

# Genome Evolution and Nitrogen Fixation in Bacterial Ectosymbionts of a Protist Inhabiting Wood-Feeding Cockroaches

Vera Tai,<sup>a,e\*</sup> Kevin J. Carpenter,<sup>b\*</sup> Peter K. Weber,<sup>b</sup> Christine A. Nalepa,<sup>c</sup> Steve J. Perlman,<sup>d,e</sup> Patrick J. Keeling<sup>a,e</sup>

Department of Botany, University of British Columbia, Vancouver, BC, Canada<sup>a</sup>; Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California, USA<sup>b</sup>; Department of Entomology, North Carolina State University, Raleigh, North Carolina, USA<sup>c</sup>; Department of Biology, University of Victoria, Victoria, BC, Canada<sup>d</sup>; Integrated Microbial Biodiversity Program, Canadian Institute for Advanced Research, Toronto, ON, Canada<sup>e</sup>

## ABSTRACT

By combining genomics and isotope imaging analysis using high-resolution secondary ion mass spectrometry (NanoSIMS), we examined the function and evolution of *Bacteroidales* ectosymbionts of the protist *Barbulanympha* from the hindguts of the wood-eating cockroach *Cryptocercus punctulatus*. In particular, we investigated the structure of ectosymbiont genomes, which, in contrast to those of endosymbionts, has been little studied to date, and tested the hypothesis that these ectosymbionts fix nitrogen. Unlike with most obligate endosymbionts, genome reduction has not played a major role in the evolution of the *Barbulanympha* ectosymbionts. Instead, interaction with the external environment has remained important for this symbiont as genes for synthesis of transporters, outer membrane proteins, lipopolysaccharides, and lipoproteins have been retained. The ectosymbiont genome carried two complete operons for nitrogen fixation, a urea transporter, and a urease, indicating the availability of nitrogen as a driving force behind the symbiosis. NanoSIMS analysis of *C. punctulatus* hindgut symbionts exposed *in vivo* to <sup>15</sup>N<sub>2</sub> supports the hypothesis that *Barbulanympha* ectosymbionts are capable of nitrogen fixation. This genomic and *in vivo* functional investigation of protist ectosymbionts highlights the diversity of evolutionary forces and trajectories that shape symbiotic interactions.

## IMPORTANCE

The ecological and evolutionary importance of symbioses is increasingly clear, but the overall diversity of symbiotic interactions remains poorly explored. In this study, we investigated the evolution and nitrogen fixation capabilities of ectosymbionts attached to the protist *Barbulanympha* from the hindgut of the wood-eating cockroach *Cryptocercus punctulatus*. In addressing genome evolution of protist ectosymbionts, our data suggest that the ecological pressures influencing the evolution of extracellular symbionts clearly differ from intracellular symbionts and organelles. Using NanoSIMS analysis, we also obtained direct imaging evidence of a specific hindgut microbe playing a role in nitrogen fixation. These results demonstrate the power of combining NanoSIMS and genomics tools for investigating the biology of uncultivable microbes. This investigation paves the way for a more precise understanding of microbial interactions in the hindguts of wood-eating insects and further exploration of the diversity and ecological significance of symbiosis between microbes.

Bacteria form mutualistic symbioses with a great diversity of eukaryotic organisms (1–5). By providing beneficial and often essential functions for their hosts, thus expanding the range of ecological niches that can be occupied, the ecological and evolutionary importance of symbioses is increasingly clear, but the overall diversity of symbiotic interactions remains poorly explored (6–9). In particular, extracellular symbioses have been relatively understudied. These encompass incredibly diverse associations ranging from obligate to facultative for the host and/or symbiont, and include symbionts that are extracellular but inside a multicellular host, such as animal gut bacteria, as well as symbionts that reside on the surface of host cells, referred to here as ectosymbionts, such as those attached to protists (see, for example, references 10 to 14).

Genomic analyses in conjunction with culture-independent methods have provided the best insight into the functioning and evolution of bacterial symbioses. These investigations have generally focused on intracellular symbionts (referred to here as endosymbionts) of animals and, more recently, of protists, including ciliates, amoebae, and parabasalids (15–18), revealing common patterns of genome evolution. The genomes of vertically transmitted, obligate endosymbionts are typically reduced in gene content,

are AT biased, and carry an elevated number of pseudogenes compared to their free-living relatives (15, 16, 19–21). These genomic features are thought to be the result of population bottlenecks, genetic drift, and relaxed selection on genes that are no longer needed as a free-living bacterium adapts to a symbiotic lifestyle (20, 22, 23).

Received 25 February 2016 Accepted 18 May 2016

Accepted manuscript posted online 27 May 2016

Citation Tai V, Carpenter KJ, Weber PK, Nalepa CA, Perlman SJ, Keeling PJ. 2016. Genome evolution and nitrogen fixation in bacterial ectosymbionts of a protist inhabiting wood-feeding cockroaches. *Appl Environ Microbiol* 82:4682–4695. doi:10.1128/AEM.00611-16.

Editor: H. Goodrich-Blair, University of Wisconsin–Madison

Address correspondence to Vera Tai, vtai@cfenet.ubc.ca.

\* Present address: Vera Tai, BC Centre for Excellence in HIV/AIDS, St. Paul's Hospital, Vancouver, BC, Canada; Kevin J. Carpenter, Biological Sciences Department, California State Polytechnic University, Pomona, California, USA.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.00611-16>.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

While the endosymbionts of diverse hosts appear to share convergent evolutionary outcomes due to common selective pressures, far less is known about ectosymbionts. These associations are also common in certain environments, appear to be every bit as specific as intracellular associations, and can involve elaborate structural adaptations in both the host and symbiont (10, 24–26). The specificity and integration of these ectosymbionts suggest that these bacteria have evolved similarly to obligate intracellular endosymbionts. However, genome-wide analyses of ectosymbionts, and evidence for their functional roles, have lagged behind those of endosymbionts.

The hindguts of wood-eating lower termites and related *Cryptocercus* cockroaches provide an excellent model system to investigate differences between endo- and ectosymbionts and the biology of symbiosis in general. Here, a highly endemic and diverse community of microbial symbionts, including protists (mainly parabasalids and oxymonads), bacteria, and archaea, thrives solely on ingested wood lignocellulose, a highly recalcitrant and nutrient-poor substance. The symbionts produce food in the form of acetate through the fermentation of lignocellulose and essential nitrogenous nutrients for the insect host (27, 28). Within this community of hindgut symbionts, perhaps even more tightly integrated are the numerous endo- and ectosymbiotic associations between hindgut protists and bacteria. These endosymbionts (bacteria inside a protist cell) and ectosymbionts (bacteria on the surface of a protist cell) form highly specific associations often resulting in coevolution among protists, their bacterial symbionts, and also the insect hosts (29–31). In many cases, endo- or ectosymbionts of protists and free-living bacteria in the hindguts are closely related, providing the opportunity to directly compare and trace the evolution of these two modes of symbiosis (32).

Nitrogen fixation is likely an essential function for lignocellulose-digesting gut communities because of the low levels of nitrogen available from this diet. Molecular and genomic data indicate that hindgut symbionts from wood-eating termites and cockroaches possess a diversity of nitrogenase genes (16, 33–37), but actual nitrogen fixation by specific symbionts has rarely been demonstrated. For instance, a few cultivated bacteria have been shown to fix nitrogen, but these are likely not the dominant nitrogen fixers in the hindgut (37–40). As well, in using cultivation-independent methods, results of PCR, acetylene reduction, and  $^{15}\text{N}_2$  incorporation assays suggest that the majority of nitrogen fixation in the hindguts of *Hodotermopsis sjoestedti* is performed by spirochete endosymbionts of *Eucomonympha* (41).

For identifying nitrogen-fixing symbionts *in situ*, stable isotope probing (SIP) with  $^{15}\text{N}_2$  followed by high lateral-resolution imaging secondary ion mass spectrometry (SIMS) with a Cameca NanoSIMS has been shown to be a successful cultivation-independent approach (42). This combined SIP-NanoSIMS approach (also known as nanoSIP) allows direct imaging of individual cells based on their isotopic composition, thereby indicating which cells have used an isotopically labeled substrate (e.g.,  $^{15}\text{N}_2$ ). A similar approach was also successful for examining carbon transfer between protists and their symbionts in termite hindgut communities using  $^{13}\text{C}$ -labeled cellulose (43).

In this study, we combined genomics and NanoSIMS analysis to examine the function and evolution of *Bacteroidales* ectosymbionts of *Barbulanympha*, a lignocellulose-digesting parabasalid protist from the hindgut of the wood-eating cockroach *Cryptocercus punctulatus*. *Bacteroidales* have a predilection for forming in-

tracellular and extracellular associations, and complete genomes are available from endosymbiotic, host-associated, and free-living relatives, together providing a unique opportunity to investigate the evolution of diverse lifestyles within a single bacterial order. We examined the beneficial functions that *Barbulanympha* ectosymbionts may provide—in particular, whether they fix nitrogen. We also examined the selection pressures faced by ectosymbionts compared to endosymbionts of protists and the evolutionary consequences of an ectosymbiotic lifestyle.

## MATERIALS AND METHODS

**Cockroach collections and identification of *Barbulanympha* from the hindgut.** *Cryptocercus punctulatus* adults and nymphs were collected from Mountain Lake Biological Station, Giles County, VA (37.364°N, 80.519°W). Adults collected in September 2011 and September 2012 were used to isolate single *Barbulanympha* cells for genomic sequencing. Nymphs collected in April 2011 were used for stable isotope labeling experiments, scanning electron microscopy (SEM), and NanoSIMS.

Using light microscopy (for genomics) or SEM (for NanoSIMS), *Barbulanympha* cells were visually identified from the *C. punctulatus* hindgut based on published descriptions (13, 24, 44). *Barbulanympha* has a distinct morphology compared to other large protists in the hindgut (e.g., *Trichonympha* and *Eucomonympha*).

**DNA isolation and library preparation from single *Barbulanympha* cells.** To isolate cells for sequencing, the hindgut was dissected from live cockroaches and the contents were collected into Trager's solution U (45). Viewed under an inverted light microscope, *Barbulanympha* cells were individually transferred to clean Trager's solution 2 times to remove unattached bacteria and then into a microcentrifuge tube using pulled glass capillary tubes attached to a syringe via rubber tubing. Three cells (named Barb4, Barb6, and Barb7) and one cell (Barb6XT) were isolated from a cockroach collected in September 2011 and September 2012, respectively. From each cell, approximately 1 ng of DNA was isolated using the MasterPure DNA extraction kit (EpiCentre, Madison, WI).

From the DNA of one cell (Barb6XT), a sequencing library was generated using the Nextera XT kit (Illumina, San Diego, CA) and 250-bp paired-end reads were sequenced using Illumina MiSeq. Using Sickle (<https://github.com/najoshi/sickle>) (46), reads were trimmed where the quality score fell below 20, and only reads at least 150 bp in length were retained. Transposon sequences used to generate the Nextera XT library were also trimmed from the ends of the reads using a custom Perl script (removeTP.pl), available through <https://github.com/JFP-Laboratory/Genomics>.

Genomic DNA was isolated from three other cells (Barb4, Barb6, and Barb7) and amplified using phi29 DNA polymerase from the Genomiphi v2 kit per the manufacturer's protocol (GE Healthcare Life Sciences, Mississauga, ON, Canada). The DNA was mechanically sheared with an S220 focused ultrasonicator (Covaris, Woburn, MA), and sequencing libraries were constructed from 400- to 600-bp fragments using a TruSeq kit (Illumina). These libraries were sequenced using Illumina HiSeq, generating 100-bp paired-end reads. Reads were trimmed where the quality score fell below 20, and only reads at least 50 bp in length were retained.

**Sequence assembly and binning.** Sequences were assembled using Ray v2.1.1-devel with kmers ranging from 25 to 95 (47), ABySS v1.5.1 with kmers ranging from 21 to 95 (48), MIRA v4.0 (49), and the CLC Genomics Workbench (Qiagen Bioinformatics, Redwood City, CA). For the Ray and ABySS assemblies, the kmer that generated the assembly with the longest contig and the longest N50 was chosen for further analysis. The assemblies generated by the different methods were compared using QUAST (50). Depth of read coverage was determined by mapping the reads to the assembled contigs using Bowtie2 (51). Based on the lengths of the contigs generated, the different methods were equivocal (see Table S1 in the supplemental material). Ray assemblies were used for all further analyses, as Ray was able to efficiently assemble the larger HiSeq data sets from Barb4, Barb6, and Barb7.

rRNA genes were predicted from the contigs by the HMMER 3.0 program ([http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/hmm\\_rRNA/](http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/hmm_rRNA/)) (52). rRNA sequences were also assembled using EMIRGE to assess the level of potential contamination from bacteria other than the *Barbulanympha* ectosymbiont (53).

Contigs greater than 1,000 bp were binned, using VizBin, based on pentanucleotide frequencies to identify contigs belonging to the ectosymbiont, distinguishing them from host-derived contigs, other bacteria, and contaminants (54). From the VizBin plot, two bins were identified that included most of the assembled contigs (see Fig. S1 in the supplemental material). All contigs containing 16S rRNAs were contained within one of the bins, and these 16S rRNA sequences were at least 99% similar to previously sequenced *Bacteroidales* ectosymbionts of *Barbulanympha*, so these contigs were designated the ectosymbiont contigs. The contigs within this bin were analyzed again using VizBin, but the plot did not reveal clusters to further bin these contigs. These contigs were curated by using BLASTn to identify contaminating contigs. Contigs with best BLASTn hits (E value  $< 10^{-5}$ ) to plasmid vectors, eukaryotes (potentially host derived), or *Escherichia coli* in GenBank's nucleotide database were excluded.

The second bin contained contigs containing 18S rRNA genes matching *Barbulanympha*, so these contigs were designated host contigs. The GC content of all contigs was analyzed to provide further evidence distinguishing the taxonomic origin of the ectosymbiont and host contigs.

**Genome annotation and genome completeness.** Open reading frames (ORFs) were identified and annotated using Prokka v1.5.2 (55). The KEGG Automatic Annotation Server (<http://www.genome.jp/kegg/kaas/>) (56) and reverse position-specific (RPS) BLAST searches of the Cluster of Orthologous Groups (COG) database (<http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/cog/>) were also used to obtain functional annotations.

Genome completeness was estimated for each set of ectosymbiont contigs using a set of 139 conserved single copy genes (CSCG) determined from all finished bacterial genomes (57). Hidden Markov model profiles (HMMs) of the protein families (Pfam) of these genes were used to search the ORFs from each set of ectosymbiont contigs using HMMER3 (58). To calculate genome completeness, the number of CSCG Pfams found in the ectosymbiont contigs that scored above precalculated cutoffs was determined and then divided by the total number of CSCG Pfams (57). The number of significant hits to each CSCG Pfam indicated the copy number of these genes in each set of ectosymbiont contigs.

**Genome comparisons.** The ORFs annotated from each set of ectosymbiont contigs were compared by clustering the ORFs into homologous gene families using the COGtriangles algorithm (59). ORF clusters unique to an ectosymbiont were further analyzed by BLASTp searches (E value  $< 10^{-5}$ ) to a database of all *Barbulanympha* ectosymbiont ORFs and to the GenBank nonredundant (nr) nucleotide database.

**Polymorphism analysis.** Sequence reads from the 4 data sets were mapped to the Barb6XT contigs using Bowtie2 (51), including mapping the Barb6XT reads to the Barb6XT contigs to examine polymorphisms among the ectosymbionts on a single host cell. The indels were realigned to the contigs using Genome Analysis Tool Kit (GATK) v3.1-1 (60), and single nucleotide polymorphisms (SNPs) were identified using samtools v0.1.18 (61) and snpEffect v4.0b (62). Polymorphisms were identified relative to the Barb6XT ectosymbiont contigs.

**Phylogenetic analyses.** 16S rRNA sequences annotated from the ectosymbiont contigs were aligned using MAFFT L-INS-i (63) with 16S rRNA sequences obtained from GenBank representing the diversity of *Bacteroidales* from clusters most closely related to the *Barbulanympha* ectosymbionts. A separate alignment was made for 18S rRNA sequences identified from the nonectosymbiont contigs. From each alignment, short sequences ( $< 400$  bp) were removed, as were variable and ambiguously aligned sites using Gblocks ([http://molevol.cmima.csic.es/castresana/Gblocks\\_server.html](http://molevol.cmima.csic.es/castresana/Gblocks_server.html)) (64). Phylogenetic trees were generated from a maximum likelihood analysis using RAxML v8.1.3, implementing a gen-

eral time-reversible (GTR) model of nucleotide substitution with the  $\Gamma$  (gamma) model of rate heterogeneity (65). Statistical support for the best maximum likelihood tree was assessed from 1,000 bootstrap replicates.

To generate a NifH phylogeny, the amino acids of ORFs annotated as "nitrogenase iron protein" or *nifH* were aligned using MAFFT with amino acid sequences from GenBank representing the known diversity of NifH, including those from recently sequenced *Bacteroidales* genomes (66). The alignment was trimmed using GBlocks. Using RAxML, LG was determined to be the best protein model and a maximum likelihood phylogenetic tree was calculated with 1,000 bootstrap replicates. Phylogenies were also obtained for the amino acid sequences of NifD-like nitrogenase subunit, NifK-like nitrogenase subunit, urease subunit A, urease subunit B, and the substrate binding proteins of an ABC-type transporter.

**PCR analysis of potential genome rearrangements.** BLASTn searches (E value  $< 10^{-15}$ ) were used to locate syntenic regions and potential rearrangements among the ectosymbiont contigs from different host cells. To exclude the possibility that these rearrangements were due to assembly errors, two potential rearrangement sites were confirmed by PCR and sequencing. For each of the two sites, PCR primers were designed to anneal on either side of the rearrangement sites and used to amplify DNA extracted from the hindgut contents of the same *Cryptocercus punctulatus* cockroaches as used to isolate the *Barbulanympha* cells (MasterPure DNA extraction kit; EpiCentre, Madison, WI).

For the first site, primers 114F (5' TCT TGT CGG GGA TGG TAG TAG) and 116R (5' AAT GGG CTT GGA TTT CGA TGA G), specific for the Barb4, Barb6, and Barb7 ectosymbiont genomes, were designed to anneal to an ABC-type transporter permease and a transposase ORF, respectively. Primer 114F was also paired with primer 1461R (5' AAA GGG CGC GAT TGG TAT G), designed to anneal to a hypothetical ORF and amplify the Barb6XT ectosymbiont genome. A second site was analyzed using primers 172F (5' CTA TGC ACA TAT TCG CGA CAT C) and 174R (5'TCC AGG AGA AGA GAC GAA AC), which anneal to siroheme synthase and an intergenic region, respectively, and were designed to amplify the Barb4 and Barb7 ectosymbiont genomes. Primer 172F was also paired with primer 683R (5' TCC AGG AGA GAC GAA AC), designed to anneal to a transposase and amplify the Barb6 ectosymbiont genome. Standard thermal cycling conditions were used (94°C for 2 min, followed by 30 cycles of 94°C for 30 s, 56°C for 1 min (site 1 primers) or 53°C for 45 s (site 2 primers), and 72°C for 1 min or 45 s, ending with a final extension at 72°C for 10 min). The resulting PCR products were Sanger sequenced (Thermo Fisher Scientific, Waltham, MA).

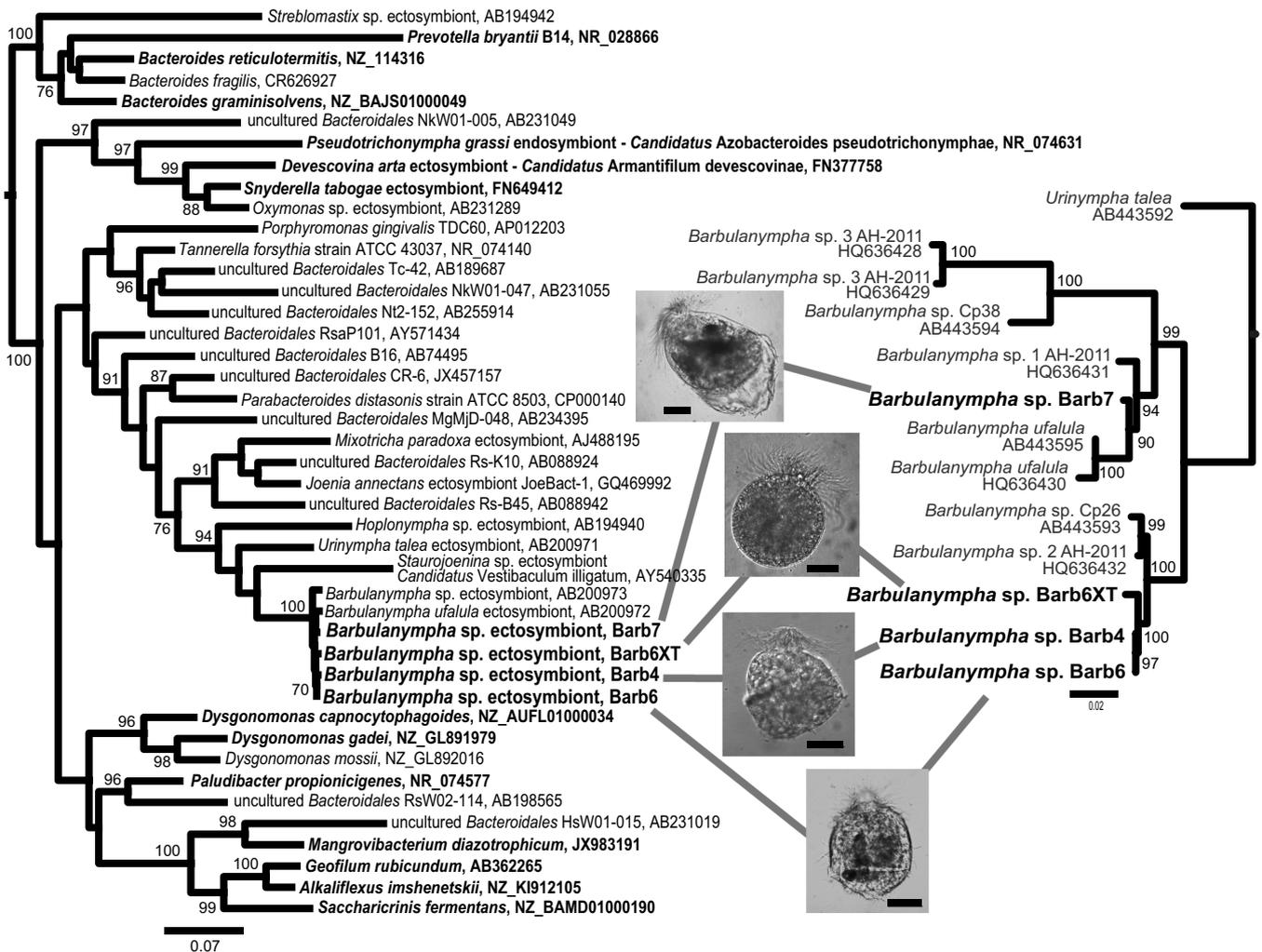
**NanoSIMS analysis of *Barbulanympha* exposed to  $^{15}\text{N}_2$ .** Three live *C. punctulatus* cockroach nymphs were placed in 150-ml glass Wheaton vials with moist cellulose. The air in the vials was equilibrated with  $^{15}\text{N}_2$  gas (Cambridge Isotope Laboratories, Tewksbury, MA) to give a final  $^{15}\text{N}_2/^{14}\text{N}_2$  ratio of 0.38 and an  $\text{O}_2$  content of 26%. During the incubation, the air was not refreshed and the gas lines were closed to the external atmosphere. Following a 2-week exposure to  $^{15}\text{N}$ -enriched air, the cockroaches were removed from the vials and dissected, and the hindgut material was pooled.

As previously described for whole-cell SEM and NanoSIMS imaging, the hindgut material was fixed with 2% glutaraldehyde, postfixed in 1%  $\text{OsO}_4$ , and then gently filtered onto 1- $\mu\text{m}$ -pore membrane filters. The filters were treated with a series of ethanol dehydration steps (50, 70, 90, and 100% ethanol), dried in a carbon dioxide critical point drier, affixed to an SEM stub, and sputter coated with 5-nm iridium (43). SEM was used to identify *Barbulanympha* cells, free-living bacteria, and other protists from the hindgut material to target for NanoSIMS analysis. The locations of these targets were mapped on the SEM stubs using a JEOL 7401F or an FEI Inspect F scanning electron microscope.

High lateral-resolution imaging secondary ion mass spectrometry (SIMS) was performed with a Cameca NanoSIMS 50 (Gennevilliers, France) to directly image the nitrogen isotopic composition of individual cells. High relative  $^{15}\text{N}$  enrichment indicates fixation of  $^{15}\text{N}_2$  or direct

## 16S rRNA (bacteria)

## 18S rRNA (protist)



**FIG 1** Comparison of the phylogenies of *Bacteroidales* ectosymbionts and their *Barbulanympha* hosts. Maximum likelihood trees of 16S rRNA sequences of the ectosymbionts from single *Barbulanympha* cells (on the left) are matched with the 18S rRNA sequences of their host cell (on the right). Light microscopic images of the four isolated cells are shown between the trees, and gray lines link to the sequences in the trees obtained from each cell. Bootstrap support is shown for branches with greater than 70% support. New sequences are in bold, as are *Bacteroidales* that are known to have nitrogen fixation genes. Scale bar = 50  $\mu\text{m}$ .

access to newly fixed  $^{15}\text{N}$ . NanoSIMS was performed with a focused 10-pA, 16-keV  $\text{Cs}^+$  primary ion beam scanned over the sample to eject secondary ions to generate quantitative isotopic images. Scans were between 10 by 10 and 20 by 20  $\mu\text{m}^2$  with 256 by 256 pixels and a dwell time of 1 ms/pixel. Secondary ions were collected for  $^{15}\text{N}^{12}\text{C}^-$  and  $^{14}\text{N}^{12}\text{C}^-$  using electron multipliers in pulse counting mode (43). Secondary electrons were collected for sample visualization. Targets were first sputtered to a depth of  $>50$  nm before data collection to enhance ion yield. Twenty to 30 serial scans were made for each imaged area to obtain sufficient ion counts for the minor isotope.

NanoSIMS ion image data were processed with a custom software package (LIMAGE; L. R. Nittler, Carnegie Institution of Washington, Washington, DC) run with PV-Wave (Rogue Software, Boulder, CO). Quantitative  $^{15}\text{N}/^{14}\text{N}$  isotope ratio images were generated from the  $^{15}\text{N}^{12}\text{C}^-$  and  $^{14}\text{N}^{12}\text{C}^-$  ion images. Isotope data for *Barbulanympha*, its ectosymbionts, and other protists were extracted from hand-drawn regions of interest (ROIs) based on secondary ion and electron images,

excluding areas of low ion counts ( $<5\%$  of maximum). Isotope data for areas of free-living bacteria were extracted based on  $^{14}\text{N}^{12}\text{C}^-$  images using the LIMAGE automated particle-finding software (verified manually). The data are expressed as  $^{15}\text{N}/^{14}\text{N}$  isotope ratios and atom percent excess (APE)  $^{15}\text{N}$  (APE =  $[R_f/(R_i + 1) - R_i/(R_i + 1)] \times 100$ , where  $R_f$  is the measured ratio and  $R_i$  is the initial ratio) (67). The  $R_i$  was determined to be 0.00367 based on uncorrected NanoSIMS measurements. ROIs are determined to be significantly enriched if the  $R_f$  is more than 3 standard errors greater than the  $R_i$ . The Student  $t$  test was used to assess the enrichment differences between *Barbulanympha* cell interiors and their ectosymbionts on a pairwise basis to allow for natural variation in activity among these consortia.

**Accession number(s).** The ectosymbiont contigs from each assembly were deposited at DDBJ/EMBL/GenBank under BioProject accession number PRJNA278755, BioSample accession numbers SAMN03428871, SAMN03428873, SAMN03428874, and SAMN03428875, and GenBank

TABLE 1 Comparison of *Barbulanympha* ectosymbiont Barb6XT genome features with related *Bacteroidales*

Bacterium	Reference	GenBank accession no.	Lifestyle	Habitat	Genome size (Mb)	GC content (%)	# CDS	Coding density (%)	Genome completeness (%) <sup>a</sup>
Barb6XT	This study	LBCW000000000	Attached extracellularly, mutualist	<i>Barbulanympha</i> in cockroach hindgut	3.43	48.7	3,133	82.3	94.2
<i>Parabacteroides distasonis</i> ATCC 8503	11	NC_009615	Host associated, mutualist	Human gut	4.81	45.1	3,849	90.1	98.6*
<i>Tannerella forsythia</i> ATCC 43037	Unpublished data	NC_016610	Host associated, pathogenic	Human periodontal pocket	3.41	47	2,665	84.7	98.6*
<i>Paludibacter propionicegens</i> WB4	68	NC_014734	Free-living	Plant residue from rice field soil	3.69	38.9	2,967	84.4	99.3*
" <i>Candidatus</i> Azobacteroides pseudotriconymphae" CFP2	16	NC_011565	Intracellular, mutualist	<i>Pseudotriconympha</i> in termite hindgut	1.11	33	758	71.2	95.0*

<sup>a</sup> The presence of a set of conserved single copy genes was used to estimate genome completeness in Barb6XT and for comparison also estimated in complete genomes (\*).

accession numbers LBCW000000000, LBCX000000000, LBCY000000000, and LBCZ000000000. The versions described in this paper are versions LBCW010000000, LBCX010000000, LBCY010000000, and LBCZ010000000.

## RESULTS

The DNA sequences obtained from single *Barbulanympha* cells were assembled (see Table S1 in the supplemental material) and a single full-length (or nearly so) 16S rRNA sequence was predicted from each assembly. Truncated 16S rRNA sequences were also predicted from the ends of two Barb6XT and one Barb7 contig, and these exactly matched the longer 16S rRNA sequences or did not overlap. All of the 16S rRNA sequences shared at least 99% identity to previously sequenced 16S rRNA clones from *Bacteroidales* ectosymbionts of *Barbulanympha* (Fig. 1). Fluorescent *in situ* hybridization (FISH) has been previously used to confirm that these sequences, representing the dominant 16S rRNA phylotype from isolated *Barbulanympha* cells, belong to the ectosymbionts of *Barbulanympha* (81). Based on EMIRGE reconstructions and HMM searches, no other 16S rRNA genes were identified, again indicating that *Bacteroidales* ectosymbionts were the dominant bacteria associated with *Barbulanympha* and were the only bacteria sequenced to a significant depth.

The identity of the host cells as *Barbulanympha* was confirmed from 18S rRNA sequences predicted from the assembled contigs. The *Barbulanympha* 18S rRNA sequences from Barb4, Barb6, and Barb6XT shared at least 99.3% identity with each other, while that of Barb7 was more divergent, sharing a mean of 95.1% identity with the other three. Phylogenetic analysis confirmed that the Barb7 host groups with a different clade of *Barbulanympha* (Fig. 1).

The assembled contigs were binned to identify a set of contigs belonging to the ectosymbiont. A distinct bin of contigs was identified that contained all annotated 16S rRNA genes, so these contigs were designated as belonging to the ectosymbiont (see Fig. S1 and Table S2 in the supplemental material). The contigs within this bin also had a distinct GC content, clearly distinguishing them from the remaining contigs, most of which likely belong to the host (see Fig. S2 in the supplemental material). The ectosymbiont contigs were further curated to exclude six contigs hitting *Escherichia coli* with nearly 100% similarity and one contig that matched a plasmid vector from the Barb6 ectosymbiont contigs. Seven contigs that had a best BLASTn hit to a eukaryote were also excluded (2 from Barb4, 3 from Barb6, and 2 from Barb7).

**Completeness of the ectosymbiont genome.** As expected for the genome of one organism, known single-copy genes were

found only once in a set of ectosymbiont contigs with the exception of 2 hits to isopentanylpyrophosphate (IPP) transferases (PFAM01715) in Barb6XT, Barb4, and Barb6. Two versions of this gene are also found in the genomes of closely related *Bacteroidetes* (e.g., *Parabacteroides distasonis*, *Parabacteroides merdae*, and *Tannerella forsythia*). This result indicates that each set of ectosymbiont contigs was from a single bacterial genome and not a mixture of genomes from multiple organisms.

The best assembly resulted from Barb6XT. Genome completeness of the Barb6XT ectosymbiont contigs is estimated at 94.2%; therefore, these contigs contain nearly the complete coding capacity of the bacterial ectosymbiont (Table 1). The total length of the Barb6XT ectosymbiont contigs, the GC content, and the number and density of ORFs fell within the range of free-living and host-associated relatives (Table 1). The Barb4 ectosymbiont contigs also represented a nearly complete genome (92.1%) with a similar total length and GC content (see Table S2 in the supplemental material). The Barb6 and Barb7 ectosymbiont contigs were less complete (87.8 and 61.9%, respectively) but were similar in GC content.

**The ectosymbiont genome content more closely resembles that of free-living bacteria than an endosymbiont from the same environment.** The genes annotated from the *Barbulanympha* ectosymbiont genomes were compared with those of other *Bacteroidales* to detect functionally important differences due to different lifestyles (Table 1; Fig. 1) (11, 16, 68). The Barb6XT ectosymbiont encoded many of the core functions expected from both free-living and symbiotic bacteria, including a complete set of ribosomal proteins except for L17 (L17 was found on a "nonectosymbiont" contig), DNA replication genes, and tRNA synthetases. Genes to synthesize all amino acids were also present, except for the chorismate mutase gene, which is needed to synthesize tyrosine and phenylalanine (Fig. 2).

Energy metabolism was also as expected for an anaerobic microbe. Genes for oxidative phosphorylation and the tricarboxylic acid (TCA) cycle were lacking, suggesting that energy is obtained mainly through glycolysis (although pyruvate kinase was absent) and potentially fumarate respiration (Fig. 2). The ectosymbiont genome did not carry genes for NADH dehydrogenase but contained an operon for five subunits of Na<sup>+</sup>-translocating NADH-quinone reductase (Na<sup>+</sup>-NQR), with a sixth subunit found on a separate contig (Fig. 2). This complex catalyzes the transfer of electrons from NADH to ubiquinone and pumps sodium ions (Na<sup>+</sup>) rather than protons (H<sup>+</sup>) (69). Six subunits encoding the

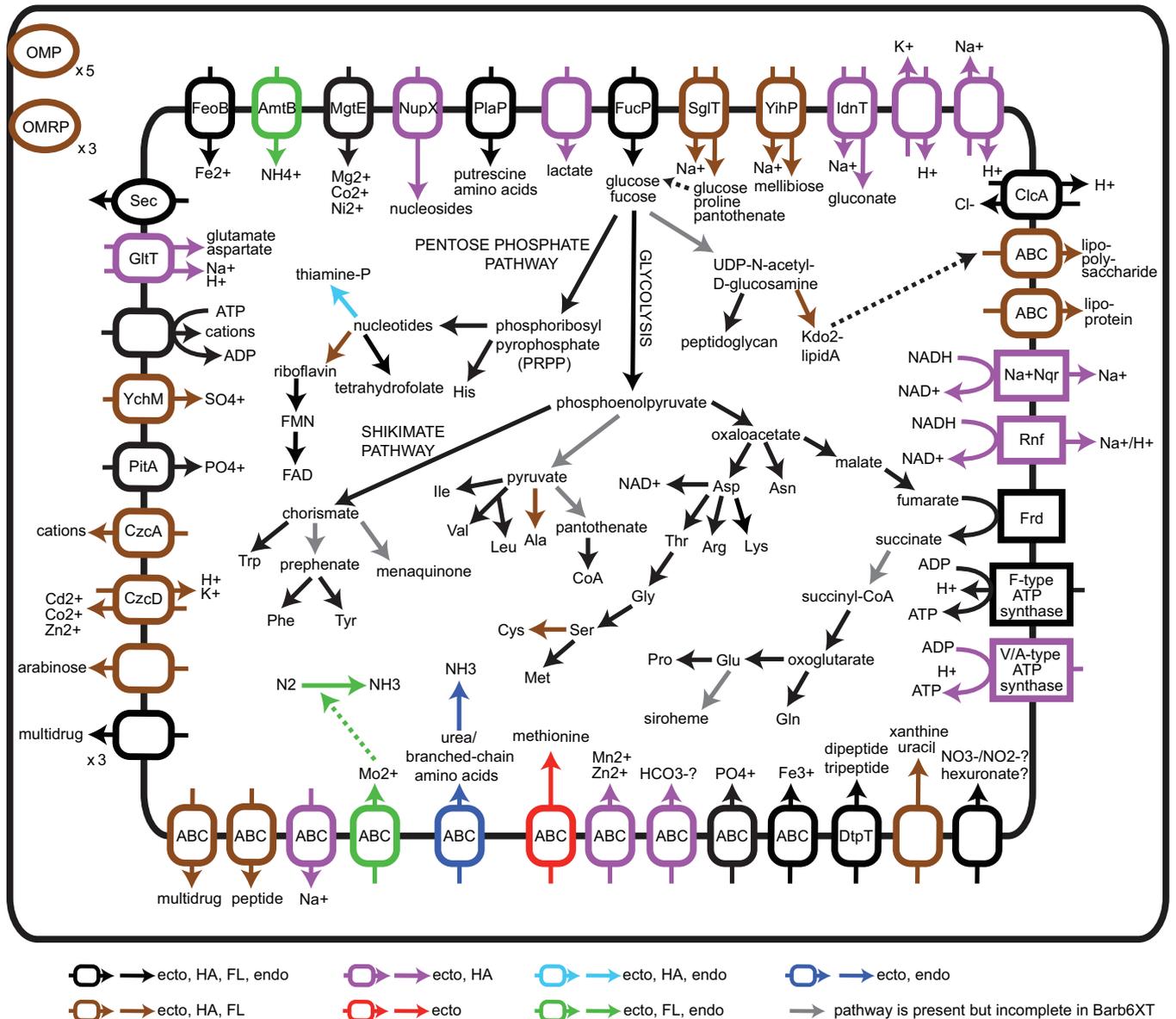
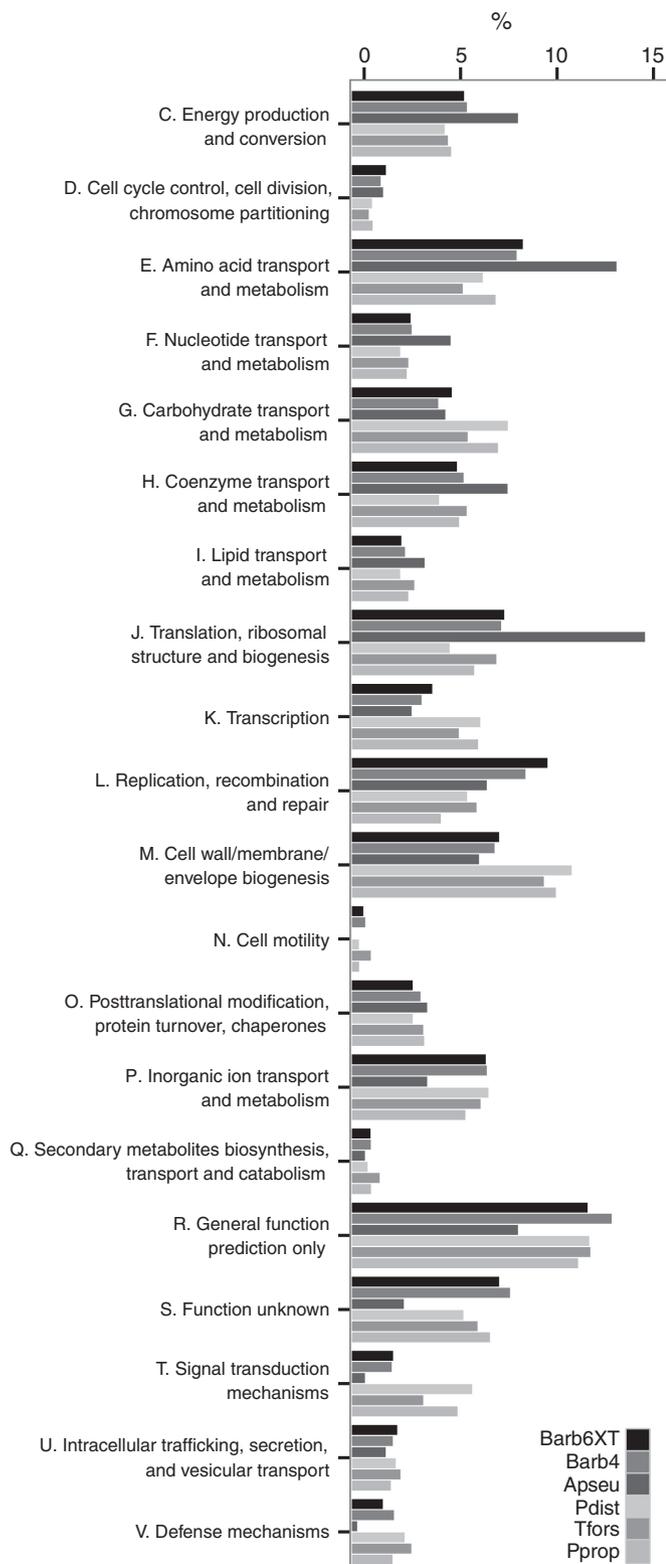


FIG 2 Metabolic pathways predicted from the Barb6XT ectosymbiont. Dotted arrows indicate pathways where an imported or synthesized product is likely to be used. For some functions, Barb6XT carries multiple COGs with similar functions (indicated by a multiplication sign). OMP, outer membrane protein; OMRP, outer membrane receptor protein; FMN, flavin mononucleotide; FAD, flavin adenine dinucleotide; CoA, coenzyme A. The pathways are colored to indicate whether they are found in other *Bacteroidales* that are free-living (FL; *Paludibacter propionigenes* WB4), free-living but host-associated (HA; *Parabacteroides distans* ATCC 8503 and/or *Tannerella forsythia* ATCC 43037), or endosymbiotic (endo; "*Candidatus* Azobacteroides pseudotrichonymphae" CFP2), in addition to Barb6XT (ecto). For descriptions of these bacteria and their genomes, see Table 1.

related RNF complex, typically an NADH:ferredoxin dehydrogenase, were also found in an operon (Fig. 2).

The *Barbulanympha* ectosymbiont genomes contained numerous ORFs shared with free-living species but absent in the endosymbiont "*Candidatus* Azobacteroides pseudotrichonymphae" of the protist *Pseudotrichonympha grassi* from the hindgut of the termite *Coptotermes formosanus*. Most significantly, the ectosymbionts retained genes to synthesize an outer membrane typical of free-living Gram-negative bacteria. The Barb6XT ectosymbiont genome encoded numerous efflux transporters (16 ORFs), outer membrane proteins (15 ORFs), and outer membrane receptor proteins (OMRPs; 58 ORFs), as well as retaining the ability to

synthesize lipopolysaccharides and lipoproteins for the outer membrane, all of which except one efflux transporter were absent from "*Ca. Azobacteroides pseudotrichonymphae*" (Fig. 2). Many of the OMRPs belonged to the SusC/RagA family and were found paired with ORFs of the SusD/RagB family. These proteins are known to bind and import large oligosaccharides (70). Also like in its free-living relatives, multiple sugar transporters were encoded, as well as riboflavin synthesis (Fig. 2). More generally, Barb6XT did not exhibit reductions in inorganic ion transport and metabolism (including outer membrane proteins), cell motility, and defense mechanism genes (particularly efflux proteins) as seen in "*Ca. Azobacteroides pseudotrichonymphae*" (Fig. 3).



**FIG 3** Comparison of COG functional profiles for genomes of *Bacteroidales* related to *Barbulanympha* ectosymbionts (Barb6XT and Barb4). The percentage of ORFs in a COG class is relative to the total number of ORFs classifiable into a COG. Apseu, “*Candidatus Azobacteroides pseudotrichonymphae*” CFP2; Pdist, *Parabacteroides distasonis* ATCC 8503; Tfors, *Tannerella forsythia* ATCC 43037; Pprop, *Paludibacter propionicigenes* WB4.

Barb6XT was reduced in the relative number of COGs relating to transcription, cell wall biosynthesis, and carbohydrate transport and metabolism similarly to “*Ca. Azobacteroides pseudotrichonymphae*” compared to free-living *Bacteroidales*. This genome was also reduced in signal transduction mechanisms but not to the same extent as “*Ca. Azobacteroides pseudotrichonymphae*” (Fig. 3).

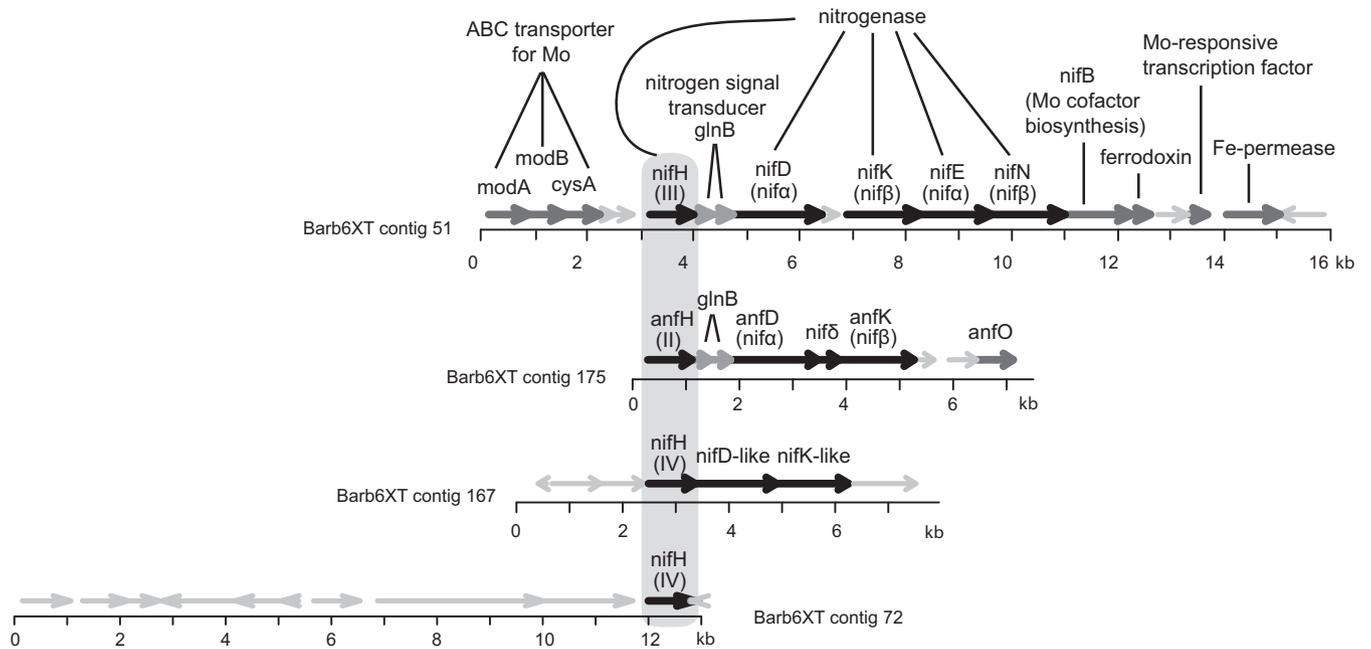
Interestingly, an increase in COGs related to replication, recombination, and repair was observed, primarily due to an increase in the number of transposases (64 ORFs assigned to 6 transposase COGs plus 54 unassigned ORFs annotated as transposases) (Fig. 3; see also Table S4 in the supplemental material). The free-living *Bacteroidales* encoded less than half of the number of transposases, and these were assigned to fewer and to different COG categories (see Table S4 in the supplemental material). Only 1 out of 6 transposase COGs in Barb6XT (COG3547) was also found in the other *Bacteroidales*. No transposases were found in “*Ca. Azobacteroides pseudotrichonymphae*.”

**Nitrogen recycling and fixation are key symbiont functions.**

The genomes of *Barbulanympha* ectosymbionts and the endosymbiont “*Ca. Azobacteroides pseudotrichonymphae*,” both symbionts of parabasalid protists in the hindguts of related wood-eating insects, possess genes involved with nitrogen recycling (i.e., to recover fixed nitrogen, usually ammonia, from waste products such as urea or uric acid). These symbionts are the only *Bacteroidales* known to encode subunits for an ABC-type transporter for urea, as well as a urease that hydrolyzes urea to produce ammonia (Fig. 2). Of the known homologues for the urea transporter and urease, the *Barbulanympha* ectosymbiont proteins are most closely related to those in “*Ca. Azobacteroides pseudotrichonymphae*,” and these have their closest homologues with members of the *Cytophagales* (*Bacteroidetes*) (see Fig. S3 in the supplemental material). Their genomes also contain genes involved in nitrogen fixation, including genes for nitrogenases, nitrogen regulatory proteins, and transporters specific for nitrogenase cofactors that are not typically found in other *Bacteroidales* (Fig. 2 and 4). These bacteria also harbor an ammonia transporter that can be found in some, but not all, *Bacteroidales* (Fig. 2).

The *Barbulanympha* ectosymbiont genomes carried two distinct nitrogenase operons, one for a molybdenum (Mo)-dependent nitrogenase and a second for an alternative iron (Fe)-only nitrogenase (Fig. 4). The Barb6XT Mo-dependent nitrogenase operon included the *nifH*, *nifD*, *nifK*, *nifE*, and *nifN* genes as well as two nitrogen PII signal transducer genes (*glnB*) (Fig. 4). An ABC transporter for molybdenum was located upstream of the nitrogenase genes, and downstream were a molybdenum cofactor biosynthesis gene and a molybdate-dependent transcriptional regulator. The Barb6XT Fe-only nitrogenase operon consisted of *anfH*, *anfD*, and *anfK* (*nifH*, *nifD*, and *nifK* homologues, respectively), two *glnB* ORFs, and a downstream nitrogenase accessory protein (*anfO*) (Fig. 4). The genes for these nitrogenase operons were also found in the Barb4, Barb6, and Barb7 assemblies, but split across multiple contigs.

The amino acid sequences of the NifH protein from these nitrogenase operons were identical, or nearly so, to either cluster II (Fe only) or cluster III-3 (Mo-dependent) NifH previously sequenced from the hindgut of *Cryptocercus punctulatus* (following the nomenclature of Yamada et al. [36]) (Fig. 5). They also grouped with other *Bacteroidales* NifH homologues, including ones from other symbionts of termite hindgut protists and re-



**FIG 4** Genetic structure of Barb6XT contigs carrying *nifH* and *nifH*-like genes (shaded). *anf* nomenclature is used for genes associated with Fe-only nitrogenases. Nitrogenase genes are shown as black arrows pointing in the direction of transcription. Genes for nitrogen regulation are in medium gray, and other genes whose functions support nitrogen fixation are in dark gray. All other genes are in light gray.

cently discovered from cultivated strains but not from *Prevotella bryantii* B<sub>1</sub>4 (37, 66).

The Barb6XT, Barb4, and Barb7 ectosymbiont genomes encoded an additional NifH protein that grouped within a clade of poorly understood NifH relatives (cluster IV) (Fig. 5). In the Barb6XT and Barb7 assemblies, this *nifH* gene was located next to homologues of *nifD* and *nifK* (Fig. 4). A second cluster IV *nifH* gene was found in the Barb6XT, Barb4, and Barb6 ectosymbiont genomes, closely related to the first. Like the first cluster, these NifH-related genes were nearly identical to each other and previously sequenced cluster IV NifH gene from *C. punctulatus* hindguts (Fig. 5).

The cluster IV NifH from the *Barbulanympha* ectosymbionts groups with the only NifH homologue from *Bacteroides reticulotermitis*, isolated from the gut of the subterranean termite *Reticulitermes speratus* (Fig. 5). *B. reticulotermitis* also has genes that encode a NifK-like protein and a truncated NifD-like protein that are located next to its *nifH* gene, and these proteins are closely related to the ones in the *Barbulanympha* ectosymbionts (see Fig. S4 and S5 in the supplemental material). Together, these proteins were distantly related to their homologues in *Endomicrobium proavitum*, an *Elusimicrobia* isolate from the hindguts of *Reticulitermes santonensis* recently shown to be linked to nitrogen fixation (71).

**Evidence for nitrogen fixation from stable isotope labeling and NanoSIMS analysis.** To assess the likelihood that *Barbulanympha* ectosymbionts actively fix nitrogen, live *C. punctulatus* nymphs were housed in a <sup>15</sup>N<sub>2</sub> atmosphere for 2 weeks, after which their hindgut contents were collected by dissection and the hindgut microbes examined using <sup>15</sup>N/<sup>14</sup>N imaging analysis with NanoSIMS. Four *Barbulanympha* cells were identified by SEM for analysis, including two that were damaged during fixation such that the cell interiors were exposed adjacent to attached *Bacteroi-*

*dales* ectosymbionts (Fig. 6). From these two exposed *Barbulanympha* cells, NanoSIMS analysis showed that the ectosymbionts were significantly more enriched in <sup>15</sup>N than their host (2.46 versus 1.47 APE <sup>15</sup>N and 1.73 versus 1.15 APE <sup>15</sup>N, respectively; *P* < 0.001 for both) (Fig. 6D). The ectosymbionts on two of the intact *Barbulanympha* cells were also enriched (4.26 and 4.68 APE <sup>15</sup>N) (Fig. 6D).

For comparison, unidentified free-living bacteria and other protists from the hindgut were also analyzed by NanoSIMS. Approximately 90% of the bacteria had no significant <sup>15</sup>N enrichment. Of the 33 bacteria with significant enrichment, only seven had enrichment levels reaching the range of the analyzed *Barbulanympha* ectosymbionts (Fig. 6D). Of the other protists examined, 75% had no significant <sup>15</sup>N enrichment (6 of 8 cells). Two protist cells were enriched to levels similar to those in *Barbulanympha* but lower than the majority of the ectosymbionts (Fig. 6D).

**Genetic content of the ectosymbionts from each host cell is variable and polymorphic.** The ectosymbiont genomes from each host cell were compared by clustering the ORFs into homologous gene clusters. A large proportion of the gene clusters were unique to an ectosymbiont. Most interestingly, four ORFs (TraG, TraJ, TraO, and TraN) that are part of a *Bacteroides* conjugative transposon were found only from the Barb7 contigs, but the large majority of unique clusters were hypothetical and did not contribute to the metabolic and functional analysis of the ectosymbionts (see Fig. S6 in the supplemental material). Nevertheless, for the unique clusters, a gene belonging to a *Bacteroidetes* was the best hit for 71 to 85% of the hits to GenBank's nonredundant (nr) nucleotide database (see Fig. S6), indicating that potential contamination of contigs from other organisms was low. Approximately one-third of the unique clusters (22 to 39%) were not truly

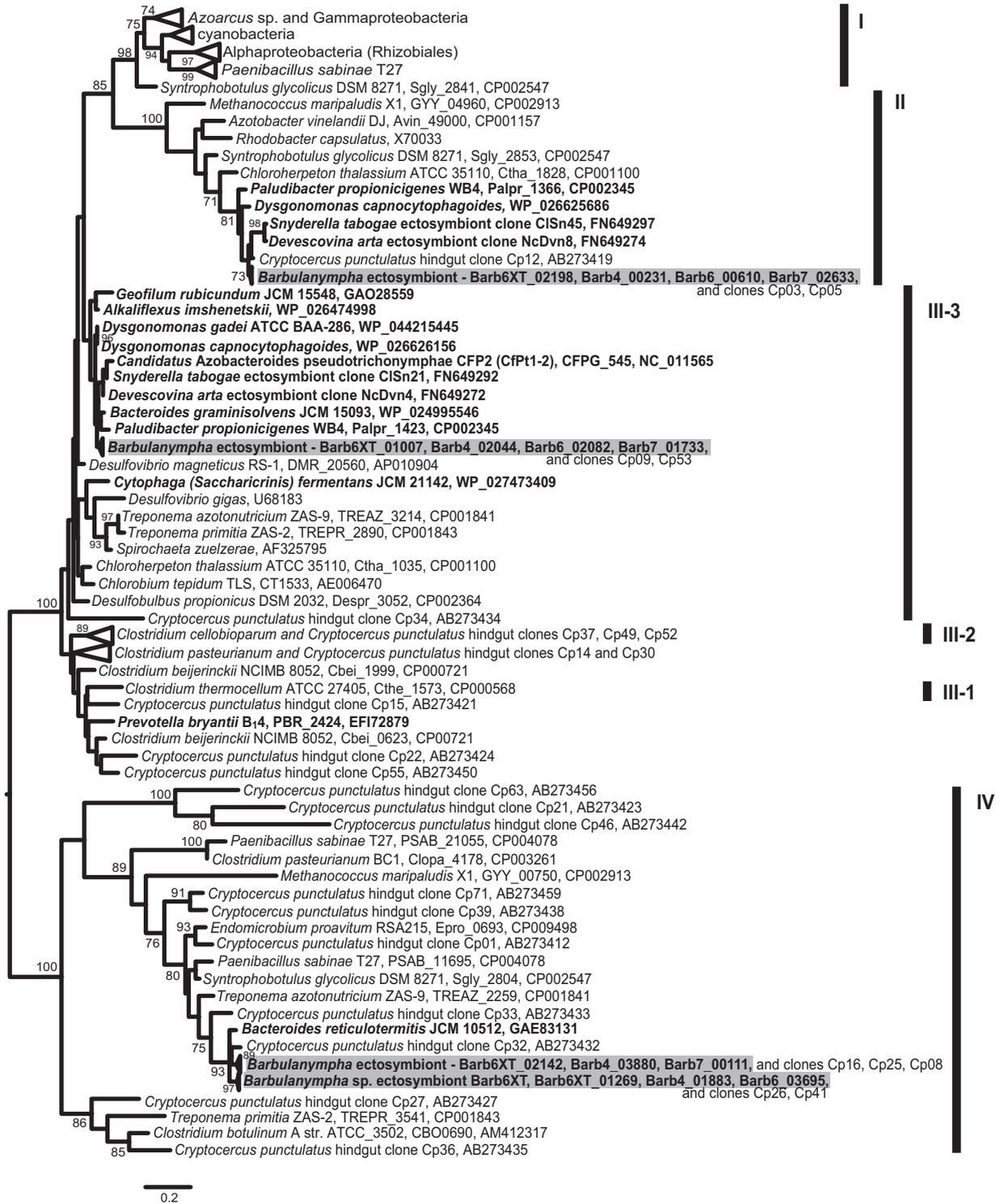
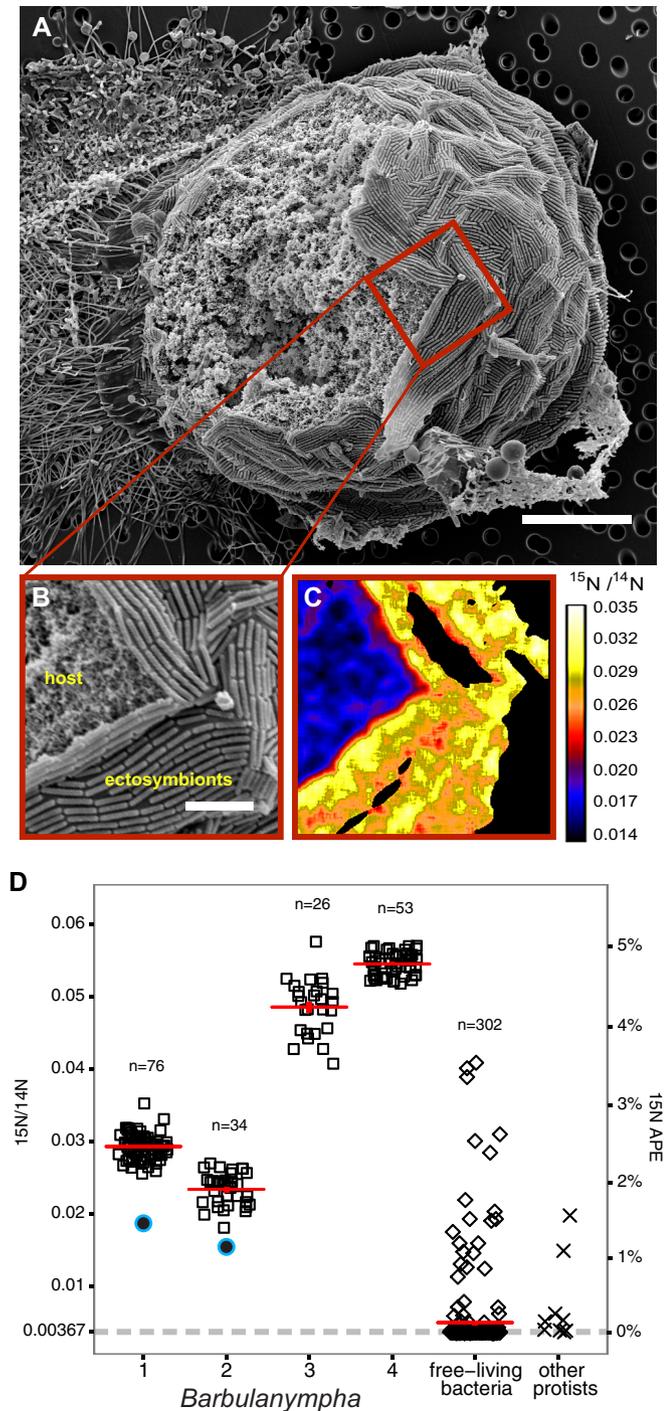


FIG 5 Maximum likelihood tree of NifH amino acid sequences. Environmental clones consisted of all sequences obtained from the hindgut of *Cryptocercus punctulatus* (Cp clones, accession numbers AB273412 to AB273459). Black vertical bars identify recognized clusters of NifH sequences. NifH sequences from *Bacteroidales* are labeled in bold, and those from *Barbulanympcha* ectosymbionts are shaded. Only bootstrap support values greater than 70% from 1,000 replicates are shown.

unique, but homologues in the other genomes were not detected because the ORFs were truncated or misannotated (see Fig. S6).

The ORFs found in common among all four sets of ectosymbiont contigs shared on average 94.9% nucleic acid similarity.

Most of the single nucleotide polymorphisms (SNPs) were unique to the ectosymbionts from a single *Barbulanympcha* cell, indicating that distinct ectosymbiont strains were harbored by each host cell (see Fig. S7 in the supplemental material). The SNPs were largely



**FIG 6** NanoSIMS analyses of nitrogen fixation in *Barbulanympha* ectosymbionts and free-living bacteria. (A) Scanning electron micrograph of a damaged *Barbulanympha* cell that was analyzed by NanoSIMS. Scale bar is 20  $\mu\text{m}$ . (B) Area of cell corresponding to the red box in panel A. Here, the interior of the host cell was exposed and host cytoplasm could be clearly differentiated from the ectosymbionts. Scale bar is 5  $\mu\text{m}$ . (C) NanoSIMS image of the same area as in panel B. Colors reflect  $^{15}\text{N}$  enrichment relative to  $^{14}\text{N}$ . Black areas are where the extracted ion counts were very low (<5% of maximum) due to surface topography and were excluded from analysis. (D) NanoSIMS  $^{15}\text{N}$  enrichment data. Solid circles ringed in blue are the mean  $^{15}\text{N}$  enrichments of ion measurements from the cytoplasm of individual *Barbulanympha* cells, open squares are from individual ectosymbionts on the cell surface, open diamonds are from individual free-living bacteria in the hindgut, and crosses are from protists in the hindgut other than *Barbulanympha*. The number of individual

synonymous and did not result in amino acid changes (data not shown). Gene order also varied on the ectosymbiont contigs, which was confirmed by PCR, suggesting that rearrangement possibly mediated by transposases is also playing a significant role in the evolution of these bacterial genomes (see Fig. S8 and S9 in the supplemental material). In contrast, there were far fewer polymorphisms among the ectosymbiont sequences from a single *Barbulanympha* cell. The sequencing reads from a single *Barbulanympha* cell originated from hundreds of ectosymbionts on its cell surface, but there was little variation in sequence, especially of the ORFs common to all four ectosymbionts (see Fig. S7). The lack of variation in the sequencing reads suggests that the ectosymbionts from a single host cell were mostly clonal.

## DISCUSSION

**The *Barbulanympha* ectosymbiont genome is not reduced.** The *Barbulanympha* ectosymbiont genomes were not reduced in size or gene content, nor were they AT biased, despite the fact that the *Barbulanympha*-*Bacteroidales* symbiosis is considered highly specific, vertically transmitted, and obligate (13, 24, 32). Under these conditions, such genomic features are typical for endosymbionts (19–21), including those of *Bacteroidales* (16), because the symbionts undergo population bottlenecks during vertical transmission resulting in relaxed selection and genetic drift. Not all endosymbionts have reduced genomes (41), but this raises the question of whether the genomes of ectosymbionts evolve in a fundamentally different way.

Much like endosymbionts, the extracellular bacteria associated with the guts of plataspid and acanthosomatid stinkbugs (Plataspidae and Acanthosomatidae) exhibit AT-biased, reduced genomes and accelerated rates of molecular evolution (72, 73). These extracellular symbionts are harbored in specialized gut crypts that are isolated from the main gut tract, similar to the stable intracellular environments of endosymbionts, and their deposition on the surface of eggs results in vertical transmission when the symbionts are probed or ingested by newborn nymphs (10, 73). Thus, extracellular symbionts can evolve similarly to endosymbionts.

In contrast, like the *Barbulanympha* ectosymbionts, extracellular symbionts of lumbricid earthworms (colonizing the lumens of nephridia) also do not have reduced, AT-biased genomes even though the association is vertically transmitted, mutualistic, highly specific, and evolutionarily ancient (~100 million years) (74, 75). These symbionts are deposited in egg capsules and transmitted to the nephridia of developing worms through a recruitment canal (76). Due to this vertical transmission, they have likely undergone population bottlenecks similarly to endosymbionts, as shown by their accelerated evolutionary rates, but genome reduction has not evolved (75, 77). These symbionts also have dynamic genomes where an expansion of mobile elements appears to have mediated genome rearrangements. With the large number of

bacterial cells measured is indicated (n). Red lines indicate the mean  $^{15}\text{N}$  enrichment  $\pm$  the standard error for the ectosymbionts on a single host cell or for free-living bacteria. The standard error line is not visible if it is smaller than the thickness of the mean line. Enrichment data for *Barbulanympha* 3 and 4 were not available because the host cell was intact and the cytoplasm was not accessible for analysis. Measurement uncertainties (standard errors) are not shown. They are smaller than the data points for *Barbulanympha* cells, their ectosymbionts, and other protists and are 1 to 4% for the free-living bacteria.

transposases, variable gene content, and preliminary PCR evidence of rearrangements, we also predict that the *Barbulanympha* ectosymbiont genomes are dynamic.

The lack of genome reduction and active genome rearrangements in the extracellular symbionts of lumbricid earthworms are attributed to their interaction with mixed microbial communities and multiple environments rather than the genetically isolated and stable environments of intracellular symbionts. Diverse and fluctuating environments may select against the loss of genes that are required for survival under these conditions and provide a milieu for genetic exchange and recombination with other microbes, which would counteract the deleterious effects of bottleneck-induced genetic drift (77).

The interaction of the *Barbulanympha* ectosymbionts with the mixed microbial community in the hindgut environment may also explain their lack of genome reduction, even though they are obligate, specific, vertically transmitted symbionts. Both genomic and morphological data indicate that they have retained their cell wall, outer membrane, and outer membrane proteins, effectively maintaining a boundary for mediating interactions with the hindgut environment (13) (Fig. 2 and 3). Efflux transporters were also retained providing a defensive capacity for living in this mixed community (Fig. 2 and 3). Hence, the extracellular symbiotic lifestyle itself with population bottlenecks does not necessarily result in converging evolutionary trajectories with endosymbionts. Rather, the extent of genetic isolation and the stability of a symbiont's environment likely play major roles in the evolution and genetic content of their genomes.

#### Specificity of the *Barbulanympha*-*Bacteroidales* symbiosis.

The selective pressures faced by *Barbulanympha* ectosymbionts and the evolutionary outcomes observed in their genome structure are most likely the result of their niche as extracellular symbionts attached to their protist host and not due to a free-living stage within the insect gut. *Barbulanympha* is usually covered in rod-shaped ectosymbionts, and morphologically identical bacteria (possibly the same bacteria as the ectosymbionts) have also been observed intracellularly in vesicles but are much less abundant (13, 24). The ectosymbionts appear to be attached to *Barbulanympha* cells through specialized, electron-dense structures in the host cytoplasm and a thickened glycocalyx (13, 24). The proteins responsible for this interaction are not known but could include the numerous outer membrane proteins, receptors, and a possible adhesin protein in the ectosymbiont. The density of ectosymbionts on *Barbulanympha* cells varies, possibly decreasing during *Barbulanympha* cell division and encystment (13), but *Barbulanympha* cells without ectosymbionts have not been observed. Repopulation of the *Barbulanympha* cell surface following division or emergence from cysts has been hypothesized to come from symbionts that have remained on the cell through its life stages or from symbionts that were retained in vesicles intracellularly (13). These ultrastructural data suggest that the symbiosis is obligate, and the symbionts are vertically transmitted during *Barbulanympha* cell division and not acquired from the hindgut environment, although the possibility of a free-living stage cannot be completely rejected. *Barbulanympha* itself is vertically transmitted from adult insects to nymphs together with other hindgut symbionts through proctodeal trophallaxis (the consumption of hindgut fluids directly from the rectal pouch of the donor) (78, 79).

Protist-*Bacteroidales* symbioses have evolved many times in the hindguts of lower termites and *Cryptocercus* cockroaches, and

different protist lineages independently acquired symbionts from the pool of bacteria in the hindgut (32). But once acquired, the protist host and symbiont appear to have cospeciated in many cases, indicating the high degree of specificity between hindgut protists and their *Bacteroidales* symbionts (29, 31). As demonstrated by the near identity of 16S rRNA sequences from *Barbulanympha* isolated from different collections of *C. punctulatus* cockroaches (Fig. 1), *Barbulanympha* consistently associates with a specific lineage of *Bacteroidales*. Molecular data were not collected from the cells analyzed by NanoSIMS, but there is little doubt these symbioses are consistent with those investigated for genomics.

**Bacteroidales as nitrogen fixers and recyclers.** Nitrogen fixation has been hypothesized from molecular and genomic data to be a key process in termite gut symbioses (16, 35–37, 41). Our NanoSIMS analysis of *in vivo*  $^{15}\text{N}$ -labeled hindgut microbes provides direct imaging evidence of a specific hindgut microbe playing a role in nitrogen fixation in its natural environment and represents a significant advance for these uncultivated microbes. The ectosymbionts were significantly more enriched in  $^{15}\text{N}$  than their host *Barbulanympha* and were much more enriched than the majority of free-living bacteria in the hindgut (Fig. 6). *Barbulanympha* is likely enriched in  $^{15}\text{N}$  because it receives fixed nitrogen products from its ectosymbionts (likely in the form of ammonia, but this is not known). Other protists in the hindgut are not similarly enriched, but those that are might also harbor nitrogen-fixing symbionts. Approximately 2% of the free-living bacteria were as enriched as the *Barbulanympha* ectosymbionts, providing evidence that free-living bacteria also fix nitrogen in the *C. punctulatus* hindgut. However, most of the bacteria (~80%) were not significantly enriched (Fig. 6).

*Bacteroidales* are often associated within the guts of animals and are typically known as degraders of polymeric carbohydrates (80), but recent genome sequencing has revealed the potential for nitrogen fixation in several members (66). It is also becoming increasingly clear that most *Bacteroidales* endo- and ectosymbionts of protists in the hindguts of termites and cockroaches are likely diazotrophs (16, 37). The hindgut of *Cryptocercus punctulatus* harbors microorganisms with a diversity of *nifH* genes (36), and we show here that four of these are from the ectosymbionts of *Barbulanympha* encoding an Fe-only nitrogenase subunit (cluster II), a Mo-dependent nitrogenase subunit (cluster III-3), and 2 cluster IV NifH proteins (Fig. 4).

The *Barbulanympha* ectosymbiont is the only known *Bacteroidales* genome with homologues from three different NifH clusters. The presence of both Fe-only and Mo-dependent nitrogenases is thought to maintain nitrogen fixation depending on the availability of Mo cofactors. However, in termite guts and in *Bacteroidales* symbionts encoding homologues to both Fe-only and Mo-dependent NifH, Fe-only NifH (AnfH) was preferentially expressed despite providing additional molybdenum (35, 37). Cluster IV NifH was not thought to contribute to nitrogen fixation, but *Endomicrobium proavitum* (belonging to the *Endomicrobia* and a free-living termite hindgut microbe) fixes nitrogen while only encoding a cluster IV NifH, and also NifD-like and NifK-like proteins (71). *Barbulanympha* ectosymbionts have multiple, diverse nitrogenase operons, suggesting that the ectosymbionts have several means to achieve nitrogen fixation. The NanoSIMS analysis demonstrates that the ectosymbionts do fix nitrogen, but the specific nitrogenase genes that are responsible for this function and the

expression and regulation of these genes have not been determined.

*Bacteroidales* with nitrogen fixation genes belong to different lineages based on 16S rRNA gene phylogeny, but their NifH proteins cluster together (Fig. 4 and 5), suggesting that their nitrogenase genes are ancestral and have generally evolved vertically but were lost in many *Bacteroidales* (66). The diversity of *nifH* homologues and associated nitrogenase genes in the *Barbulanympha* ectosymbiont suggests that a common ancestor possessed a diversity of nitrogenases that were subsequently lost or transferred among the *Bacteroidales*. Alternatively, the nitrogenases may have been independently acquired in different *Bacteroidales* ancestors that were transferred among the *Bacteroidales*, with 3 types accumulating in the *Barbulanympha* ectosymbiont lineage. The two cluster IV *nifH* genes in the *Barbulanympha* ectosymbionts likely arose through duplication. Notably, the NifH from *Prevotella bryantii* B<sub>1</sub>4 did not cluster with the majority of the *Bacteroidales* NifH proteins, suggesting that *P. bryantii* B<sub>1</sub>4 *nifH* has a different origin and that the occurrence of *nifH* genes in the *Bacteroidales* is not entirely due to common ancestry. Deviations from a strict vertical evolution of nitrogenase genes in the *Bacteroidales* have also been noted previously (66).

By providing bioavailable nitrogen, nitrogen fixation in the *Bacteroidales*, and perhaps also urea transport and metabolism, has likely been key to their evolution as protist symbionts in the hindguts of wood-eating termites and cockroaches. Low nitrogen availability is the most plausible driving force behind these protist-*Bacteroidales* symbioses because wood, which is consumed by many of the large protists in the hindguts, is nitrogen poor. The prevalence of nitrogen fixation by bacterial symbionts of protists supports this hypothesis, but the amount of nitrogen supplied to protists by their symbionts, the mechanism, the type of molecule supplied, and the dependency of the protists on this nitrogen supply are not known. Free-living hindgut bacteria probably also fix nitrogen (38–40), but most of the nitrogen fixation activity in the hindguts is likely performed by bacterial symbionts of protists (37, 41), a conclusion supported by our analysis of free-living bacteria in the hindguts of *Cryptocercus* (Fig. 6D). Consequently, these protist-bacterium symbioses are key to the hindgut ecosystem, which functions to digest wood and supply nutrition to their insect hosts.

**Conclusions.** The combination of culture-independent genomics and NanoSIMS analyses provides a powerful set of complementary tools for investigating the biology of uncultivable microbes. We used these tools to study the *Bacteroidales* ectosymbionts of *Barbulanympha*, providing new insights into the functional interactions and evolution of ectosymbiosis. Like *Bacteroidales* that are protist endosymbionts, *Barbulanympha* ectosymbionts are diazotrophs and probably also recycle nitrogenous compounds. The ecological pressures faced by these ectosymbionts, however, clearly differ from those for endosymbionts, which reside intracellularly, resulting in divergent evolutionary outcomes. The dynamic environment experienced by ectosymbionts is in sharp contrast to the stable intracellular environment of endosymbionts, which we suggest favors gene retention and genome recombination. However, the ecological and evolutionary conditions that result in intracellular symbiosis as opposed to extracellular symbiosis remain unclear. It is tempting to speculate that the *Barbulanympha* ectosymbionts might eventually evolve into endosymbionts, as may have occurred with the spirochete endosym-

bionts of the termite gut-dwelling protist *Eucomonympha* (41), but the ectosymbionts are clearly adapted to interact with a diverse extracellular environment, and their genomes, though not reduced, are in a process of dynamic evolution.

## ACKNOWLEDGMENTS

We thank Erick James (University of British Columbia) for assistance with isolating cells for genomic analysis, M. Lee Davisson (Lawrence Livermore National Laboratory [LLNL]) for assistance with <sup>15</sup>N stable isotopic labeling, Christina Ramon (LLNL) for imaging assistance, Patrik D'Haeseleer (LLNL) for helpful comments on the manuscript, Jean-François Pombert (Illinois Institute of Technology) for bioinformatics support, and Moriya Ohkuma (RIKEN Bioresource Center, Wako-Saitama) for helpful discussions about termite gut microbial ecology. Work at LLNL was performed under the auspices of the U.S. Department of Energy (DOE) under contract DE-AC52-07NA27344.

## FUNDING INFORMATION

This work was supported by a grant (227301) to Patrick Keeling from the Natural Sciences and Engineering Research Council of Canada (NSERC). NanoSIMS work was supported by an LLNL Laboratory Directed Research and Development grant (011-LW-039) to Kevin Carpenter. Peter Weber at LLNL was funded in part by DOE Genome Sciences Program grant SCW1039. Vera Tai was supported as a Global Scholar with the Canadian Institute for Advanced Research (CIFAR) and through a Postdoctoral Fellowship from NSERC. CIFAR supports Patrick Keeling as a Senior Fellow and Steve Perlman as a Fellow of the Integrated Microbial Biodiversity Program.

## REFERENCES

- Dale C, Moran NA. 2006. Molecular interactions between bacterial symbionts and their hosts. *Cell* 126:453–465. <http://dx.doi.org/10.1016/j.cell.2006.07.014>.
- Dubilier N, Bergin C, Lott C. 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Microbiol* 6:725–740. <http://dx.doi.org/10.1038/nrmicro1992>.
- Nowack ECM, Melkonian M. 2010. Endosymbiotic associations within protists. *Philos Trans R Soc Lond B Biol Sci* 365:699–712. <http://dx.doi.org/10.1098/rstb.2009.0188>.
- Yang T, Chen Y, Wang X-X, Dai C-C. 2012. Plant symbionts: keys to the phytosphere. *Symbiosis* 59:1–14.
- Anderson OR. 2014. Living together in the plankton: a survey of marine protist symbioses. *Acta Protozool* 53:29–38.
- Gast RJ, Sanders RW, Caron DA. 2009. Ecological strategies of protists and their symbiotic relationships with prokaryotic microbes. *Trends Microbiol* 17:563–569. <http://dx.doi.org/10.1016/j.tim.2009.09.001>.
- Feldhaar H. 2011. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol Entomol* 36:533–543. <http://dx.doi.org/10.1111/j.1365-2311.2011.01318.x>.
- Hansen AK, Moran NA. 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol Ecol* 23:1473–1496.
- McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* 110:3229–3236. <http://dx.doi.org/10.1073/pnas.1218525110>.
- Fukatsu T, Hosokawa T. 2002. Capsule-transmitted gut symbiotic bacterium of the Japanese common plataspid stinkbug, *Megacopta punctatissima*. *Appl Environ Microbiol* 68:389–396. <http://dx.doi.org/10.1128/AEM.68.1.389-396.2002>.
- Xu J, Mahowald MA, Ley RE, Lozupone CA, Hamady M, Martens EC, Henrissat B, Coutinho PM, Minx P, Latreille P, Cordum H, Van Brunt A, Kim K, Fulton RS, Fulton LA, Clifton SW, Wilson RK, Knight RD, Gordon JL. 2007. Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol* 5:e156. <http://dx.doi.org/10.1371/journal.pbio.0050156>.
- Strassert JFH, Desai MS, Radek R, Brune A. 2010. Identification and localization of the multiple bacterial symbionts of the termite gut flagellate *Joenia annectens*. *Microbiology* 156:2068–2079. <http://dx.doi.org/10.1099/mic.0.037267-0>.
- Carpenter KJ, Horak A, Chow L, Keeling PJ. 2011. Symbiosis, morphol-

- ogy, and phylogeny of Hoplonymphidae (Parabasalia) of the wood-feeding roach *Cryptocercus punctulatus*. *J Eukaryot Microbiol* 58:426–436. <http://dx.doi.org/10.1111/j.1550-7408.2011.00564.x>.
14. Altura MA, Heath-Heckman EAC, Gillette A, Kremer N, Krachler A-M, Brennan C, Ruby EG, Orth K, McFall-Ngai MJ. 2013. The first engagement of partners in the *Euprymna scolopes-Vibrio fischeri* symbiosis is a two-step process initiated by a few environmental symbiont cells. *Environ Microbiol* 15:2937–2950.
  15. 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. <http://dx.doi.org/10.1073/pnas.0801389105>.
  16. Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Genome of an endosymbiotic coupling N<sub>2</sub> fixation to cellulolysis within protist cells in termite gut. *Science* 322:1108–1109. <http://dx.doi.org/10.1126/science.1165578>.
  17. Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M. 2010. The genome of the amoeba symbiont “*Candidatus Amoebophilus asiaticus*” reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* 192:1045–1057. <http://dx.doi.org/10.1128/JB.01379-09>.
  18. Boscaro V, Felletti M, Vannini C, Ackerman MS, Chain PS, Malfatti S, Vergez LM, Shin M, Doak TG, Lynch M. 2013. *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. *Proc Natl Acad Sci U S A* 110:18590–18595. <http://dx.doi.org/10.1073/pnas.1316687110>.
  19. Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, Matsuyama K, Takishita K, Kawato M, Uematsu K, Fujiwara Y, Sato T, Kato C, Kitagawa M, Kato I, Maruyama T. 2007. Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* 17:881–886. <http://dx.doi.org/10.1016/j.cub.2007.04.039>.
  20. Toft C, Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* 11:465–475. <http://dx.doi.org/10.1038/nrg2798>.
  21. McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26.
  22. Moran NA. 1996. Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* 93:2873–2878. <http://dx.doi.org/10.1073/pnas.93.7.2873>.
  23. Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet* 3:850–861. <http://dx.doi.org/10.1038/nrg931>.
  24. Bloodgood RA, Fitzharris TP. 1976. Specific associations of prokaryotes with symbiotic flagellate protozoa from the hindgut of the termite *Reticulitermes* and the wood-eating roach *Cryptocercus*. *Cytobios* 17:103–122.
  25. Yubuki N, Edgcomb VP, Bernhard JM, Leander BS. 2009. Ultrastructure and molecular phylogeny of *Calkinsia aureus*: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria. *BMC Microbiol* 9:16. <http://dx.doi.org/10.1186/1471-2180-9-16>.
  26. Orsi W, Charvet S, Vd’acný P, Bernhard JM, Edgcomb VP. 2012. Prevalence of partnerships between bacteria and ciliates in oxygen-depleted marine water columns. *Front Microbiol* 3:341.
  27. Ohkuma M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends Microbiol* 16:345–352. <http://dx.doi.org/10.1016/j.tim.2008.04.004>.
  28. Brune A, Ohkuma M. 2011. Role of the termite gut microbiota in symbiotic digestion, p 439–475. *In* Bignell D, Roisin Y, Lo N (ed), *Biology of termites: a modern synthesis*. Springer, New York, NY.
  29. Noda S, Kitade O, Inoue T, Kawai M, Kanuka M, Hiroshima K, Hongoh Y, Constantino R, Uys V, Zhong J, Kudo T, Ohkuma M. 2007. Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotrichonympha* spp.), their hosts, and their bacterial endosymbionts. *Mol Ecol* 16:1257–1266. <http://dx.doi.org/10.1111/j.1365-294X.2006.03219.x>.
  30. 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.
  31. Desai MS, Strassert JFH, Meuser K, Hertel H, Ikeda-Ohtsubo W, Radek R, Brune A. 2010. Strict cospeciation of descovinid flagellates and Bacteroidales ectosymbionts in the gut of dry-wood termites (Kalotermitidae). *Environ Microbiol* 12:2120–2132.
  32. Noda S, Hongoh Y, Sato T, Ohkuma M. 2009. Complex coevolutionary history of symbiotic Bacteroidales bacteria of various protists in the gut of termites. *BMC Evol Biol* 9:158. <http://dx.doi.org/10.1186/1471-2148-9-158>.
  33. Ohkuma M, Noda S, Usami R, Horikoshi K, Kudo T. 1996. Diversity of nitrogen fixation genes in the symbiotic intestinal microflora of the termite *Reticulitermes speratus*. *Appl Environ Microbiol* 62:2747–2752.
  34. Ohkuma M, Noda S, Kudo T. 1999. Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Appl Environ Microbiol* 65:4926–4934.
  35. Noda S, Ohkuma M, Usami R, Horikoshi K, Kudo T. 1999. Culture-independent characterization of a gene responsible for nitrogen fixation in the symbiotic microbial community in the gut of the termite *Neotermes koshunensis*. *Appl Environ Microbiol* 65:4935–4942.
  36. Yamada A, Inoue T, Noda S, Hongoh Y, Ohkuma M. 2007. Evolutionary trend of phylogenetic diversity of nitrogen fixation genes in the gut community of wood-feeding termites. *Mol Ecol* 16:3768–3777. <http://dx.doi.org/10.1111/j.1365-294X.2007.03326.x>.
  37. Desai MS, Brune A. 2012. Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. *ISME J* 6:1302–1313. <http://dx.doi.org/10.1038/ismej.2011.194>.
  38. French JR, Turner GL, Bradbury JF. 1976. Nitrogen fixation by bacteria from the hindgut of termites. *J Gen Microbiol* 96:202–206.
  39. Potriuk CJ, Breznak JA. 1977. Nitrogen-fixing *Enterobacter agglomerans* isolated from guts of wood-eating termites. *Appl Environ Microbiol* 33:392–399.
  40. Lilburn TG, Kim KS, Ostrom NE, Byzek KR, Leadbetter JR, Breznak JA. 2001. Nitrogen fixation by symbiotic and free-living spirochetes. *Science* 292:2495–2498. <http://dx.doi.org/10.1126/science.1060281>.
  41. Ohkuma M, Noda S, Hattori S, Iida T, Yuki M, Starns D, Inoue J-I, Darby AC, Hongoh Y. 2015. Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *Proc Natl Acad Sci U S A* 112:10224–10230. <http://dx.doi.org/10.1073/pnas.1423979112>.
  42. Lechene CP, Luyten Y, McMahon G, Distel DL. 2007. Quantitative imaging of nitrogen fixation by individual bacteria within animal cells. *Science* 317:1563–1566. <http://dx.doi.org/10.1126/science.1145557>.
  43. Carpenter KJ, Weber PK, Davisson ML, Pett-Ridge J, Haverty MI, Keeling PJ. 2013. Correlated SEM, FIB-SEM, TEM, and NanoSIMS imaging of microbes from the hindgut of a lower termite: methods for in situ functional and ecological studies of uncultivable microbes. *Microsc Microanal* 19:1490–1501. <http://dx.doi.org/10.1017/S1431927613013482>.
  44. Cleveland LR, Hall SR, Sanders EP, Collier J. 1934. The wood-feeding roach *Cryptocercus*, its protozoa, and the symbiosis between protozoa and roach. *Mem Am Acad Arts Sci* 17:185–342.
  45. Trager W. 1934. The cultivation of a cellulose-digesting flagellate, *Trichomonas termopsidis*, and of certain other termite protozoa. *Biol Bull* 66:182–190. <http://dx.doi.org/10.2307/1537331>.
  46. Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). GitHub, San Francisco, CA.
  47. Boisvert S, Laviolette F, Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol* 17:1519–1533. <http://dx.doi.org/10.1089/cmb.2009.0238>.
  48. Simpson JT, Wong K, Jackman SD, Schein JE, Jones S, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res* 19:1117–1123. <http://dx.doi.org/10.1101/gr.089532.108>.
  49. Chevreur B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics*.
  50. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <http://dx.doi.org/10.1093/bioinformatics/btt086>.
  51. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie2. *Nat Methods* 9:357–359. <http://dx.doi.org/10.1038/nmeth.1923>.
  52. Huang Y, Gilna P, Li W. 2009. Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* 25:1338–1340. <http://dx.doi.org/10.1093/bioinformatics/btp161>.
  53. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. 2011. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* 12:R44. <http://dx.doi.org/10.1186/gb-2011-12-5-r44>.
  54. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margosian H, Coronado S, van der Maaten L, Vlassis N, Wilmes P. 2015. VizBin—an application for reference-independent visualization and hu-

- man-augmented binning of metagenomic data. *Microbiome* 3:1. <http://dx.doi.org/10.1186/s40168-014-0066-1>.
55. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>.
  56. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAA5: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185. <http://dx.doi.org/10.1093/nar/gkm321>.
  57. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499: 431–437. <http://dx.doi.org/10.1038/nature12352>.
  58. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <http://dx.doi.org/10.1371/journal.pcbi.1002195>.
  59. Contreras-Moreira B, Vinuesa P. 2013. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 79:7696–7701. <http://dx.doi.org/10.1128/AEM.02411-13>.
  60. DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernysky A, Sivachenko A, Cibulskis K, Gabriel S, Altshuler D, Daly M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498. <http://dx.doi.org/10.1038/ng.806>.
  61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics* 25:2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>.
  62. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3. *Fly* 6:80–92. <http://dx.doi.org/10.4161/fly.19695>.
  63. Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286–298. <http://dx.doi.org/10.1093/bib/bbn013>.
  64. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334>.
  65. Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <http://dx.doi.org/10.1093/bioinformatics/btl446>.
  66. Inoue J-I, Oshima K, Suda W, Sakamoto M, Iino T, Noda S, Hongoh Y, Hattori M, Ohkuma M. 2015. Distribution and evolution of nitrogen fixation genes in the phylum *Bacteroidetes*. *Microbes Environ* 30:44–50. <http://dx.doi.org/10.1264/jisme2.ME14142>.
  67. Popa R, Weber PK, Pett-Ridge J, Finzi JA, Fallon SJ, Hutcheon ID, Neelson KH, Capone DG. 2007. Carbon and nitrogen fixation and metabolite exchange in and between individual cells of *Anabaena oscillarioides*. *ISME J* 1:354–360.
  68. Gronow S, Munk C, Lapidus A, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng J-F, Tapia R, Han C, Goodwin L, Pitluck S, Liolios K, Ivanova N, Mavromatis K, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Chang Y-J, Jeffries CD, Brambilla E, Rohde M, Göker M, Detter JC, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk H-P. 2011. Complete genome sequence of *Paludibacter propionicipigenes* type strain (WB4T). *Stand Genomic Sci* 4:36–44. <http://dx.doi.org/10.4056/signs.1503846>.
  69. Reyes-Prieto A, Barquera B, Juárez O. 2014. Origin and evolution of the sodium-pumping NADH: ubiquinone oxidoreductase. *PLoS One* 9:e96696. <http://dx.doi.org/10.1371/journal.pone.0096696>.
  70. Shipman JA, Berleman JE, Salyers AA. 2000. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J Bacteriol* 182:5365–5372. <http://dx.doi.org/10.1128/JB.182.19.5365-5372.2000>.
  71. 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. <http://dx.doi.org/10.1111/1462-2920.12960>.
  72. Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. 2006. Strict host-symbiont speciation and reductive genome evolution in insect gut bacteria. *PLoS Biol* 4:e337. <http://dx.doi.org/10.1371/journal.pbio.0040337>.
  73. Kikuchi Y, Hosokawa T, Nikoh N, Meng X-Y, Kamagata Y, Fukatsu T. 2009. Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biol* 7:2. <http://dx.doi.org/10.1186/1741-7007-7-2>.
  74. Lund MB, Davidson SK, Holmstrup M, James S, Kjeldsen KU, Stahl DA, Schramm A. 2010. Diversity and host specificity of the *Verminephrobacter*-earthworm symbiosis. *Environ Microbiol* 12:2142–2151.
  75. Kjeldsen KU, Bataillon T, Pinel N, De Mita S, Lund MB, Panitz F, Bendixen C, Stahl DA, Schramm A. 2012. Purifying selection and molecular adaptation in the genome of *Verminephrobacter*, the heritable symbiotic bacteria of earthworms. *Genome Biol Evol* 4:307–315. <http://dx.doi.org/10.1093/gbe/evs014>.
  76. Davidson SK, Stahl DA. 2008. Selective recruitment of bacteria during embryogenesis of an earthworm. *ISME J* 2:510–518. <http://dx.doi.org/10.1038/ismej.2008.16>.
  77. Lund MB, Kjeldsen KU, Schramm A. 2014. The earthworm-*Verminephrobacter* symbiosis: an emerging experimental system to study extracellular symbiosis. *Front Microbiol* 5:128.
  78. Nalepa CA, Bignell DE, Bandi C. 2001. Detritivory, coprophagy, and the evolution of digestive mutualisms in Dictyoptera. *Insectes Soc* 48:194–201. <http://dx.doi.org/10.1007/PL00001767>.
  79. Nalepa CA. 2015. Origin of termite eusociality: trophallaxis integrates the social, nutritional, and microbial environments. *Ecol Entomol* 40:323–335. <http://dx.doi.org/10.1111/een.12197>.
  80. Thomas F, Hehemann J-H, Rebuffet E, Czejek M, Michel G. 2011. Environmental and gut bacteroidetes: the food connection. *Front Microbiol* 2:93.
  81. Noda S, Inoue T, Hongoh Y, Kawai M, Nalepa CA, Vongkaluang C, Kudo T, Ohkuma M. 2006. Identification and characterization of ectosymbionts of distinct lineages in Bacteroidales attached to flagellated protists in the gut of termites and a wood-feeding cockroach. *Environ Microbiol* 8:11–20. <http://dx.doi.org/10.1111/j.1462-2920.2005.00860.x>.