



Simultaneous Single-Cell Genome and Transcriptome Sequencing of Termite Hindgut Protists Reveals Metabolic and Evolutionary Traits of Their Endosymbionts

Michael E. Stephens,^{a,b}  Jacquelynn Benjamino,^{a,c}  Joerg Graf,^a  Daniel J. Gage^a

^aDepartment of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA

^bAzenta Life Sciences, South Plainfield, New Jersey, USA

^cCommunication Partners Group, New York, New York, USA

Michael E. Stephens and Jacquelynn Benjamino share first authorship. Both contributed equally to work described in the manuscript. Order was determined by which did the majority of the manuscript writing.

ABSTRACT Some of the protist species which colonize the hindguts of wood-feeding *Reticulitermes* termites are associated with endosymbiotic bacteria belonging to the genus *Endomicrobium*. In this study, we focused on the endosymbionts of three protist species from *Reticulitermes flavipes*, as follows: *Pyronympha vertens*, *Trichonympha agilis*, and *Dinenympha* species II. Since these protist hosts represented members of different taxa which colonize separate niches within the hindguts of their termite hosts, we investigated if these differences translated to differential gene content and expression in their endosymbionts. Following assembly and comparative genome and transcriptome analyses, we discovered that these endosymbionts differed with respect to some possible niche-specific traits, such as carbon metabolism. Our analyses suggest that species-specific genes related to carbon metabolism were acquired by horizontal gene transfer (HGT) and may have come from taxa which are common in the termite hindgut. In addition, our analyses suggested that these endosymbionts contain and express genes related to natural transformation (competence) and recombination. Taken together, the presence of genes acquired by HGT and a putative competence pathway suggest that these endosymbionts are not cut off from gene flow and that competence may be a mechanism by which members of *Endomicrobium* can acquire new traits.

IMPORTANCE The composition and structure of wood, which contains cellulose, hemicellulose, and lignin, prevent most organisms from using this common food source. Termites are a rare exception among animals, and they rely on a complex microbiota housed in their hindguts to use wood as a source of food. The lower termite, *Reticulitermes flavipes*, houses a variety of protists and prokaryotes that are the key players in the disassembly of lignocellulose. Here, we describe the genomes and the gene expression profiles of five *Endomicrobium* endosymbionts living inside three different protist species from *R. flavipes*. Data from these genomes suggest that these *Endomicrobium* species have different mechanisms for using carbon. In addition, they harbor genes that may be used to import DNA from their environment. This process of DNA uptake may contribute to the high levels of horizontal gene transfer noted previously in *Endomicrobium* species.

KEYWORDS carbon metabolism, endosymbionts, genetic competence, metagenomics, protists, single-cell methods, symbiosis, termites

Among the wood-feeding lower termites, symbiotic protists which reside in the hindgut are often colonized by endosymbionts (1–4). In *Reticulitermes* spp. termites, both Oxymonadida (order) and Parabasalia (class) protists associate with

Editor Lifeng Zhu, Nanjing Normal University

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

Address correspondence to Daniel J. Gage, daniel.gage@uconn.edu.

The authors declare no conflict of interest.

Received 12 January 2022

Accepted 13 January 2022

Published 2 February 2022

endosymbiotic bacteria belonging to the genus *Endomicrobium* (phylum *Elusimicrobia*, class *Endomicrobia*) (2, 5–7). Members of *Endomicrobium* have been shown to comprise a significant portion of the core bacterial community in wood-feeding termites, such as *Reticulitermes flavipes* (8, 9). These endosymbiotic lineages are thought to have initiated their associations with hindgut protists approximately 40 to 70 million years ago (10) and arose from free-living relatives during multiple independent acquisition events (11). Vertical passage from one protist cell to its progeny has resulted in cospeciation, as inferred from congruent rRNA phylogenies (7, 10, 12, 13).

In addition to colonizing the cytoplasm of certain hindgut protist species, *Endomicrobium* spp. are ectosymbionts of protists (14) and can be free living as well (11, 15–17). Because of their distribution across these separate niches, they provide an opportunity for studying bacterial genome evolution across different association lifestyles, namely, free living, endosymbiotic, and ectosymbiotic.

To determine the differences between two *Endomicrobium* species that are closely related but with distinct lifestyles, a previous study compared genomes of a free-living *Endomicrobium*, *Endomicrobium proavitum* strain Rsa215 (16), and “*Candidatus Endomicrobium trichonymphae*” strain Rs-D17 (3), an endosymbiont (3, 18). The findings suggested that the transition from the free-living state to an intracellular lifestyle involved genome reduction, similar to that of endosymbionts of sap-feeding insects and many obligate intracellular pathogens. However, the intracellular strain Rs-D17 also incorporated genes, possibly from other termite gut inhabitants, by horizontal gene transfer (HGT) (18). For example, the genome of “*Ca. Endomicrobium trichonymphae*” Rs-D17 appeared to have acquired several pathways, including those that encode sugar and amino acid transporters and genes involved in amino acids biosynthesis (18). These findings suggested that, unlike the endosymbionts of sap-feeding insects, *Endomicrobium* species may not be completely cut off from gene flow (18).

We expand upon these studies by presenting and comparing near-complete draft genome and transcriptome sequences of three *Endomicrobium* organisms, which were assembled from single protist cells of three different species that inhabit the hindgut of *R. flavipes*. One of these protist species, *Pyronympha vertens*, lives attached to the oxic gut wall (19, 20), while the other two, namely, *Trichonympha agilis* and *Dinenympha* species II, are found in the more anoxic hindgut lumen. In addition, *P. vertens* and *Dinenympha* species II are both oxymonads, while *T. agilis* is a parabasalid.

The analyses indicate that these *Endomicrobium* species have differences in their gene content and expression, which are related to carbon usage and metabolism. Also, as seen previously in “*Ca. Endomicrobium trichonymphae*” Rs-D17, they have likely acquired genes from putative donor taxa that are commonly associated with termites. In addition, we describe data suggesting that these *Endomicrobium* species have retained competence genes which may allow them to import exogenous DNA and that perhaps have contributed to HGT. The genes involved in this pathway are conserved across several *Endomicrobium* species and were expressed in the endosymbionts examined in this study.

RESULTS

Phylogeny of protist hosts. Protist 18S rRNA genes were retrieved from metagenome assemblies and confirmed (when possible) independently by PCR and Sanger sequencing. A maximum likelihood (ML) phylogenetic tree was made that indicated that the species of the protist cells used in this study were *Trichonympha agilis* (cells TA21 and TA26), *Pyronympha vertens* (cells PV1 and PV7), and *Dinenympha* species II (cell DS12) (Fig. 1). These protist species have been confirmed previously to live associated with *R. flavipes*, the termite species used in this study.

Taxonomic composition of the assemblies. Reads from each of the single-cell metagenomes were mapped against a database of V4 amplicon sequence variants (ASVs) from previous work and from the DictDB and SILVA rRNA databases (Tab5 in Data Set S1 in the supplemental material). Each assembly, except PV7, contained reads that mapped only to the expected single-cell eukaryotic host. The PV7 metagenome contained no eukaryotic

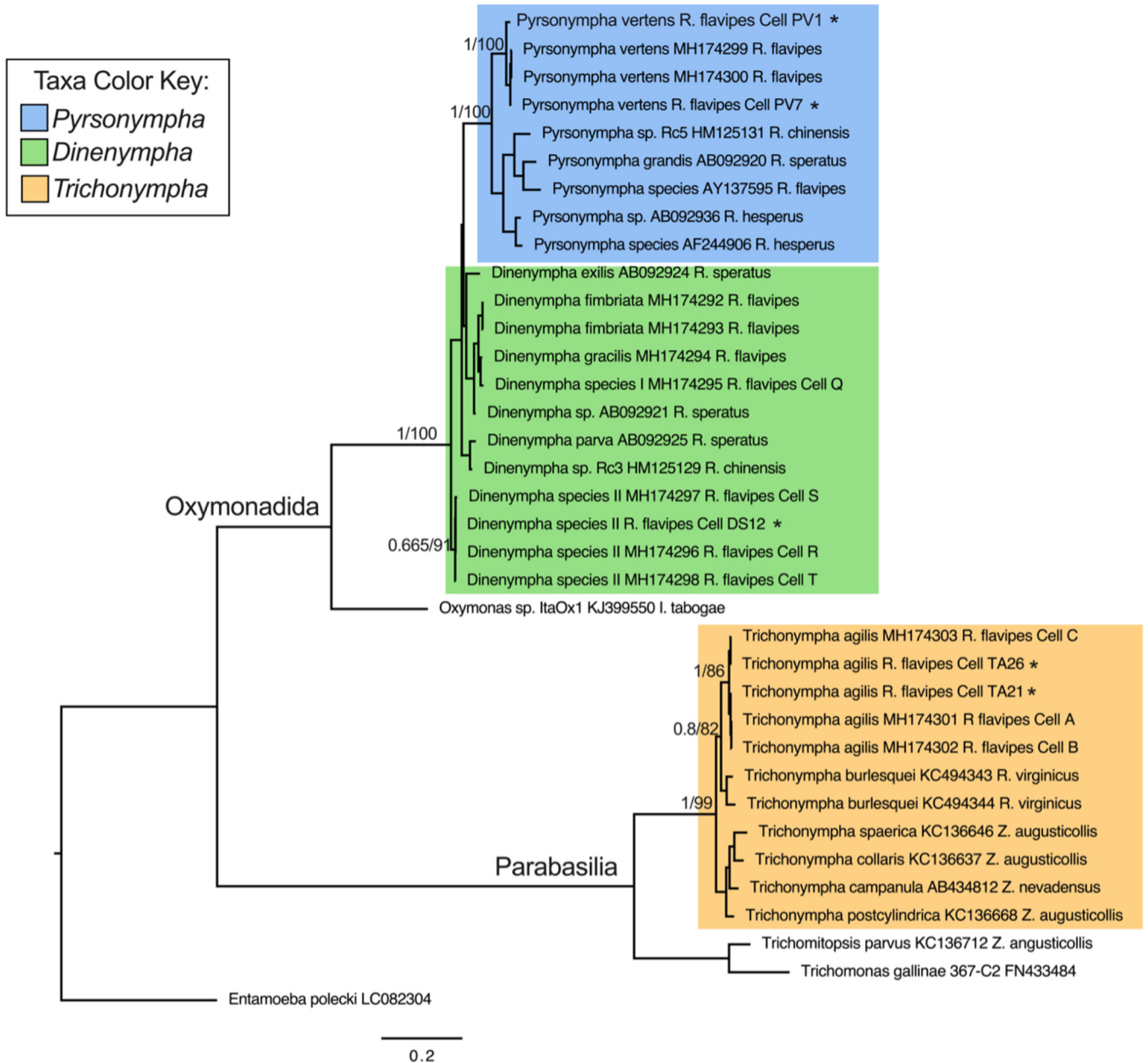


FIG 1 Protist 18S rRNA gene phylogeny. 18S rRNA genes were retrieved from single protist cell metagenome assemblies and aligned to references, and a maximum likelihood (ML) phylogenetic tree was made using IQ-TREE using substitution model TIM2+G4. All 18S rRNA gene sequences obtained in this study (denoted by *) are shown grouped with their respective references. Branch support values represent the Bayesian posterior probability and bootstrap support values, respectively.

ASVs. The taxonomic composition of the PV1 and PV7 assemblies were the least diverse and contained the same major *Endomicrobium* ASV (Elusimicrobia_ASV024). This ASV was found associated with *Pyrronympha* sp. previously (7). The PV7 assembly also contained single reads that matched Elusimicrobia_ASV020 and Elusimicrobia_ASV022. These ASVs are usually associated with *Trichonympha* sp. but may be associated with *Pyrronympha* sp. at low levels or may be contaminants acquired during the single-cell isolation. The DS12 assembly was the most diverse and had reads mapping to a variety of spirochete species. This protist, *Dinonympha* sp. II, is known to associate with a large number of spirochete ectosymbionts (7). DS12 contained Elusimicrobia_ASV023 as the only *Endomicrobium* ASV. This ASV was seen previously to associate with *Dinonympha* sp. II (7). The *Trichonympha* assemblies TA21 and TA21 differed from each other in that reads from each hit different *Trichonympha* ASVs (18S_rRNA::

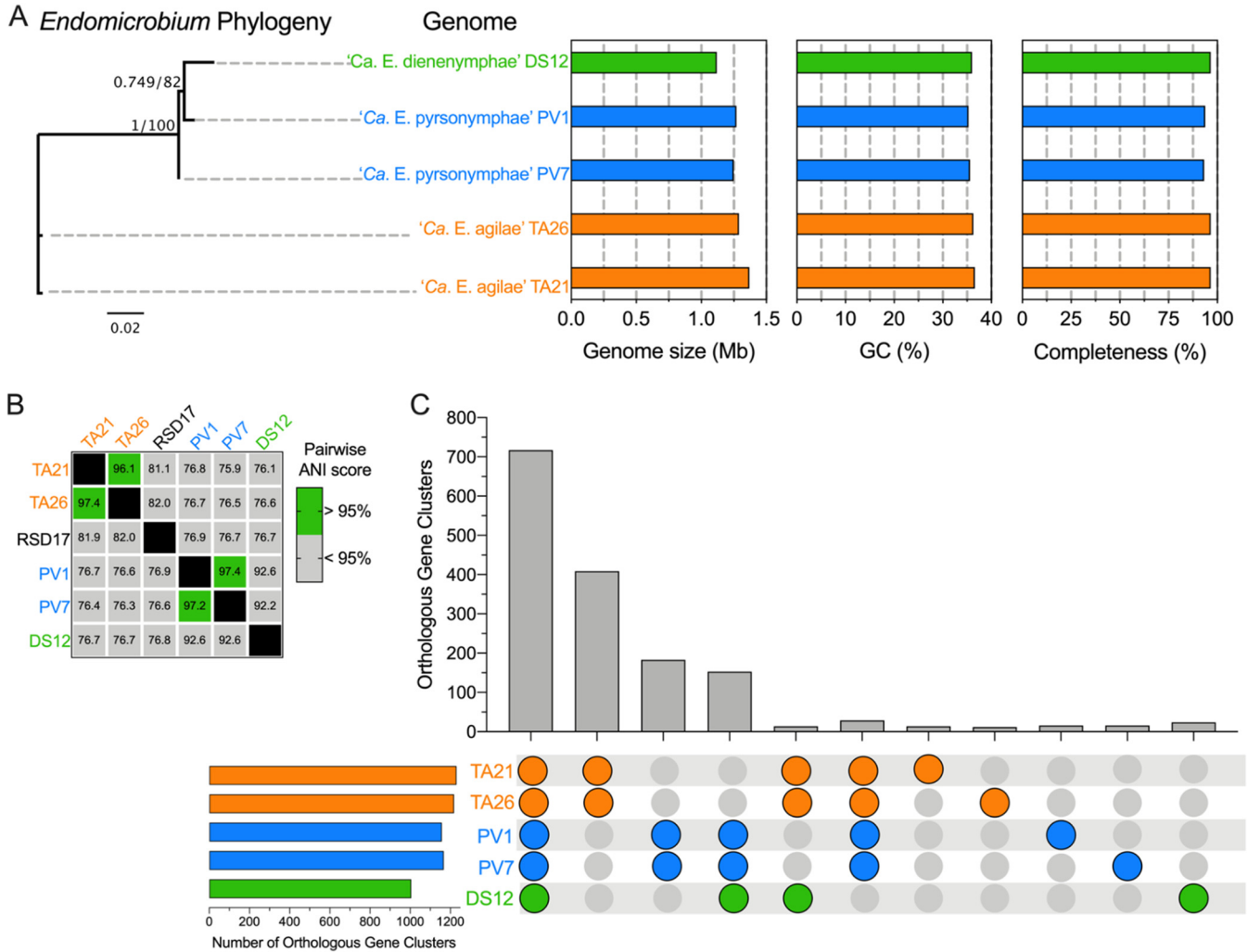


FIG 2 *Endomicrobium* draft genomes statistics, species identification, and shared gene content. (A) 16S rRNA gene maximum likelihood tree (unrooted) of the three *Endomicrobium* species, genome sizes, percent G+C content, and estimated percent genome completeness. (B) Pairwise genomic ANI scores of *Endomicrobium* genomes obtained by this study and a previously sequenced relative, Rs-D17. (C) UpSet graph of the number of orthologous gene clusters (OGCs) of protein-coding sequences within and across each of the *Endomicrobium* draft genomes.

TA21_scaffold229 and KC494354.1_*Trichonympha*, respectively) and different *Endomicrobium* ASVs (Elusimicrobia_ASV020 and Elusimicrobia_ASV019). It has been observed by us, and others, that there are at least two *Trichonympha* taxa in *R. flavipes* that are morphologically similar but that can be differentiated by their rRNA sequences and those of their resident *Endomicrobium* species (7, 21); TA21 and TA26 may be an example of each of these two taxa.

Endomicrobium genome statistics. Five near-complete *Endomicrobium* genomes were obtained from single protist cell metagenomic assemblies. The five draft genomes contained 25 to 229 contigs, ranged from 1.12 to 1.37 Mb long, and had a G+C content from 35.3% to 36.6%. NCBI annotated between 10 and 37 pseudogenes in the genomes. They are listed in Tab12 of Data Set S1. The genomes were 93.3% to 96.6% complete and contained between 0.0% and 5.2% contamination as measured by CheckM (Fig. 2A; also Tab3 in Data Set S1. Gene content, as determined by NCBI, is also provided in this table.

To determine if these genomes were from the same or different *Endomicrobium* species, we calculated pairwise genomic distances using average nucleotide identity (ANI) methods as implemented by JSpeciesWS (22). An ANI score of 95% or greater was used as a marker for species-level cutoff (Fig. 2B) (23). From *T. agilis* samples, we

assembled two *Endomicrobium* draft genome sequence which had ANI scores of greater than 96% to one another but less than 90% to “*Ca. Endomicrobium trichonymphae*” Rs-D17, indicating that they are likely different species. Based on this analysis, we refer to the draft genomes as coming from “*Candidatus Endomicrobium agilae*” TA21 and “*Candidatus Endomicrobium agilae*” TA26. We also assembled two *Endomicrobium* genomes from *P. vertens* samples which had an ANI score greater than 97% identity to each another (Fig. 2B) and whose 16S rRNA genes were greater than 98% identical to a previously described species, “*Candidatus Endomicrobium pyrsonymphae*” (6), which is the *Candidatus* species designation that we use for PV1 and PV7. One additional *Endomicrobium* genome sequence was assembled from *Dinenympha* species II. This genome did not share an ANI score greater than 95% to other *Endomicrobium* genomes and was thus given a new *Candidatus* species designation, namely, “*Candidatus Endomicrobium dinenymphae*” DS12 (Fig. 2B).

Individually, these *Endomicrobium* genomes contained between 1,005 and 1,230 orthologous gene clusters (OGCs), of which 717 were found in all 5 genomes (Fig. 2C). Additionally, 409 OGCs were unique to “*Ca. Endomicrobium agilae*” TA21 and TA26, and another 183 OGCs were unique to “*Ca. Endomicrobium pyrsonymphae*” PV1 and PV7 (Fig. 2C). Although the genome of “*Ca. Endomicrobium dinenymphae*” DS12 had only 24 unique OGCs, it shared 153 with “*Ca. Endomicrobium pyrsonymphae*” PV1 and PV7 (Fig. 2C) which may reflect similar selective pressures for gene retention in their oxymonad hosts *Dinenympha* and *Pyrsonympha* sp., or their more recent shared history, compared with the *Endomicrobium* sp. (TA21 and TA26) that associated with the parabasalid *T. agilis*.

Analyses using HGTector indicated that the 5 *Endomicrobium* genomes contained between 72 (TA26) and 100 (PV1) genes that had likely been acquired by HGT. This result represented roughly 7% of the protein-coding genes of these genomes. A total of 35 core genes appeared to have been acquired by HGT in all 5 organisms (Tab6 in Data Set S1). For comparison, HGTector identified 75 genes acquired by HGT in *Endomicrobium* strain RsD17 representing 7% of its genome.

Biosynthesis of amino acids, vitamins, and peptidoglycan. The presence of genes for the various functions discussed below were either annotated by NCBI or detected by tblastn using the queries listed in Tab9 and Tab10 of Data Set S1. When genes were not detected in this manner, read mapping using Geneious and read mapping using MEGAN (24) were done as well to identify reads from genes that may not have assembled into contigs. In general, each of the five *Endomicrobium* genome sequences assembled in this study had similar gene contents for processes involved in the biosynthesis of amino acids (see Fig. S2A in the supplemental material), vitamins (Fig. S2B), glycolysis and the pentose phosphate pathway (Fig. S2C), and peptidoglycan (Fig. S2D). Each genome possessed complete pathways for alanine (from cysteine) aspartate, arginine, glutamine, glutamate, glycine (from imported serine), isoleucine, leucine, valine, lysine, tyrosine, phenylalanine, and tryptophan biosynthesis (Fig. S2A). Interestingly, the *Endomicrobium* symbionts of oxymonad protists (PV1, PV7 and DS12) lacked at least one gene in the biosynthesis pathway for histidine (*hisG*) (Fig. S2A). The histidine biosynthetic pathway was complete in the genomes of “*Ca. Endomicrobium agilae*” TA21 and TA26 (Fig. S2A). Conversely, it is likely that the *Endomicrobium* symbionts of oxymonad protists (PV1 and PV7) can make proline, while the symbionts represented by genomes TA21, TA26, DS12, and RsD17 cannot (Fig. S2A). The five genomes encoded incomplete pathways for the synthesis of cysteine and methionine. The three genomes isolated from oxymonad protists encoded a methionine transporter (*MetT*) and all contained a gene encoding a B12-dependent methionine synthase system comprised of *MetH* and an activation protein *Meth2* (Fig. S2A). However, it is unclear how methionine can be synthesized or transported in TA21 and TA26. A similar situation exists in the *Endomicrobium* strain RsD17 (18). Also incomplete in all five genomes were pathways for the synthesis of serine and asparagine. Each genome encoded a serine transporter (*SdaC*) and a proline transporter (*ProT*), and PV1, PV7, and DS12 each encoded a glutamate transporter (*GltP*) (Fig. S2A).

The five *Endomicrobium* genomes also had similar gene contents for processes involved in the biosynthesis of vitamins and cofactors, with the pathways to pantothenate, coenzyme A (CoA), NAD, and NADP being complete and other pathways being incomplete (Fig. S2B). Interestingly, the biotin biosynthesis pathways in the five genomes are missing just a single gene (*bioW*) needed to convert pimelate to pimelate-CoA suggesting that pimelate-CoA may be synthesized by another enzyme or imported (Fig. S2B). Several genes in the thiamine biosynthesis pathway were also missing in each of these genomes (Fig. S2B). As noted previously for *E. proavitum* and “*Candidatus Endomicrobium trichonymphae*” strain Rs-D17, the five genomes described here were also missing the steps in the folate pathway needed to make 4-aminobenzoate, which may be transported into the cells (18). The pathways for pyridoxine (B6) and vitamin B12 were also incomplete, although each of the five *Endomicrobium* genomes appeared to encode ABC transport systems for vitamin B12 and heme.

Regarding peptidoglycan synthesis, each *Endomicrobium* genome was missing a gene encoding the enzyme (BacA) which typically dephosphorylates undecaprenyl pyrophosphate (Fig. S2D). Since these different *Endomicrobium* species, including the free-living *E. proavitum*, are missing the same gene, it may be that these bacteria utilize an alternate phosphatase to carry out the same function as BacA.

Differences in carbon metabolism. Some of the more interesting differences between these *Endomicrobium* genomes pertained to carbon metabolism. Each of the five *Endomicrobium* genomes encoded relatively simple pathways for importing and using different wood-derived carbon sources. Each had a complete phosphotransferase system (PTS) for importing sugars. Present were two EIIA genes encoding sugar specific phosphorylation proteins related most closely to those of the mannose and fructose type EIIA proteins (Fig. S2C). Zheng et al. reported that *E. proavitum*, which contains a very similar PTS pathway, did not grow on mannose or fructose but did grow on glucose, suggesting that glucose may be the carbohydrate transported by the PTS in that *Endomicrobium* species and perhaps in the ones described here as well (16, 18).

Based on the gene content in the five genomes analyzed here, the carbon sources capable of being catabolized by endosymbiotic *Endomicrobium* species may often differ from each other and from their free-living relatives. For example, “*Ca. Endomicrobium agilae*” TA21 and TA26 encoded all the genes necessary to import and use both glucuronate and glucose-6-phosphate (Fig. 3A and 3B). The closely related “*Ca. Endomicrobium trichonymphae*” Rs-D17 (3) also contained these genes. Interestingly, genome analyses suggest that these two carbon sources cannot be used by the other *Endomicrobium* species studied here which lack genes encoding the glucuronate transporter ExuT, the glucuronate isomerase UxaA, and the glucose-6-phosphate transporter UhcP. The other *Endomicrobium* genomes encoded either arabinose (“*Ca. Endomicrobium pyrsonymphae*” PV1 and PV7) or xylose (“*Ca. Endomicrobium dinenymphae*” DS12) import and catabolism proteins that were not encoded in the TA21, TA26, *E. proavitum*, or “*Ca. Endomicrobium trichonymphae*” Rs-D17 genomes. (Fig. 4 and 5, respectively). Transcriptome data indicated that each of the genes involved in these carbon usage pathways were expressed in the respective *Endomicrobium* while they resided in their protist hosts (Fig. 3C, 4C, and 5C). Metabolites from these carbon sources are typically fed into both the nonoxidative pentose phosphate pathway and glycolysis, of which both are complete in the five genomes described here (Fig. 3B, 4B, and 5B; Fig. S2C).

Other likely differences in the carbon metabolism of these *Endomicrobium* species included the production of fermentation end products (Fig. S2C). An analysis of the five *Endomicrobium* genomes suggested that following glycolysis, pyruvate can be fermented to acetate; however, only the genomes of “*Ca. Endomicrobium agilae*” and “*Ca. Endomicrobium dinenymphae*” encoded AdhE, which can convert acetate to ethanol. In addition, genes encoding lactate dehydrogenase (Ldh) were in the genomes of both “*Ca. Endomicrobium dinenymphae*” and “*Ca. Endomicrobium pyrsonymphae*” but not in “*Ca. Endomicrobium agilae*” or “*Ca. Endomicrobium trichonymphae*” Rs-D17 (Fig. 3B, 4B, and 5B). Differences in these fermentation pathways between free-living *E.*

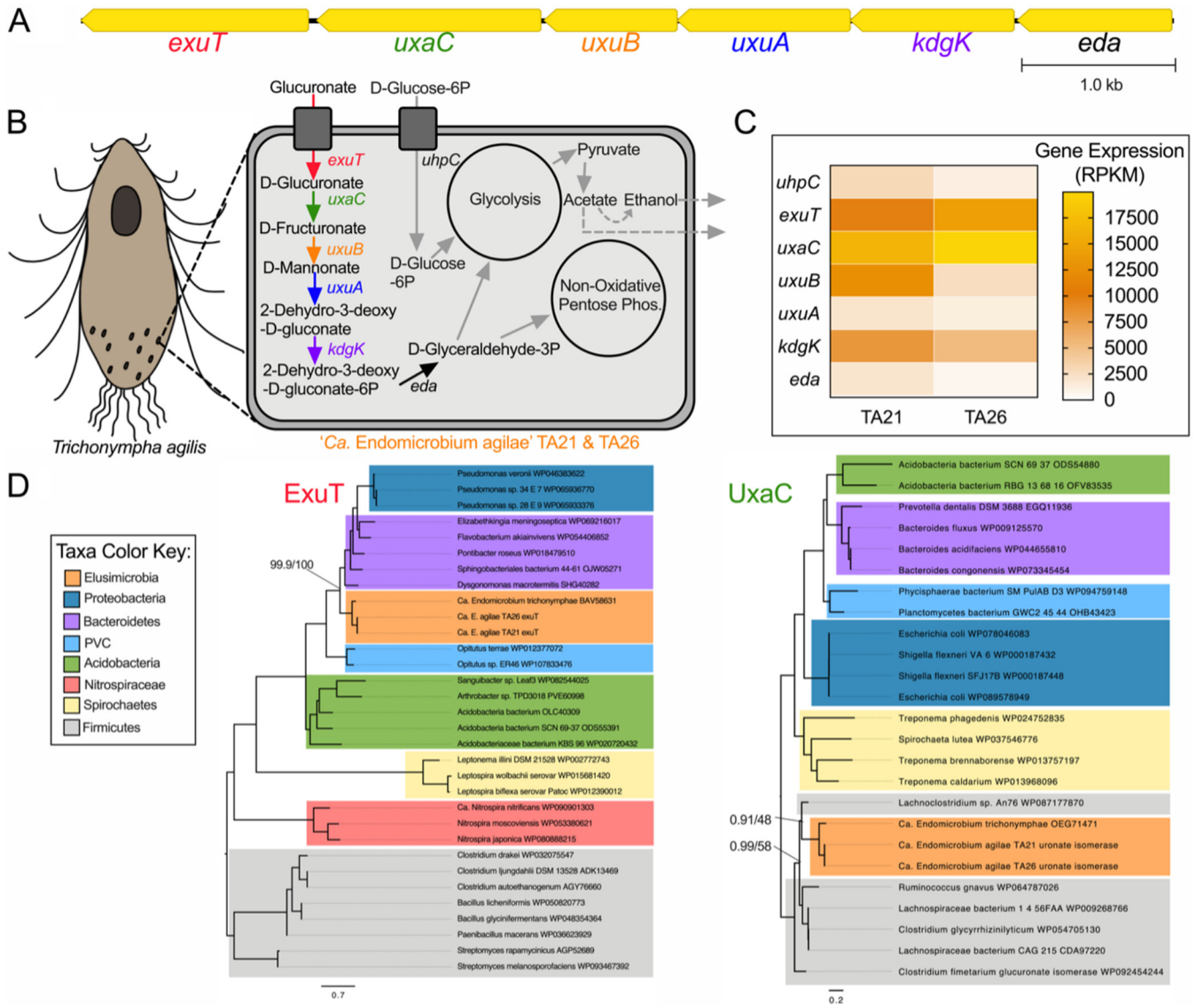


FIG 3 Carbon metabolism and HGT in “*Ca. Endomicrobium agilae.*” (A) Gene neighborhood of the genes involved in the metabolism of glucuronate in the “*Ca. Endomicrobium agilae.*” TA21 and TA26 genomes. (B) Diagram of a protist host and an *Endomicrobium* cell showing the inferred metabolic conversions of carbon sources based on gene content data. (C) Gene expression data of genes of interest (rows) pertaining to carbon metabolism in “*Ca. Endomicrobium agilae.*” TA21 and TA26 (columns). (D) Maximum likelihood phylogenetic trees of amino acid sequences of the transporter (ExuT, using substitution model LG+F+G4) and isomerase (UxaC, using substitution model LG+I+G4) in the glucuronate metabolism pathway. Support values represent the Bayesian posterior probability and bootstrap support values, respectively.

proavutum and “*Ca. Endomicrobium trichonymphae.*” Rs-D17 were described earlier by Zheng et al. (18).

Previous studies identified genes acquired by horizontal gene transfer (HGT) in other *Endomicrobium* species (18); therefore, we tested whether HGT could, at least in part, explain the differences seen in carbon metabolism across the genome sequences presented in this study. Phylogenetic trees were made for each of the transport and isomerase proteins in the glucuronate, arabinose, and xylose degradation pathways (Fig. 3D, 4D, and 5D), and the phylogenies were compared with the *Endomicrobium* 16S rRNA gene phylogeny to determine if they were congruent (see Fig. S4 in the supplemental material). In each case, these phylogenies were not congruent, suggesting that these genes were acquired by HGT (Fig. 3D, 4D, and 5D). Likely donor taxa include *Bacteroidetes*, *Actinobacteria*, and *Firmicutes* (Fig. 3D, 4D, and 5D), which are all part of the hindgut community of *R. flavipes* (8). An additional analysis of these genes by HGTector (25) indicated that they were likely acquired by HGT from the same donor

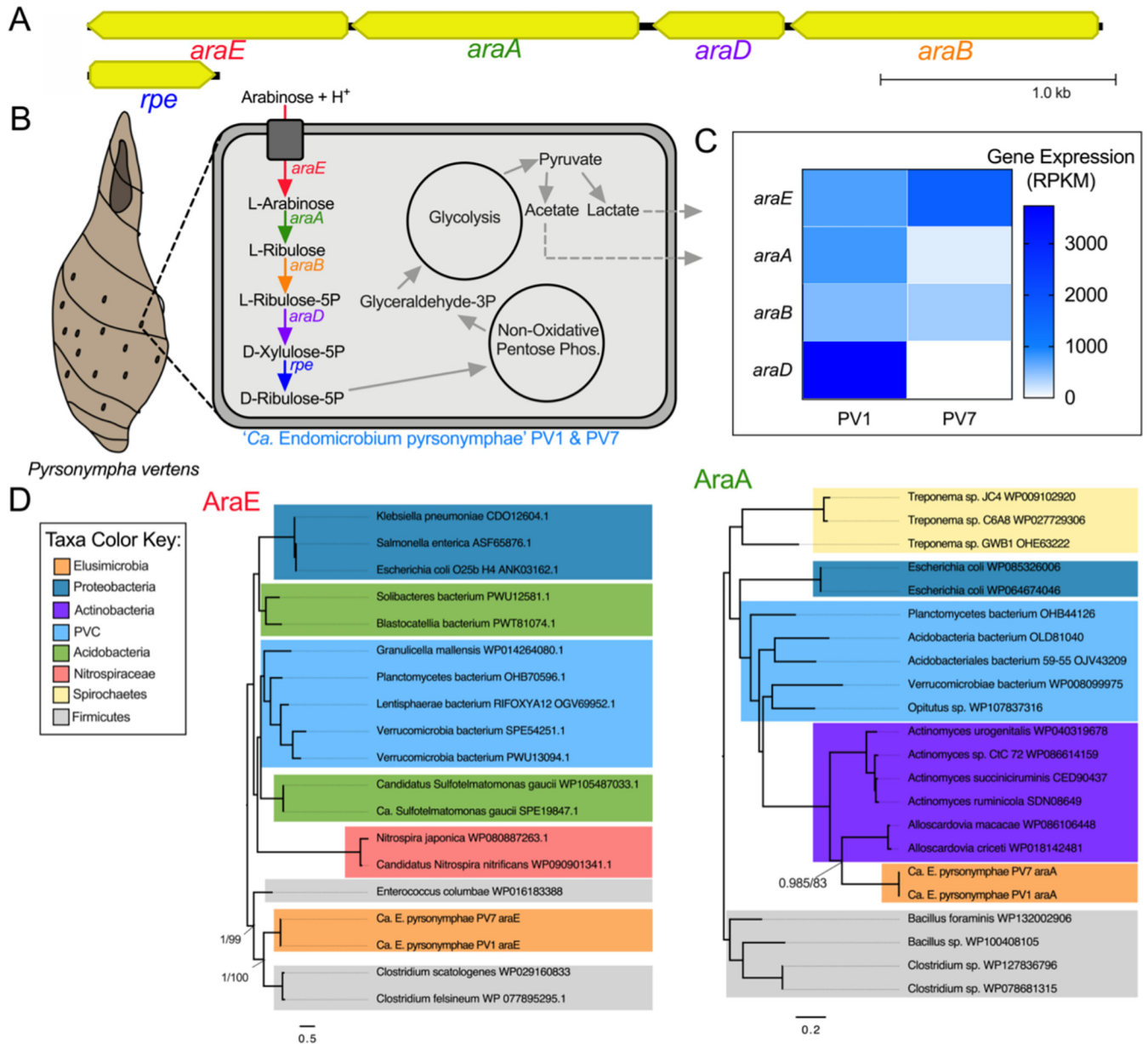


FIG 4 Carbon metabolism and HGT in “*Ca. Endomicrobium pyrsonymphae*.” (A) Gene neighborhood of the genes involved in the metabolism of arabinose in the “*Ca. Endomicrobium pyrsonymphae*” PV1 and PV7 genomes. (B) Diagram of a protist host and an *Endomicrobium* cell showing the inferred metabolic conversions of carbon sources based on gene content data. (C) Gene expression data of genes of interest (rows) pertaining to carbon metabolism in “*Ca. Endomicrobium pyrsonymphae*” PV1 and PV7 (columns). (D) Maximum likelihood phylogenetic trees of amino acid sequences from the transporter (AraE, using substitution model LG+F+G4) and isomerase (AraA, using substitution model LG+I+G4) in the arabinose metabolism pathway. Support values represent the Bayesian posterior probability and bootstrap support values, respectively.

taxa that were suggested by the phylogenies (Tab6 in Data Set S1). Similar data regarding HGT in “*Ca. Endomicrobium trichonymphae*” Rs-D17 have been reported and suggest that *Endomicrobium* symbionts are not cut off from gene flow and HGT (18). This finding contrasts with information known about the endosymbionts of sap-feeding insects, which are traditionally thought to experience little to no gene flow; however, recent analyses suggested that HGT may occur more frequently than previously thought in these symbionts (26).

Natural transformation and competence as a possible mechanism for acquiring genes. Analyses of sequenced genomes of endosymbiotic *Endomicrobium* lineages imply that the acquisition of genes by HGT is relatively common. Interestingly, compared with other endosymbionts, the *Endomicrobium* genomes were enriched in genes

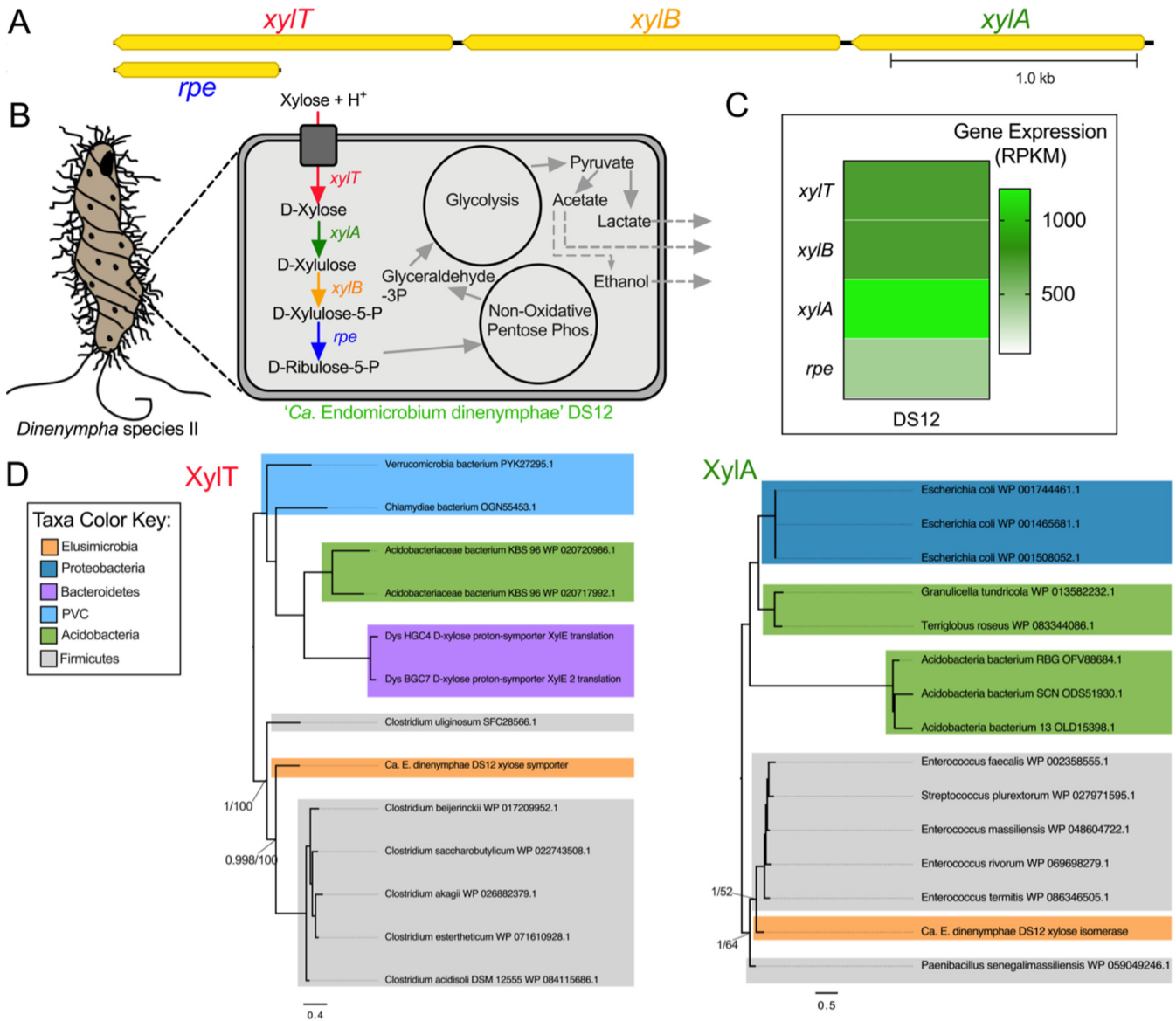


FIG 5 Carbon metabolism and HGT in “*Ca. Endomicrobium dinenymphae.*” (A) Gene neighborhood of the genes involved in the metabolism of xylose in the “*Ca. Endomicrobium dinenymphae.*” DS12 genome. (B) Diagram of a protist host and an *Endomicrobium* cell showing the inferred metabolic conversions of carbon sources based on gene content data. (C) Gene expression data of genes of interest (rows) pertaining to carbon metabolism in “*Ca. Endomicrobium dinenymphae.*” DS12 (column). (D) Maximum likelihood phylogenetic trees of amino acid sequences from the transporter (*XyIT*, using substitution model LG+F+G4) and isomerase (*XylA*, using substitution model LG+G4) in the xylose metabolism pathway. Support values represent the Bayesian posterior probability and bootstrap support values, respectively.

related to the uptake of exogenous DNA and recombination (natural transformation/competence) (see Fig. S5 in the supplemental material). Of special interest are the *Endomicrobium* genes *comEC*, *comEB*, *comF*, *comM*, *ssb*, *drpA*, and *recA* which are all involved in natural transformation in bacteria, such as *Vibrio cholerae* (27).

The ratio of nonsynonymous to synonymous evolutionary changes (*dN/dS* ratio) of these genes supported the hypothesis that selection was acting to maintain the amino acid sequences of their corresponding gene products (*dN/dS* ratio, <1.0), with the exception of *ssb* from TA21 (Fig. 6A). In addition, a transcriptome analysis indicated that these genes were expressed (Fig. 6B). Expression of *comEC*, which encodes a transporter that imports single-stranded DNA across the inner membrane and into the cytoplasm of Gram-negative bacteria (27, 28), was verified in “*Ca. Endomicrobium agillae*” by reverse transcriptase PCR (RT-PCR) and sequencing by using *comEC*-specific primers on a protist cell fraction sample prepared from 20 worker termite hindguts (Fig. 6C).

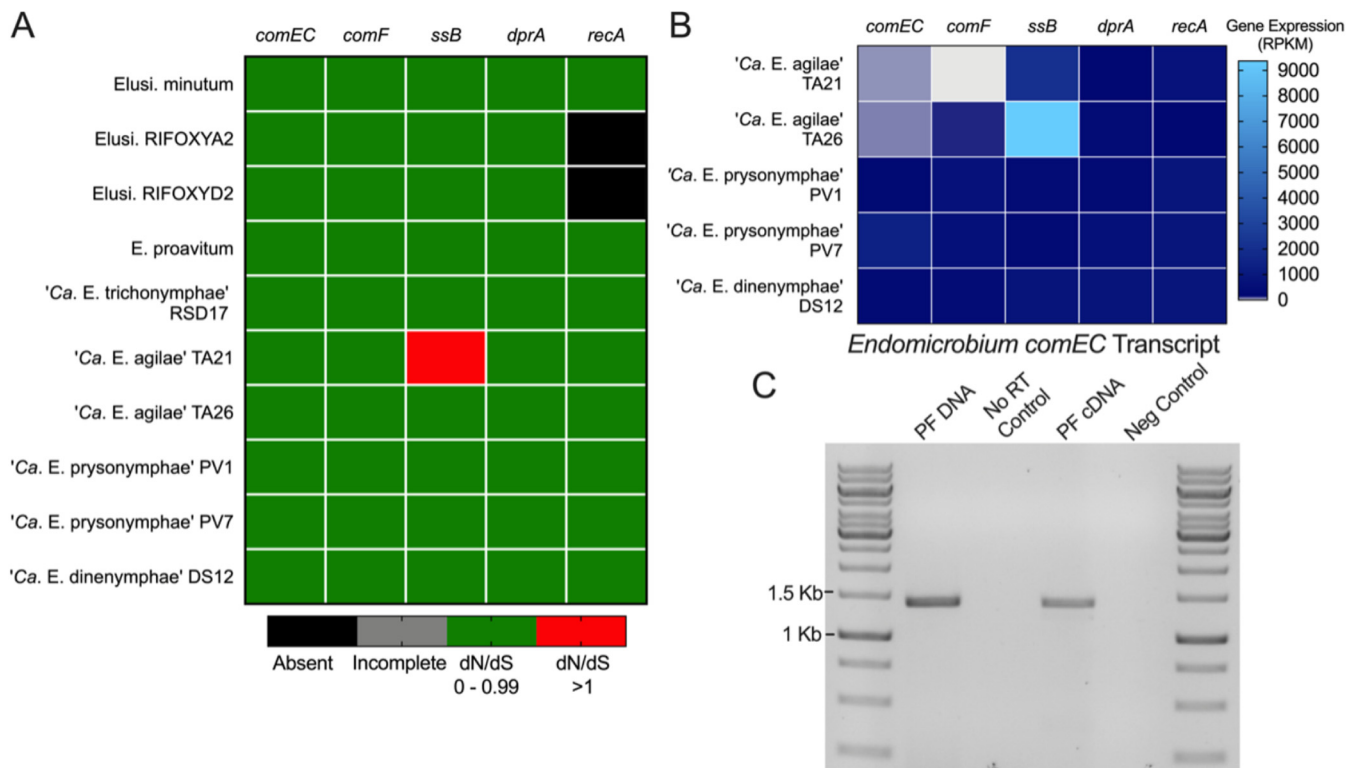


FIG 6 An analysis of the genes involved in a putative competence pathway in *Endomicrobium* spp. (A) Heatmap showing the results of *dN/dS* analyses of genes involved in competence and recombination (columns) from *Endomicrobium* spp. and *Elusimicrobium* relatives (rows). (B) Gene expression data of those genes (columns) in the *Endomicrobium* spp. (rows) presented in this study. (C) RT-PCR gel image of *Endomicrobium comEC* transcript. Samples consisted of protist fraction (PF) DNA (positive control), no-RT control, PF cDNA, and molecular-grade water (negative control). Accession numbers for reference genomes used can be found in Tab2 of Data Set S1.

Together, these data support the hypothesis that genes involved in this competence pathway are both conserved and expressed in these *Endomicrobium* symbionts of hindgut protists of *R. flavipes*.

The competence genes discussed above are involved in the translocation of single-stranded DNA across the inner membrane of Gram-negative bacteria and subsequent recombination. Also present in the genomes of all five *Endomicrobium* species analyzed in this study are genes which encode proteins that are similar to type IV pilins. Some pilins from classes type IV and type II can bind and import double-stranded DNA across the outer membrane and have been shown to work in conjunction with ComEC-type proteins (27, 29). The TA21 and TA26 genomes contained a large chromosomal region devoted to type IV Tad-like pilus synthesis as does *E. proavitum* and *Elusimicrobium minutum*. The bacterium "*Ca. Endomicrobium trichonymphae*" Rs-D17 has a similar region, but it appears that many of the genes have become pseudogenes. The *P. vertens* and *Dinenympha* species II *Endomicrobium* symbionts had genes encoding pilins similar to the type II PulG pilins. In addition, all five genomes possessed a pre-pilin peptidase (PilD) (Fig. S5). A graphical summary of these findings along with a model of how competence may work in these organisms is provided as Fig. S6 in the supplemental material. A list of genes and their putative function can be found in Tab10 and Tab11 of Data Set S1.

Transcriptome analysis of *Endomicrobium* populations inside single protist cells.

Reads from each of the five single-cell metatranscriptomes were then mapped to all contigs in their matching metagenome and to their respective *Endomicrobium* draft genome. For the vast majority of *Endomicrobium* genes, the reads per kilobase per million (RPKM) values were the similar for both mappings, indicating most *Endomicrobium* reads did not match contigs that were not part of *Endomicrobium* draft genomes (Tab7 of Data Set S1).

Transcriptome analyses of the *Endomicrobium* populations inside single protist cells

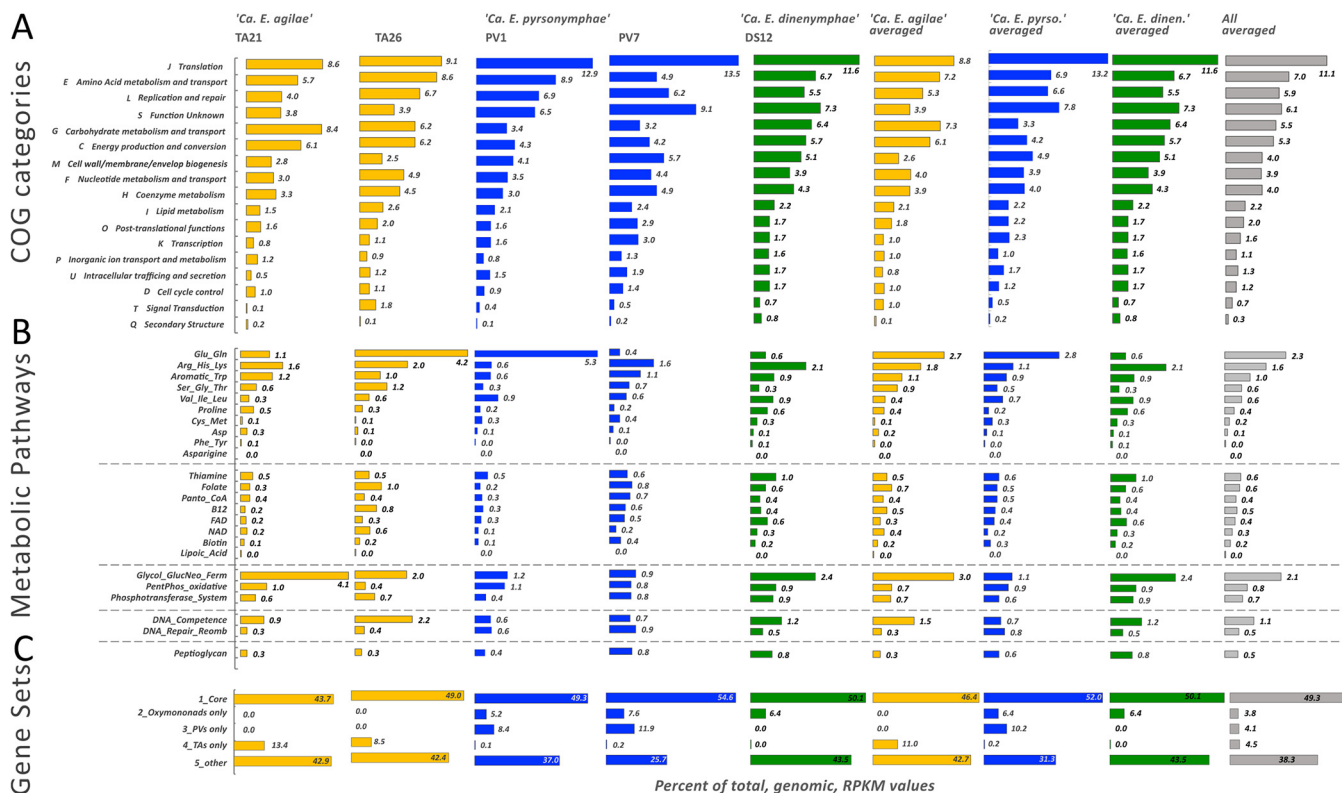


FIG 7 Transcriptome analysis of *Endomicrobium* populations from individual protist cells. (Top) Expression analysis of genes grouped into Clusters of Orthologous Groups (COG) functional categories. (Middle) Expression analysis of genes in certain metabolic pathways pertaining to amino acid and cofactor biosynthesis, DNA processing, and peptidoglycan biosynthesis. (Bottom) Expression of genes grouped by their distributions among the different *Endomicrobium* species. RPKM values for individual cells and averages of genera and all cells are in the various columns.

(from which the five genomes were derived) revealed similar gene expression profiles with a few notable exceptions (Fig. 7). While our sample size for this work was necessarily small and while we were unable to do time-resolved sampling of the hindgut community from single termites, some general trends did appear in the transcriptomic data. Among Clusters of Orthologous Groups (COG) categories, which are quite broad, the expression by each endosymbiont population was relatively similar with one exception being that there was a higher expression of genes related to carbohydrate transport and metabolism in *Trichonympha* hosts and *Dinenympha* (TA21, TA26 and DS12) than that of the *Pyrrsonympha* hosts (PV1 and PV7) (Fig. 7A).

Our analyses which focused on narrower categories, such as genes in related biosynthetic pathways, carbohydrate transport and break down, peptidoglycan synthesis, and DNA uptake and repair, revealed further differences not only between the endosymbionts of different protist species but also between the populations of endosymbionts of individual protist cells of the same type (Fig. 7B). These expression differences between bacterial populations in the same protist cell type are demonstrated by the differences related to the expression of genes in the glutamine and glutamate biosynthesis pathway (Fig. 7B). Overall, this pathway is more highly expressed by the endosymbionts of *Pyrrsonympha* hosts than that of endosymbionts in other protist species, but there was variation in the expression of this pathway between individual protist host cells. For example, in host cell PV1, this pathway represented 5.3% of the total transcriptome reads, mostly from the gene *glnN* encoding a glutamine synthase, whereas in host cell PV7, it comprised only 0.4% (Fig. 7B). A similar variation in this pathway was seen in the TA21 and TA26 transcriptomes. These data suggest that even populations of the bacterium residing in different host cells may not be expressing the same functions at any given point in time.

Core genes, which were shared between all five of the *Endomicrobium* species,

represented an average of 30% to 36% of the transcripts of each endosymbiont population (Fig. 7C). The expressions of genes that were specific to each *Endomicrobium* species ranged from 11% in “*Ca. Endomicrobium pyrsonymphae*” to 22% in “*Ca. Endomicrobium agilae*,” indicating that there is differential gene content and gene expression of endosymbionts of different protist host species (Fig. 7C). Transcriptomic data are in Tab7 and Tab8 in Data Set S1.

DISCUSSION

Single-cell protist metagenomics has enabled the assembly of genome sequences from several protist-associated bacterial symbionts from termite hindguts (1, 3, 4, 30, 31). In this study, we present near-complete draft genome and transcriptome sequences of five endosymbiotic *Endomicrobium* samples, from three different protist species. Those endosymbionts displayed differences with regard to their gene content and its expression. For example, these organisms possessed different carbon usage pathways. One hypothesis that may explain such differences in carbon utilization is that different by-products from protists hydrolyzing and fermenting wood may be provided as carbon sources to the *Endomicrobium* endosymbionts. For example, glucuronate may be present in the cytoplasm of *Trichonympha* spp. because they possessed the enzymes needed to cleave those monomers from polysaccharides found in wood, whereas the other protists, *P. vertens* and *D. species II*, may not be able to generate such monomers, or they may use them for other purposes (see below). If true, this hypothesis suggests that there may be specialization among the protists with regard to polysaccharide hydrolysis in the hindgut of *R. flavipes*. A recent study demonstrated a division of labor among symbiotic protist species in a different termite, namely, *Coptotermes formosanus*. There, protist species produced different hydrolytic enzymes to degrade polysaccharides found in wood (32). These differences in protist functions may also explain why there was a higher expression of genes related to carbohydrate transport and metabolism in the endosymbionts of *Trichonympha* hosts than that of *Pyrsonympha* hosts (Fig. 7).

However, an alternative hypothesis is that metabolites can be partitioned within the host and some are specifically provided to certain symbionts. This hypothesis, if true, may allow the host to control endosymbiont population densities through selective carbon source provision. Such host control of carbon provisioning is thought to operate in nitrogen-fixing root nodule symbioses ensuring that bacterial symbionts continue to provide fixed nitrogen in return for plant-provided carbon. In support of this second hypothesis, the membrane-embedded symbiont “*Ca. Desulfovibrio trichonymphae*,” which cocolonizes the same *Trichonympha* host as “*Ca. Endomicrobium trichonymphae*” Rs-D17, uses malate and citrate as carbon sources, whereas its coinhabitant Rs-D17 likely uses glucuronate and glucose-6-phosphate (31).

Evidence suggests that *Endomicrobium* species have acquired genes by HGT and some of the donor taxa may include termite-associated bacteria. Endosymbiotic lineages of *Endomicrobium*, and their free-living relatives, possess many genes involved in DNA uptake, repair, and recombination (Fig. S5). Our analyses showed that the genes *comEC*, *comEB*, *comF*, *comM*, *ssb*, *drpA*, and *recA* are usually conserved within the *Elusimicrobia* phylum and were expressed in the endosymbiotic *Endomicrobium* species characterized in this study (Fig. 6). Collectively, these genes have been shown to be involved in the translocation of single-stranded DNA across the inner membrane of other Gram-negative bacteria and in homologous recombination (27). The gene *comEC*, in particular, has an important function in this process, as it encodes an essential part of the DNA transporter (27, 28). Using transcriptome data, RT-PCR, and sequencing, we were able to show that the “*Ca. Endomicrobium agilae*” *comEC* gene is expressed in these endosymbionts (Fig. 6). Collectively, the genes involved in the competence pathway comprised between 0.4% and 1.3% of the total transcriptome reads of each endosymbiont population (Fig. 7). These data suggest that *Endomicrobium*

species may have the ability to become competent which may allow them to acquire DNA from the wider termite gut community and could result in HGT.

It is not clear how these organisms transport DNA across their outer membranes. None of these *Endomicrobium* species possessed all the components of a type IV pili-based DNA-translocation system, but “*Ca. Endomicrobium agilae*” TA21 and TA26 contained a near-complete type IV *tad* system which may allow DNA uptake (Fig. S5 and 6). It is also puzzling why each of the five genomes has retained *pilD*, a prepilin peptidase, as well as genes encoding type II and type IV pilins. They may carry out some function in DNA uptake or they may be nonfunctional and are in the process of being lost. If competence is a common trait among the *Endomicrobium* members, it could explain why these organisms have many genes acquired through HGT, and this capability may allow for their rapid adaptation to new and diverse niches. Because hindgut protists phagocytize wood, wood-associated bacteria, and perhaps free-living hindgut bacteria (33), it may be a route through which endosymbiotic *Endomicrobium* members could be exposed to exogenous DNA.

However, it is worth noting that competence is not the only plausible avenue for DNA acquisition in these endosymbionts. HGT could also occur by bacteriophage transduction, conjugation, or other routes. Several lines of evidence indicate that these endosymbionts are susceptible to molecular parasites, such as bacteriophages and plasmids. Previous studies have reported that *Endomicrobium* species possessed several intact defense mechanisms to combat molecular parasites, such as CRISPR-Cas and restriction-modification systems (34, 35). The *Endomicrobium* species sequenced in this study also contained those defense systems. The complete genome sequence of a bacteriophage of an endosymbiont (“*Candidatus Azobacteroides pseudotrichonymphae*”) of a termite hindgut protist has been published previously, indicating that phage infection is not limited to *Endomicrobium* endosymbionts and may be common in termite hindguts (36).

Our analysis of *Endomicrobium* genomes and transcriptomes obtained from single protist cell metagenomes highlight important differences between protist hosts which have led to hypotheses that warrant further investigation. In each case, the major hurdle of testing these hypotheses is the current inability to culture the protist hosts, which restricts their experimental tractability. However, the use of additional approaches should further our understanding of these symbioses by focusing on the host and their symbiont genes, mRNA, and metabolic and protein contents (37–42).

MATERIALS AND METHODS

Termite collection and species identification. *R. flavipes* termites were collected using cardboard traps placed under logs for 2 to 4 weeks at the UConn Campus at Storrs, Connecticut (longitude, -72.262216 ; latitude, 41.806543), and their identity was verified as described previously (7) by amplifying and sequencing the mitochondrial cytochrome oxidase II gene. Termites were maintained in the lab with moistened sand and spruce wood that were initially sterilized.

Single protist cell isolation. Termites from the worker caste were brought into an anaerobic chamber, and their hindguts were dissected with sterile forceps. Hindguts were ruptured in ice-cold Trager’s solution U (TU) (43) and washed three times by centrifuging in $500\ \mu\text{L}$ of TU at 3,000 rpm in an Eppendorf microcentrifuge for 90 seconds. This washed cell suspension was then diluted 10-fold in TU buffer on ice. A $1\text{-}\mu\text{L}$ aliquot of the washed and diluted cell suspension was added to a $9\text{-}\mu\text{L}$ droplet on a glass slide treated with RNase Away reagent (Life Technologies) and UV light. Individual protist cells were isolated using a micromanipulator (Eppendorf CellTram Vario) equipped with a hand-drawn glass capillary. Individual cells were washed three times in $10\text{-}\mu\text{L}$ droplets of TU via micromanipulation, transferring approximately $0.1\ \mu\text{L}$ each time, and finally placed in $10\ \mu\text{L}$ molecular grade phosphate-buffered saline (PBS), flash frozen on dry ice, and immediately stored at -80°C .

Whole-genome and transcriptome amplification and sequencing. The metagenome (DNA) and metatranscriptome (cDNA) from individual protist cells and their associated bacteria were amplified simultaneously at 12 to 24 hours after isolation. Cell lysis and amplification were performed using the Repli-g whole-genome amplification/whole-transcriptome amplification (WGA/WTA) kit (Qiagen). Cells were lysed using a Qiagen lysis buffer followed immediately by incubation on ice. Two samples from each lysed cell were taken and used for whole-genome amplification and whole-transcriptome amplification. These amplifications were carried out using the manufacturer’s standard protocol with the exception that random hexamer primers were used to amplify DNA and cDNA. DNA and cDNA were sheared using a Covaris M220 ultrasonicator according to the manufacturer’s protocol. WGA samples were sheared to a 550-bp insert size using 200 ng of DNA. WTA samples were sheared to a 350-bp insert

size using 100 ng of cDNA. Sequencing libraries were prepared using the TruSeq Nano DNA library prep kit from Illumina according to the manufacturer's protocol. Each sample was prepared with a forward and reverse barcode such that samples could be multiplexed on the same sequencing run. The samples were sequenced using an Illumina NextSeq 1 × 150-bp midoutput run and two NextSeq 1 × 150-bp high-output runs. Metadata regarding WGA/WTA yields can be found in Tab1 in Data Set S1.

Genomic read processing and assembly. Reads were preprocessed before assembly using BMap (44). Reads were filtered for contaminating sequences by mapping reads to reference genomes of potential contamination sources, such as human DNA, human-associated microbiota, and organisms used commonly in our research laboratories. A list of reference genomes used for contamination filtering is provided in Tab2 in Data Set S1. Using BMap scripts, adaptor sequences were trimmed from reads and the last base pair of 151-bp reads was removed. Reads were then trimmed at both ends using a quality score cutoff of Q15. Homopolymers were removed by setting an entropy cutoff of 0.2 and a max G+C cutoff of 90% and by removing reads which possessed stretches of Gs equal to or greater than 23 bases long. In addition, reads which were below a minimum average quality of Q15 and/or less than 50 bases long were removed. Genomic reads were then normalized to a minimum coverage of 2× and a maximum coverage of 50× and then deduplicated using BBnorm. Genomic reads were assembled using the A5 assembly pipeline (45) on the KBase Web server (46). Data regarding metagenome and metatranscriptome read numbers can be found in Tab3 in Data Set S1.

Taxonomic composition of the assemblies. In order to identify the major taxonomic groups in the metagenome assemblies, reads in the assemblies were treated as amplicon sequence variants (ASVs) and mapped against a set of 16S rRNA V4 ASVs from previous work (7) in addition to V4 ASVs from DictDb v3.0 (42), SILVA v138 (prok), and SILVA v132 (euk) (41) rRNA databases. The DictDb and SILVA V4 ASVs were selected because they differed from the in-house ASVs and at least 0.5% of reads in the five metagenomes mapped to them at greater than 98% identity. In total, there were 8 eukaryotic 18S V4 ASVs and 139 16S ASVs in this reference data set (Tab5 in Data Set S1). Reads from the metagenomes were mapped in Geneious Prime 2021 and had to match at 100% identity to a reference ASV over its entire 149- to 150-base length in order to be counted as mapped. A less stringent 98% identity cutoff was also used in order to determine if reads failed to map because our reference set was incomplete.

Genomic binning, draft genome assessment, and annotation. Metagenomic assemblies from single protist host cells and their bacterial symbionts were binned using either 4-mer, 5-mer, or 6-mer frequencies with VizBin (47) and scaffolds at least 1 Kb long. Clustered scaffolds in genomic bins of interest (low G+C content) were selected in VizBin (see Fig. S1 in the supplemental material). Scaffolds from these bins were used in a blastn (48) search against previously sequenced *Elusimicrobia* genomes (Tab4 in Data Set S1). Scaffolds which had a positive hit to other *Elusimicrobia* (at least 70% identity over a 1-kb alignment) were retained in the draft genomes, and scaffolds which did not have a significant hit to other *Elusimicrobia* genomes were used in a second blastn search against the nonredundant (NR) database. From these searches, scaffolds which had positive hits to other *Elusimicrobia* in the NR database were retained. Draft genomes were iteratively polished with the program Pilon (49). These draft genomes were then assessed for contamination and completeness using CheckM which uses lineage-specific marker genes to perform analyses (50). The resulting near-complete draft genomes were annotated initially on the RAST server using a customized RASTtk workflow with options selected to call insertion sequences and prophages (51, 52). Metabolic pathways pertaining to carbon metabolism, amino acid biosynthesis, vitamin biosynthesis, and peptidoglycan biosynthesis were reconstructed from the annotated genomes using pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (53).

Analysis of ribosomal gene phylogeny and average nucleotide identities. Ribosomal 16S genes from each of the *Endomicrobium* draft genomes were trimmed and aligned to references using MUSCLE (54), evolutionary models were tested, and a maximum likelihood (ML) phylogenetic tree was made using IQ-TREE (55). JSpeciesWS (22) was used for determining the genomic average nucleotide identities based on BLAST+ searches (ANi) between the *Endomicrobium* draft genomes and the genome of "*Ca. Endomicrobium trichonymphae*" Rs-D17, which is a close relative (3).

Assembled 18S rRNA genes were retrieved from metagenome assemblies by performing a BLAST+ search using previously published 18S rRNA reference sequences for each protist species as queries (7). When possible, protist 18S rRNA genes were amplified using leftover DNA from WGA samples using universal primers 18SFU (5'-ATGCTTGTCTCAAAGGRYTAAGCCATGC-3') and 18SRU (5'-CWGGTTCACWACGGAAACCTTGTACG-3') (56) as described previously (7) and were sequenced by Sanger sequencing. This confirmation PCR was done on samples TA21, TA26, and DS12. Assembled 18S rRNA genes were aligned to references using MUSCLE (54), and a maximum likelihood (ML) phylogenetic tree was generated using IQ-TREE (55) with model testing. The aligned 18S rRNA sequences used for the construction of the ML tree is available in Text S1 in the supplemental material.

Detection of horizontally acquired genes. Genes that may have been acquired by horizontal gene transfer were tentatively identified using HGTector (25). Briefly, the program took proteins greater than 30 amino acids long from the TA21, TA26, PV1, PV7, and DS12 genomes and compared them using Diamond BLAST (57) to a database derived from single representative species in RefSeq release 208 plus archaeal, fungal, and protist species from the NCBI Reference and Representative databases. This database contained 139,627,894 proteins from 32,128 species. HGTector used 3 *Endomicrobium* species (taxid 1408194) as the "self" data set and 11 *Elusimicrobia* species (74152) as the "close" data set. All other species were considered "distal." In order to be considered a possible horizontally transferred gene, the test protein had to have matches in the self data set and in the distal data set but none in the close data set, as well as pass other statistical tests as outlined in Zhu et al. (25). Each query protein from

the five genomes had to match with an E value of less than $1.0E-5$, have a coverage of $>75\%$, and have $>50\%$ amino acid identity to its matched target sequences in order to be considered for further analysis. These parameters and others are in the HGTector “HGTconfig.yml” and HGTector shell script file in Text S2 in the supplemental material. Putative HGT candidates and putative donor groups are included in Tab6 in Data Set S1.

Analysis of genes involved in competence and recombination. Genes known to be involved in DNA uptake, competence, and recombination were identified in each *Endomicrobium* draft genome based on their NCBI-assigned annotations and homology to reference sequences. The distribution of these genes was then compared across draft genomes and references which included free-living relatives and other endosymbionts. To assess if these genes were complete and if the encoded proteins likely retained their putative functions, homologs of each gene were obtained from genomes of bacteria belonging to the phylum *Elusimicrobia* and aligned with MUSCLE (54). Phylogenetic trees were generated using IQ-TREE (55) with model testing, and support values were generated using the “-abayes” and “-bb 1000” commands. The resulting phylogenetic trees were used along with the MUSCLE (54) alignments to perform a *dN/dS* analysis using the program Codeml which is a part of the PAML and PAMLX packages (58, 59). Genes likely involved with competence and DNA repair are in Tab10 in Data Set S1.

Mapping transcriptome reads to draft genomes. Transcriptome sequencing (RNA-seq) metatranscriptome reads were quality trimmed and filtered as described above and error corrected in Geneious R11 (60) using BBNorm with default settings. To remove rRNA reads before mapping, rRNA sequences were identified from each metagenome assembly using RNAmmer (61); reads were mapped to these as well as rRNA references from RefSeq (62), SILVA (63), and DictDb (64) databases using BBMap (44). Remaining metatranscriptome reads were then mapped to all contigs in their matching metagenome and to the respective *Endomicrobium* draft genome. These mappings were done in Geneious R11 using Bowtie 2 (65) with alignment type set to “End to End” and using the “Medium Sensitivity” preset. Expression levels were then calculated in Geneious R11; ambiguously mapped reads were excluded from the calculations. RPKM values for genes in from each mapping are given in Tab7 and Tab8 in Data Set S1.

Verification of *comEC* expression by RT-PCR. Primers were designed to amplify *comEC* from “*Ca. Endomicrobium agilae*” in Geneious R11 using Primer3 (66) (primers were endo_comec_F [5'-ATTGCTGTGTTGAGAGT-3'] and endo_comec_R [5'-CCTGTCCTGTGCTTTAG-3']). Twenty termites were used to prepare RNA and cDNA samples for RT-PCR analysis. Termite hindguts were dissected and ruptured in TU on ice in an anaerobic chamber. Hindgut contents were washed with ice-cold TU three times at 3,000 rpm in an Eppendorf microcentrifuge for 90 seconds and then lysed in 1 mL of TRIzol reagent (Thermo Fisher Scientific). RNA was isolated per the manufacturer's protocol and treated with Turbo DNase (Thermo Fisher Scientific) following the manufacturer's protocol for 50- μ L reactions using ~ 5.6 mg of total RNA in a 20- μ L volume. A total of 20 μ L of the DNase-treated RNA was then used as the template for cDNA synthesis using SuperScript IV reverse transcriptase (Thermo Fisher Scientific) following the manufacturer's protocol for first-strand synthesis primed with random hexamers. The resulting cDNA was treated with *Escherichia coli* RNaseH for 20 minutes at 37°C.

RT-PCRs were performed using the *Endomicrobium comEC* primers with RNaseH-treated cDNA serving as the template and the no-RT control consisting of DNase-treated RNA that did not undergo cDNA synthesis. RT-PCR was performed using Phusion polymerase (Thermo Fisher Scientific) with high-fidelity (HF) buffer and dimethyl sulfoxide (DMSO). Cycling conditions were as follows: initial denaturation at 94°C for 3 minutes, followed by 35 cycles of 94°C for 45 seconds, annealing at 59°C for 30 seconds, and extension at 72°C for 45 seconds. Final extension was done 72°C for 10 minutes. Hindgut DNA (washed protist cell fractions from five hindguts in molecular-grade Tris-EDTA buffer) was used as a positive PCR control. RT-PCR products were visualized using a 1% agarose gel with ethidium bromide. Products were purified using the Monarch DNA gel purification kit (New England Biolabs) and Sanger sequenced.

Data availability. Raw reads and assemblies were submitted to NCBI GenBank under BioProject PRJNA644342 (whole-genome assemblies are JAIXMW01, JAIXMX01, JAIXMY01, JAIXMZ01, and JAIXNA01).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 3.2 MB.

TEXT S1, TXT file, 0.1 MB.

TEXT S2, TXT file, 0.01 MB.

FIG S1, JPG file, 0.8 MB.

FIG S2, JPG file, 1 MB.

FIG S3, JPG file, 1.5 MB.

FIG S4, JPG file, 0.9 MB.

FIG S5, JPG file, 0.4 MB.

FIG S6, JPG file, 0.7 MB.

TABLE S1, DOCX file, 0.1 MB.

ACKNOWLEDGMENT

This research was funded by the National Science Foundation (NSF) division of Emerging Frontiers in Research and Innovation in Multicellular and Inter-kingdom Signaling award number 1137249 (D.J.G. and J.G.).

REFERENCES

- 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 N₂ fixation to cellulolysis within protist cells in termite gut. *Science* 322:1108–1109. <https://doi.org/10.1126/science.1165578>.
- Ohkuma M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends Microbiol* 16:345–352. <https://doi.org/10.1016/j.tim.2008.04.004>.
- 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>.
- 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, 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>.
- 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>.
- Stephens ME, Gage DJ. 2020. Single-cell amplicon sequencing reveals community structures and transmission trends of protist-associated bacteria in a termite host. *PLoS One* 15:e0233065-19. <https://doi.org/10.1371/journal.pone.0233065>.
- Benjamino J, Graf J. 2016. Characterization of the core and caste-specific microbiota in the termite, *Reticulitermes flavipes*. *Front Microbiol* 7:171. <https://doi.org/10.3389/fmicb.2016.00171>.
- Boucias DG, Cai Y, Sun Y, Lietze VU, Sen R, Raychoudhury R, Scharf ME. 2013. The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Mol Ecol* 22:1836–1853. <https://doi.org/10.1111/mec.12230>.
- 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>.
- 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>.
- 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>.
- Zheng H, Dietrich C, Thompson C, 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/jsm2.ME14169>.
- 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>.
- 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>.
- 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>.
- 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>.
- 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>.
- Brune A, Emerson D, Breznak JA. 1995. The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. *Appl Environ Microbiol* 61:2681–2687. <https://doi.org/10.1128/aem.61.7.2681-2687.1995>.
- Yang H, Schmitt-Wagner D, Stingl U, Brune A. 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ Microbiol* 7:916–932. <https://doi.org/10.1111/j.1462-2920.2005.00760.x>.
- James ER, Tai V, Scheffrahn RH, Keeling PJ. 2013. *Trichonympha burlesquei* n. sp. from *Reticulitermes virginicus* and evidence against a cosmopolitan distribution of *Trichonympha agilis* in many termite hosts. *Int J Syst Evol Microbiol* 63:3873–3876. <https://doi.org/10.1099/ijs.0.054874-0>.
- Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32:929–931. <https://doi.org/10.1093/bioinformatics/btv681>.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.
- Huson DH, Beier S, Flade I, Górská A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R. 2016. MEGAN community edition—interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 12:e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>.
- Zhu Q, Kosoy M, Dittmar K. 2014. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. *BMC Genomics* 15:717. <https://doi.org/10.1186/1471-2164-15-717>.
- López-Madrugal S, Gil R. 2017. Et tu, brute? Not even intracellular mutualistic symbionts escape horizontal gene transfer. *Genes (Basel)* 8:247. <https://doi.org/10.3390/genes8100247>.
- Seitz P, Blokesch M. 2013. DNA-uptake machinery of naturally competent *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 110:17987–17992. <https://doi.org/10.1073/pnas.1315647110>.
- Pimentel ZT, Zhang Y. 2018. Evolution of the natural transformation protein, ComEC, in bacteria. *Front Microbiol* 9:2980. <https://doi.org/10.3389/fmicb.2018.02980>.
- Ellison CK, Dalia TN, Vidal Ceballos A, Wang JCY, Biais N, Brun YV, Dalia AB. 2018. Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in *Vibrio cholerae*. *Nat Microbiol* 3: 773–780. <https://doi.org/10.1038/s41564-018-0174-y>.
- Utami YD, Kuwahara H, Igai K, Murakami T, Sugaya K, Morikawa T, Nagura Y, Yuki M, Deevong P, Inoue T, Kihara K, Lo N, Yamada A, Ohkuma M, Hongoh Y. 2019. Genome analyses of uncultured TG2/ZB3 bacteria in “*Margulisbacteria*” specifically attached to ectosymbiotic spirochetes of protists in the termite gut. *ISME J* 13:455–467. <https://doi.org/10.1038/s41396-018-0297-4>.
- Sato T, Hongoh Y, Noda S, Hattori S, Ui S, Ohkuma M. 2009. *Candidatus* *Desulfobivrio trichonymphae*, a novel intracellular symbiont of the flagellate *Trichonympha agilis* in termite gut. *Environ Microbiol* 11:1007–1015. <https://doi.org/10.1111/j.1462-2920.2008.01827.x>.
- Nishimura Y, Otagiri M, Yuki M, Shimizu M, Inoue J, Moriya S, Ohkuma M. 2020. Division of functional roles for termite gut protists revealed by single-cell transcriptomes. *ISME J* 14:2449–2460. <https://doi.org/10.1038/s41396-020-0698-z>.

33. Brune A. 2014. Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol* 12:168–180. <https://doi.org/10.1038/nrmicro3182>.
34. 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/eww227>.
35. 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>.
36. Pramono AK, Kuwahara H, Itoh T, Toyoda A, Yamada A, Hongoh Y. 2017. Discovery and complete genome sequence of a bacteriophage from an obligate intracellular symbiont of a cellulolytic protist in the termite gut. *Microbes Environ* 32:112–117. <https://doi.org/10.1264/jsm2.ME16175>.
37. Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, Yang EC, Duffy S, Bhattacharya D. 2011. Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332:714–717. <https://doi.org/10.1126/science.1203163>.
38. Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Zárský V, Barlow LD, Herman EK, Soukal P, Hroudová M, Doležal P, Stairs CW, Roger AJ, Eliáš M, Dacks JB, Vlček Č, Hampl V. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol* 26:1274–1284. <https://doi.org/10.1016/j.cub.2016.03.053>.
39. Mangot JF, Logares R, Sánchez P, Latorre F, Seeleuthner Y, Mondy S, Sieracki ME, Jaillon O, Wincker P, Vargas C, De, Massana R. 2017. Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. *Sci Rep* 7:41498. <https://doi.org/10.1038/srep41498>.
40. Vacek V, Novák L, Treitli SC, Táborský P, Cepicka I, Kolísko M, Keeling PJ, Hampl V. 2018. Fe-S cluster assembly in oxymonads and related protists. *Mol Biol Evol* 35:2712–2718. <https://doi.org/10.1093/molbev/msy168>.
41. Kolísko M, Boscaro V, Burki F, Lynn DH, Keeling PJ. 2014. Single-cell transcriptomics for microbial eukaryotes. *Curr Biol* 24:R1081–R1082. <https://doi.org/10.1016/j.cub.2014.10.026>.
42. Hamann E, Tegetmeyer HE, Riedel D, Littmann S, Ahmerkamp S, Chen J, Hach PF, Strous M. 2017. Syntrophic linkage between predatory Carpediemonas and specific prokaryotic populations. *ISME J* 11:1205–1217. <https://doi.org/10.1038/ismej.2016.197>.
43. Trager W. 1934. The cultivation of a cellulose-digesting flagellate, trichomonas termopsidis, and of certain other termite protozoa. *Biol Bull* 66:182–190. <https://doi.org/10.2307/1537331>.
44. Bushnell B. 2014. BMAP: a fast, accurate, splice-aware aligner.
45. Coil D, Jospin G, Darling AE. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31:587–589. <https://doi.org/10.1093/bioinformatics/btu661>.
46. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Dvoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, et al. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.
47. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, der Maaten LV, Vlassis N, Wilmes P. 2015. VizBin—an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* 3:1. <https://doi.org/10.1186/s40168-014-0066-1>.
48. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
49. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
50. 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. *Cold Spring Harb Lab Press Method* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
51. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
52. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
53. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28:27–30. <https://doi.org/10.1093/nar/28.1.27>.
54. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
55. Nguyen LT, 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>.
56. Tikhonenkov DV, Janoušková J, Keeling PJ, Mylnikov AP. 2016. The morphology, ultrastructure and SSU rRNA gene sequence of a new freshwater flagellate, *Neobodo borokensis* n. sp. (Kinetoplastea, Excavata). *J Eukaryot Microbiol* 63:220–232. <https://doi.org/10.1111/jeu.12271>.
57. Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
58. Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556. <https://doi.org/10.1093/bioinformatics/13.5.555>.
59. Xu B, Yang Z. 2013. PamlX: a graphical user interface for PAML. *Mol Biol Evol* 30:2723–2724. <https://doi.org/10.1093/molbev/mst179>.
60. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
61. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>.
62. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35:61–65. <https://doi.org/10.1093/nar/gkl842>.
63. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and Web-based tools. *Nucleic Acids Res* 41:590–596. <https://doi.org/10.1093/nar/gks1219>.
64. Mikaelyan A, Köhler T, Lampert N, Rohland J, Boga H, Meuser K, Brune A. 2015. Classifying the bacterial gut microbiota of termites and cockroaches: a curated phylogenetic reference database (DictDb). *Syst Appl Microbiol* 38:472–482. <https://doi.org/10.1016/j.syapm.2015.07.004>.
65. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
66. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res* 40:e115. <https://doi.org/10.1093/nar/gks596>.