogenase of *El. minutum* is most likely not involved in this metabolism but may be required to maintain the redox balance between NADH and reduced ferredoxin.

However, strain Rs-D17 lacks a [NiFe]-hydrogenase, and ethanol production via AdhE would disturb the stoichiometric balance between NADH and ferredoxin required by its electron-bifurcating [FeFe]-hydrogenase. Also the fermentation of glucuronic acid produces more reduced ferredoxin than NADH (Fig. 1). This calls for an as-yet-undetected mechanism that converts reduced ferredoxin to NADH in strain Rs-D17. An Rnf-like complex, which serves this purpose in other anaerobes (46), seems to be absent from all members of the *Elusimicrobia* phylum.

Membrane transport systems for other metabolites. The detection of mostly complete gene sets for the biosynthesis of amino acids and cofactors is in agreement with the ability of *En. proavitum* to grow on mineral medium lacking organic growth factors other than vitamins. Although a large number of the anabolic pathways are conserved in strain Rs-D17, the loss or pseudogenization of the genes involved in ammonia uptake and assimilation indicates that the endosymbiont became dependent on amino acids as a nitrogen source. In view of the previously recognized breakdown of the pathways for the biosynthesis of several amino acids in the endosymbiont (12), it is likely that the newly acquired amino acid transporters compensate for the loss of the respective biosynthetic functions. However, the supply routes for several essential amino acids are far from clear, because orthologs of the glutamine and methionine transporters encoded in the genome of *En. proavitum* were apparently lost in strain Rs-D17, and the importers for glutamine, methionine, and cysteine required by strain Rs-D17 remain unidentified.

It has been proposed that *Trichonympha* flagellates have domesticated *Endomicrobia* as suppliers of essential amino acids and other cofactors required by the host (12). Unless these products are harvested by digestion of entire symbionts by the flagellate host, this would require the presence of suitable transport systems. In that context, it is of interest that the transporters newly acquired by strain Rs-D17 include a permease for aromatic amino acids (AroP; APC superfamily transporter). Considering that the

corresponding biosynthetic pathways are entirely intact and even experienced gene duplications (12), the presence of AroP would allow delivery of upgraded amino acids back to the host.

Nothing is known about the permeability of the second, putatively host-derived membrane surrounding the symbionts (6, 7), which represents another potential barrier for the uptake of amino acids and inorganic nutrients, providing the host with the possibility to regulate the growth of its symbionts by controlling the supply of these metabolites, as shown for the bacteriocytes of pea aphids (51).

Outlook. There is a growing body of evidence for convergent evolution among unrelated endosymbionts of termite gut flagellates (39–41). *Endomicrobia* provide an excellent model for future studies on the evolutionary mechanisms of symbiogenesis. One reason is the availability of close free-living relatives that are amenable to cultivation, which allows the study of phenotypic traits. Another reason is the presence of multiple lineages of endosymbionts in the radiation of the genus *Endomicrobium*, which colonize different lineages of gut flagellates and were most likely derived independently from an ancestral, free-living line of descent. This provides the opportunity to investigate convergent evolution of closely related symbionts in phylogenetically distant hosts. Conversely, the codiversification of *Endomicrobium* strains with flagellates of the genus *Trichonympha* (10) should result in the parallel evolution of relevant traits in endosymbionts that are isolated within closely related hosts. In view of the presence of several plasmids in *Ca*. Endomicrobium trichonymphae (12, 32) and their apparent absence from the free-living *En. proavitum*, it will also be of great interest study the acquisition of plasmids in other lineages of *Endomicrobia* and a possible role of plasmid-borne genes in symbiogenesis.

MATERIALS AND METHODS

Genome annotation. The genomic DNA of *En. proavitum* was extracted using the cetyltrimethylammonium bromide buffer method and sequenced on a Pacific Biosciences RS platform using three SMRT cells (15). The finished genome was assembled using the Hierarchical Genome Assembly Process (HGAP) version 3, which has been developed to produce nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data with an accuracy >99.999%, despite the high error rate inherent to the raw data produced by this platform (52). The final assembly yielded a circular chromosome with a mean coverage of 76-fold (15). The genome was annotated using the IMG-ER (53) and MicroScope (54) platforms. Annotation results for each coding sequence were compared, and gene predictions were verified (and adjusted when necessary) by inspecting the top BLAST hits among all complete genomes in the entire MicroScope database. Protein-coding genes were assigned to Clusters of Orthologous groups (55) using the COGNiTOR software (56). The annotated genome sequence of *En. proavitum* was deposited with the National Center for Biotechnology Information under accession number CP009498 (15). The pseudogenes of *En. proavitum* were identified using the "Fusion/Fission" function in the MicroScope platform and then checked manually.

Comparative genomics. The complete genome sequences of *El. minutum* (16), strain RsD17 (12), and the composite genome of CET450 (GenBank accession no. LNVX01000001 to LNVX01000866; IMG Project ID Gi01566) were uploaded to the MicroScope platform. The presence and distribution of orthologous genes among the different genomes was explored using the gene phyloprofile tool (57) integrated in the MicroScope platform, with a threshold of ≥50% similarity and ≥80% of matched length, as described previously (17), and by manual searches for specific genes using reciprocal BLAST analysis.

Phylogenetic analyses. Gene sequences were translated into amino acid sequences and aligned using MAFFT (version 7) (58). Unambiguously aligned residues were used in phylogenetic analyses. Maximum-likelihood trees were inferred with MEGA (version 7.0) (59), always using the substitution model that was suggested by MEGA for the respective alignment (the model is indicated in the figure legends of the respective trees). The confidence of tree topologies was estimated by bootstrap analyses (1,000 resamplings).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.00656-17.

SUPPLEMENTAL FILE 1, PDF file, 1.3 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

This study was funded by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 987 and by the Max Planck Society. H.Z. received a

doctoral fellowship of the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-Mic), Marburg, Germany.

REFERENCES

1. Moran NA, Bennett GM. 2014. The tiniest tiny genomes. Annu Rev Microbiol 68:195–215. https://doi.org/10.1146/annurev-micro-091213-112901.
2. McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10:13–26.
3. Brune A. 2014. Symbiotic digestion of lignocellulose in termite guts. Nat Rev Microbiol 12:168–180. https://doi.org/10.1038/nrmicro3182.
4. 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, vol 19. Springer, Heidelberg, Germany.
5. Ohkuma M, Brune A. 2011. Diversity, structure, and evolution of the termite gut microbial community, p 413–438. *In* Bignell DE, Roisin Y, Lo N (ed), Biology of termites: a modern synthesis. Springer, Dordrecht, Netherlands.
6. 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.
7. 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.
8. Brune A. 2012. Endomicrobia: intracellular symbionts of termite gut flagellates. J Endocytobiosis Cell Res 23:11–15.
9. 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.
10. Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and "*Candidatus* Endomicrobium trichonymphae." Mol Ecol 18:332–342. https://doi.org/10.1111/j.1365-294X.2008.04029.x.
11. Zheng H, Dietrich C, Thompson CL, Meuser K, Brune A. 2015. Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). Microbes Environ 30:92–98. https://doi.org/10.1264/jsme2.ME14169.
12. 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.
13. 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.
14. Geissinger O, Herlemann DPR, Morschel 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.
15. Zheng H, Brune A. 2015. Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum *Elusimicrobia*). Genome Announc 3:e00679-15. https://doi.org/10.1128/genomeA.00679-15.
16. 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.
17. 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.
18. Ikeda-Ohtsubo W. 2007. Endomicrobia in termite guts: symbionts within a symbiont. PhD thesis. Philipps-Universität Marburg, Marburg, Germany.
19. 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.
20. Ashwell G, Wahba AJ, Hickman J. 1960. Uronic acid metabolism in bacteria. I. Purification and properties of uronic acid isomerase in *Escherichia coli*. J Biol Chem 235:1559–1565.
21. Suvorova IA, Tutukina MN, Ravcheev DA, Rodionov DA, Ozoline ON, Gelfand MS. 2011. Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria. J Bacteriol 193:3956–3963. https://doi.org/10.1128/JB.00277-11.
22. 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 H_2 is a widely utilised energy source for microbial growth and survival. ISME J 10:761–777. https://doi.org/10.1038/ismej.2015.153.
23. Jacobi A, Rossmann R, Böck A. 1992. The *hyp* operon gene products are required for the maturation of catalytically active hydrogenase isoenzymes in *Escherichia coli*. Arch Microbiol 158:444–451. https://doi.org/10.1007/BF00276307.
24. Kim M, Le H, McInerney MJ, Buckel W. 2013. Identification and characterization of *Re*-citrate synthase in *Syntrophus aciditrophicus*. J Bacteriol 195:1689–1696. https://doi.org/10.1128/JB.02185-12.
25. Dixon R, Kahn D. 2004. Genetic regulation of biological nitrogen fixation. Nat Rev Microbiol 2:621–631. https://doi.org/10.1038/nrmicro954.
26. Mukherjee T, Hanes J, Tews I, Ealick SE, Begley TP. 2011. Pyridoxal phosphate: biosynthesis and catabolism. Biochim Biophys Acta 1814:1585–1596. https://doi.org/10.1016/j.bbapap.2011.06.018.
27. Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. Annu Rev Biochem 78:569–603. https://doi.org/10.1146/annurev.biochem.78.072407.102340.
28. 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 endosymbiont coupling N_2 fixation to cellulolysis within protist cells in termite gut. Science 322:1108–1109. https://doi.org/10.1126/science.1165578.
29. Yuki M, Kuwahara H, Shintani M, Izawa K, Sato T, Starns D, Hongoh Y, Ohkuma M. 2015. Dominant ectosymbiotic bacteria of cellulolytic protists in the termite gut also have the potential to digest lignocellulose. Environ Microbiol 17:4942–4953. https://doi.org/10.1111/1462-2920.12945.
30. Nakamura K, Mizushima S. 1975. In vitro reassembly of the membranous vesicle from *Escherichia coli* outer membrane components. Role of individual components and magnesium ions in reassembly. Biochim Biophys Acta 413:371–393.
31. Clavel T, Germon P, Vianney A, Portalier R, Lazzaroni JC. 1998. TolB protein of *Escherichia coli* K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp, and OmpA Mol Microbiol 29:359–367.
32. 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.
33. Fuchs TM, Eisenreich W, Heesemann J, Goebel W. 2012. Metabolic adaptation of human pathogenic and related nonpathogenic bacteria to extra- and intracellular habitats. FEMS Microbiol Rev 36:435–462. https://doi.org/10.1111/j.1574-6976.2011.00301.x.
34. Elwell C, Mirrashidi K, Engel J. 2016. *Chlamydia* cell biology and pathogenesis. Nat Rev Microbiol 14:385–400. https://doi.org/10.1038/nrmicro.2016.30.
35. Karkar S, Facchinelli F, Price DC, Weber APM, Bhattacharya D. 2015. Metabolic connectivity as a driver of host and endosymbiont integration. Proc Natl Acad Sci U S A 112:10208–10215. https://doi.org/10.1073/pnas.1421375112.
36. Chico-Calero I, Suarez M, Gonzalez-Zorn B, Scortti M, Slaghuis J, Goebel W, Vazquez-Boland JA. 2002. Hpt, a bacterial homolog of the microsomal glucose-6-phosphate translocase, mediates rapid intracellular prolifera-

tion in Listeria. Proc Natl Acad Sci U S A 99:431–436. https://doi.org/10.1073/pnas.012363899.
37. Stoll R, Goebel W. 2010. The major PEP-phosphotransferase systems (PTSs) for glucose, mannose and cellobiose of *Listeria monocytogenes*, and their significance for extra- and intracellular growth. Microbiology 156:1069–1083. https://doi.org/10.1099/mic.0.034934-0.
38. Strassert JFH, Köhler T, Wienemann THG, Ikeda-Ohtsubo W, Faivre N, Franckenberg S, Plarre R, Radek R, Brune A. 2012. "*Candidatus* Ancillula trichonymphae," a novel lineage of endosymbiotic *Actinobacteria* in termite gut flagellates of the genus *Trichonympha*. Environ Microbiol 14:3259–3270. https://doi.org/10.1111/1462-2920.12012.
39. Strassert JF, 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 https://doi.org/10.1111/1758-2229.12451.
40. 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.
41. Ohkuma M, Noda S, Hattori S, Iida T, Yuki M, Starns D, Inoue J, Darby AC, Hongoh Y. 2015. Acetogenesis from H_2 plus CO_2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. Proc Natl Acad Sci U S A 112:10224–10230. https://doi.org/10.1073/pnas.1423979112.
42. 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.
43. Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB, King PW, Adams MW. 2015. [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. Biochim Biophys Acta 1853:1350–1369. https://doi.org/10.1016/j.bbamcr.2014.11.021.
44. Schut GJ, Adams MW. 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. J Bacteriol 191:4451–4457. https://doi.org/10.1128/JB.01582-08.
45. Huang CY, Sabree ZL, Moran NA. 2012. Genome sequence of *Blattabacterium* sp. strain BGIGA, endosymbiont of the *Blaberus giganteus* cockroach. J Bacteriol 194:4450–4451. https://doi.org/10.1128/JB.00789-12.
46. Buckel W, Thauer RK. 2013. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. Biochim Biophys Acta 1827:94–113. https://doi.org/10.1016/j.bbabio.2012.07.002.
47. Kengen SWM, Stams AJM, de Vos WM. 1996. Sugar metabolism of hyperthermophiles. FEMS Microbiol Rev 18:119–137. https://doi.org/10.1111/j.1574-6976.1996.tb00231.x.
48. Pester M, Brune A. 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. ISME J 1:551–565. https://doi.org/10.1038/ismej.2007.62.
49. Thauer RK, Kaster AK, Goenrich M, Schick M, Hiromoto T, Shima S. 2010. Hydrogenases from methanogenic archaea, nickel, a novel cofactor, and H_2 storage. Annu Rev Biochem 79:507–536. https://doi.org/10.1146/annurev.biochem.030508.152103.
50. Soboh B, Linder D, Hedderich R. 2004. A multisubunit membrane-bound [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis*. Microbiology 150:2451–2463. https://doi.org/10.1099/mic.0.27159-0.
51. Whitehead LF, Douglas AE. 1993. A metabolic study of *Buchnera*, the intracellular bacterial symbionts of the pea aphid *Acyrthosiphon pisum*. J Gen Microbiol 139:821–826. https://doi.org/10.1099/00221287-139-4-821.
52. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
53. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics 25:2271–2278. https://doi.org/10.1093/bioinformatics/btp393.
54. Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fevre F, Longin C, Mornico D, Roche D, Rouy Z, Salvignol G, Scarpelli C, Thil Smith AA, Weiman M, Medigue C. 2013. MicroScope: an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res 41:D636–D647. https://doi.org/10.1093/nar/gks1194.
55. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36. https://doi.org/10.1093/nar/28.1.33.
56. Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin EV, Mushegian A. 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics 26:1481–1487. https://doi.org/10.1093/bioinformatics/btq229.
57. Engelen S, Vallenet D, Médigue C, Danchin A. 2012. Distinct co-evolution patterns of genes associated to DNA polymerase III DnaE and PolC. BMC Genomics 13:69. https://doi.org/10.1186/1471-2164-13-69.
58. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
59. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.