Revealing the metabolic capacity of *Streblomastix strix* and its bacterial symbionts using single-cell metagenomics

Sebastian C. Treitli[a], Martin Kolisko[b], Filip Husník[c], Patrick J. Keeling[c], and Vladimír Hampl[a,1]

[a]Department of Parasitology, Faculty of Science, Charles University, BIOCEV, 252 42 Vestec, Czech Republic; [b]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 370 05 České Budějovice, Czech Republic; and [c]Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Edited by Nancy A. Moran, University of Texas at Austin, Austin, TX, and approved August 14, 2019 (received for review June 26, 2019)

Lower termites harbor in their hindgut complex microbial communities that are involved in the digestion of cellulose. Among these are protists, which are usually associated with specific bacterial symbionts found on their surface or inside their cells. While these form the foundations of a classic system in symbiosis research, we still know little about the functional basis for most of these relationships. Here, we describe the complex functional relationship between one protist, the oxymonad *Streblomastix strix*, and its ectosymbiotic bacterial community using single-cell genomics. We generated partial assemblies of the host *S. strix* genome and *Candidatus* Ordinivivax streblomastigis, as well as a complex metagenome assembly of at least 8 other Bacteroidetes bacteria confirmed by ribosomal (r)RNA fluorescence in situ hybridization (FISH) to be associated with *S. strix*. Our data suggest that *S. strix* is probably not involved in the cellulose digestion, but the bacterial community on its surface secretes a complex array of glycosyl hydrolases, providing them with the ability to degrade cellulose to monomers and fueling the metabolism of *S. strix*. In addition, some of the bacteria can fix nitrogen and can theoretically provide *S. strix* with essential amino acids and cofactors, which the protist cannot synthesize. On the contrary, most of the bacterial symbionts lack the essential glycolytic enzyme enolase, which may be overcome by the exchange of intermediates with *S. strix*. This study demonstrates the value of the combined single-cell (meta)genomic and FISH approach for studies of complicated symbiotic systems.

oxymonads | *Streblomastix* | termite | ectosymbionts | Bacteroidetes

All lower termites harbor complex hindgut communities consisting of protists, bacteria, and archaea (1, 2), which are collectively involved in the digestion of cellulose (3). Most of the protists present in the gut of termites belong to parabasalids or oxymonads, and many are in close association with various bacterial or archaeal symbionts. This complex animal–protist–bacteria–archaea community is a classic example of symbioses, but we lack much detail on the functional interactions between the partners, even after over a century of study. Genomic analyses are beginning to reveal the roles of symbiotic bacteria, which have retained metabolic pathways involved in the synthesis of vitamins, amino acids, and nucleotides, as well as the ability to fix nitrogen (4, 5), suggesting that they provide critical nutrients to the protist and animal hosts. However, much of this information comes from a taxonomically narrow group of hosts, mainly species of *Trichonympha* (4–7), *Pseudotrichonympha* (8), *Barbulanympha* (9), and *Eucomonympha* (10), all members of the same parabasalian class. Few genomes of bacterial symbionts of oxymonads are available, namely the genome of *Candidatus* Symbiothrix dinenymphae, a symbiont of *Dinenympha* (11), and several partial genomes of *Candidatus* Termititenax symbionts of spirochetes attached to the cells of *Pyrsonympha*, *Dinenympha leidyi*, and *Oxymonas* (12). However, in the case of *Ca.* Termititenax, only the symbiont genomes associated with *Oxymonas* are close to be complete; the other genomes are less than 10% characterized (12).

While we are beginning to learn potential functions of the bacteria, the role of flagellate hosts remains essentially unknown. Besides partial ribosomal (r)RNA genes used for diversity studies, there are almost no molecular or biochemical data available for any of these protists. Early studies showed that *Trichomitopsis termopsidis* and several *Trichonympha* species, from the hindgut of *Zootermopsis* (13–16), have the capacity to degrade cellulose (17, 18), but for other flagellates, including all oxymonads, the involvement in cellulose digestion remains unclear. Recent results have shown that the bacterial symbionts of oxymonads have the ability to degrade cellulose (11), but there is still a critical lack in understanding the role of oxymonads themselves.

Streblomastix strix is an oxymonad that lives in the gut of 3 species of *Zootermopsis* (14, 19, 20), but some representatives of the genus *Streblomastix* were also identified in the gut of the termites from genus *Archotermopsis* (21). The flagellate has a stellar shape (19), and it is completely covered with epibiotic bacteria. Previous studies have shown that there are at least 3 different phylotypes of Bacteroidetes bacteria attached to the surface of *S. strix* (21),

Significance

Lower termites digest wood with the help of a network of interacting microbes, but how each individual microbe contributes to this process is poorly understood. We reconstructed the energy metabolism of a complex superorganism, consisting of a eukaryotic protist whose surface is coated by at least 8 distinct kinds of bacteria. The bacteria produce enzymes that degrade cellulose from wood fibers into simple sugars, which are used by both the bacteria and their protist host for energy. The protist may also acquire nitrogenous compounds, amino acids, and vitamins by engulfing and digesting its surface symbionts. We speculate that the protist controls the surface bacteria by complementing one missing enzyme of their core energy metabolism.

Author contributions: V.H. designed research; S.C.T. and F.H. performed research; P.J.K. and V.H. contributed new reagents/analytic tools; S.C.T. and M.K. analyzed data; and S.C.T., P.J.K., and V.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence data have been deposited in GenBank, https://www.ncbi.nlm.nih.gov/genbank, databases under National Center for Biotechnology Information (NCBI) BioProject PRJNA524138. The raw reads of the entire Whole Genome Shotgun project have been deposited in the Sequence Read Archive (SRA), https://www.ncbi.nlm.nih.gov/sra (accession nos. SRR8661006–SRR8661008). The raw reads for the single-cell transcriptomes of *S. strix* have been deposited in the SRA (accession nos. SRR9265627–SRR9265632). The scaffolds assigned to the draft genome of "*Ca.* Ordinivivax streblomastigis" draft genome of *S. strix* and miscellaneous Bacteroidales metagenome have been deposited in the NCBI genome database, https://www.ncbi.nlm.nih.gov/genome (accession nos. SNRX00000000, SNRW00000000, and SNRY00000000, respectively). The amplified 16S rRNA sequences have been deposited in the GenBank database (accession nos. MK585202–MK585215).

[1]To whom correspondence may be addressed. Email: vlada@natur.cuni.cz.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1910793116/-/DCSupplemental.

First published September 6, 2019.

which is in accordance with other studies that identified 3 different morphotypes of bacteria attached to *S. strix* (19). The role of the bacteria is unknown, but when removed from the surface, the *S. strix* cell changes to a teardrop shape (19, 22), suggesting that the stellar shape is an adaptation to harbor bacterial symbionts. The role of *S. strix* in the gut of the *Zootermopsis* termites is also unclear: there is no evidence from microscopy that *S. strix* does ingest wood particles (14, 19), and selective defaunation experiments have suggested that *S. strix* might not be involved in cellulose digestion in the gut of the termite (14).

We used single-cell genomics to sequence the metagenome of *S. strix* and all of its physically associated prokaryotes from the gut of the termite *Zootermopsis angusticollis*, resulting in a partial eukaryotic genome together with the genomes from the complex community of Bacteroidetes bacteria on its surface. The functional annotation of these genomes suggests that while the eukaryote itself is not involved in the digestion of lignocellulose, the community of bacteria covering its enlarged surface is able to digest lignocellulose and to fix nitrogen.

Results

Composition of the *S. strix* Surface Community. In cross-section, *S. strix* has a stellar shape with surface folds accommodating attachment sites for several morphotypes of elongate bacteria (19) (Fig. 1). We picked a single host cell including its bacterial symbionts, amplified its total DNA, and sequenced it. From the initial assembly, 5 main bins were retrieved using the tetraESOM (23) binning method: a eukaryotic bin, containing the genome of *S. strix*; an initial Bacteroidetes bin, containing the mixed genomes of Bacteroidetes bacteria; bins of *Pseudomonas* sp., *Endomicrobium* sp., *Treponema* sp.; and finally a bin containing sequences from various other bacteria (e.g., *Agrobacterium* and *Variovorax*). We did a second binning of the Bacteroidetes bin using Anvi'o v4 (24) and separated a single genome of a species described here as *Candidatus* Ordinivivax streblomastigis, while the rest of the Bacterioidetes sequences formed a bin of closely related miscellaneous Bacteroidales symbionts.

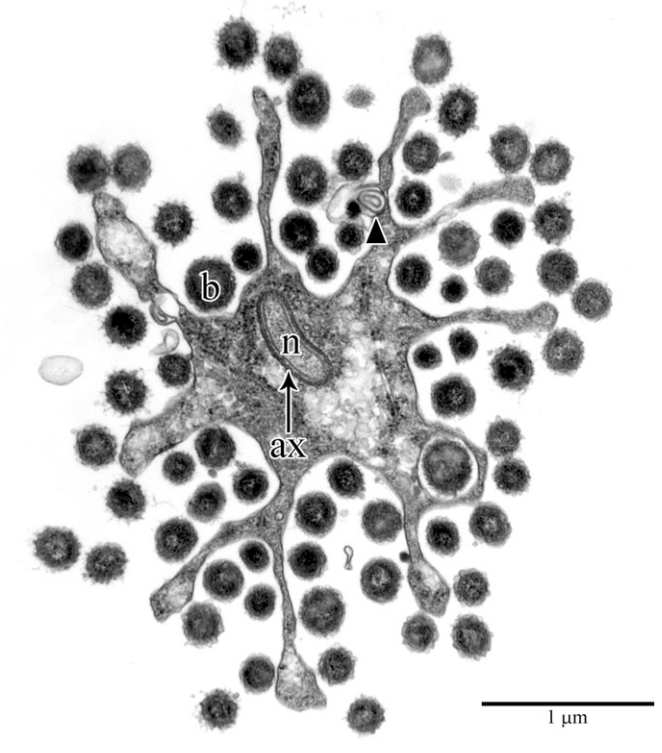

Fig. 1. Transmission electron micrograph of *S. strix* from the gut of *Z. angusticollis* showing the stellar-shaped eukaryotic cell with bacteria (b) attached to the surface. The cell nucleus (n) surrounded by the axostyle (ax) is clearly visible. The arrowhead indicates the putative process of phagocytosis of a bacterium by *S. strix*. (Scale bar, 1 μm.)

We tested whether all putative symbionts were indeed attached to *S. strix* using specific 16S rRNA fluorescence in situ hybridization (FISH) probes. Probes for *Pseudomonas*, *Treponema*, *Agrobacterium*, and *Variovorax* showed no signal on the cells of *S. strix*, and since they were also underrepresented in the sequencing, they were treated as contaminants and not considered further. Only 1 out of 3 16S rRNA FISH probes designed for *Endomicrobium* sp. showed signal on some cells of *S. strix*, but we were not able to confirm these results using the general TG1End1023T1 (15) probe for Endomicrobia, even though this probe shows a perfect match in the identified 16S rRNA sequence. For this reason, and also since the read coverage of the genome assembly is very low, we do not consider *Endomicrobium* sp. to be a symbiont of *S. strix*. In contrast, the specific and general FISH probes against 16S rRNA from *Ca.* O. streblomastigis, as well as members of the miscellaneous Bacteroidales bin, all showed clear signals strictly associated with *S. strix* (see below), and so we consider these to be bona fide symbionts. The final metagenomic dataset therefore contains 3 bins: *S. strix*, *Ca.* O. streblomastigis, and multiple species of miscellaneous Bacteroidales (Table 1).

To evaluate the diversity within the miscellaneous Bacteroidales bin, we initially identified 7 complete and 16 partial small subunit (SSU) rRNA sequences in the assembly. In the next step, we used specific primers (*SI Appendix*, Table S1) and successfully amplified 14 complete sequences, which linked all of the partial sequences from the metagenome. Not all amplified sequences were identical to the sequences in the assembly, but they were covered by reads along their whole length, indicating that they are correct, whereas sequences in the assembly are probably chimeric. Only these 14 PCR-amplified versions are considered as bona fide SSU rRNA genes of Bacteroidetes lineages compiling the miscellaneous Bacteroidales bin and were used in further steps.

Phylogenetic analysis of the 16S SSU rRNA gene (Fig. 2) showed that the sequence of the bacterium *Ca.* O. streblomastigis grouped within the Cluster V of the order Bacteriodales, a cluster containing sequences specifically from the gut of termites (21, 25, 26). Multigene phylogeny also confirmed this result (*SI Appendix*, Fig. S1) as *Ca.* O. streblomastigis robustly grouped together with *Candidatus* Azobacteroides pseudotrichonymphae and *Ca.* Symbiothrix dinenymphae.

All sequences from the miscellaneous Bacteroidales bin grouped as a sister clade to the genus *Bacteroides*, *Paraprevotella*, *Alloprevotella*, and *Prevotella* with full support. They formed 4 major clades, 3 of which contained a previously identified phylotype of a symbiont of *Streblomastix* sp. from *Archotermopsis* sp. (21).

Localization of Bacterial Symbionts by rRNA FISH. We designed 3 probes (ST-BACT-137, ST-BACT-850, and ST-BACT-1261) to target the 16S rRNAs of the 3 major clades of the miscellaneous Bacteroidales bin (*SI Appendix*, Table S2). FISH results showed that all of these bacteria are in physical contact with *S. strix* (Fig. 3 *A*–*C*). The colocalization further suggested that the probes target different bacterial cells and not versions of 16S rRNA genes within one cell (Fig. 3 *D* and *E*). We also designed specific probes for the individual sequence of clade I (ST-BACT-1020) and *Ca.* O. streblomastigis (ST-BACT-81). FISH showed that they are also associated with *S. strix* (Fig. 3 *F* and *G*). Interestingly, the probe targeting *Ca.* O. streblomastigis showed a slightly different spatial pattern of somewhat more pronounced rows of shorter and slightly wider rods (Fig. 3*G*).

As 3 of the 4 Bacteroidetes clades were composed of 3 or more sequences, we tried to establish if they represent physical ribotypes with different spatial locations by designing 2 probes

Table 1. General features of the genomes obtained in this study

Sample	No. of scaffolds	Total length, bp	N50, kbp	Completeness, %	Redundancy, %	Predicted genome size, Mbp	G+C content, %	Putative CDS	rRNA
Ca. O. streblomastigis	247	4,846,489	87.70	97.8	11.5	4.95	42.7	4,044	16S, 23S, 5S
Miscellaneous Bacteroidales symbionts	12,036	39,722,639	7.58	97.8	1,004	N/A	39.9	44,132	16S, 23S, 5S
S. strix	50,889	152,152,197	5.18	53.38/69.6	N/A	N/A	26.6	56,706	18S, 28S, 5.8S, 5S

For the bacterial datasets, the genome completeness was estimated using Anvi'o v4. The genome completeness for *S. strix* was estimated with CEGMA (first number) and BUSCO using the odb9 database (second number). N/A, not applicable.

with narrower specificity within the more diversified clades II and III (Fig. 3 *K*, *L*, *P*, and *Q*). Our FISH colocalization suggests that each probe distinguishes different cells of Bacteroidetes (Fig. 3 *M*, *N*, *R*, and *S*). In both cases, there are cells that hybridize to the general clade-specific probe, but none of the more specific probes (Fig. 3 *J* and *O*), suggesting that one or more additional ribotypes not targeted by specific probes are associated with *S. strix*.

Genome of *S. strix*. In order to understand the role of each individual organism in the community associated with *S. strix*, we analyzed their genomic bins with particular focus on the metabolic pathways. The draft genome of *S. strix* was reassembled from the metagenomic reads mapping to the bin of *S. strix* contigs. The final assembly consisted of 50,889 scaffolds with a total size of ∼152 Mbp and G+C content of 26.6% with an N50 of 5,180 bp (Table 1). The genome uses the noncanonical genetic code with TAA and TAG codons encoding the amino acid glutamine, as has been described previously (27). Out of the 56,706 predicted coding sequences, only 34,455 were found to contain start and stop codons; the rest are incomplete gene models. The number of predicted genes is most probably an overestimate due to the fragmented assembly. Automatic annotation assigned functions to 6,716 genes. The intron density within the genome is high, with 101,011 predicted introns with an average of 1.78 introns per gene. Using the Core Eukaryotic Genes Mapping Approach (CEGMA) (28) and Benchmarking Universal Single-Copy Orthologs (BUSCO), we estimated a genome completeness of 53.63 and 69.6%, respectively.

We also sequenced 6 single-cell transcriptomes from *S. strix*. The transcriptomes were manually decontaminated, but they seem to be contaminated with data from parabasalids, and we were unable to remove all contigs of parabasalid origin as no suitable reference genomes are available. Thus, for completeness estimates and annotation, we used only transcripts with the *S. strix* genetic code. The BUSCO estimates for transcriptome completeness ranges between 16.5 and 46.3%. Combining the genome and transcriptome datasets yielded a predicted protein set that is, according to BUSCO, 76.9% complete. This result is close to the estimated genome completeness of the only oxymonad genome, from *Monocercomonoides exilis*, at 63.3% with CEGMA (29) and 76.6% with BUSCO.

For the metabolic predictions, we used both the genome and transcriptomes, and we focused on the energy, amino acid, nucleotide, vitamin, and fatty acid metabolism (Dataset S1). We did not identify any exo- or endocellulases in the genome of *S. strix* (Dataset S1), but annotation of the glycosyl hydrolases (GHs) from the transcriptomes identified several cellulases from the GH5 and GH7 family. Manual investigation of these genes showed that none of them uses *S. strix* genetic code and most of them have TAA or TAG as stop codon, which in *S. strix*, encode for glutamine. While we cannot exclude the possibility that the genes that use TGA as a stop codon or have incomplete 3′ end, originated from *S. strix* (Dataset S2), their GC content (41.6 to 64.4%) is much higher than the average GC content of *S. strix* coding sequences (33.5%). In the genome, we also identified several β-glucosidases, which are involved in degradation of the disaccharide cellobiose, which is produced during cellulose degradation.

Energy metabolism apparently consists of extended glycolysis and fermentation, with the end products being pyruvate, acetate, and ethanol. Oxidative phosphorylation is absent, as is expected for an oxymonad. Several sugar transporters were identified (Dataset S1) suggesting that *S. strix* uses sugars from its environment. Several drug metabolite transporters (DMTs) were also identified in the genome, and phylogenetic classification suggests that they might be nucleotide sugar transporters (*SI Appendix*, Fig. S2). *S. strix* can synthesize 5 amino acids de novo (Cys, Ser, Ala, Glu, Gln), but it contains enzymes to interconvert several other amino acids and 4 genes encoding 2 types of transporters that could be used to import amino acids (Fig. 4). Regarding the nucleotide metabolism, *S. strix* probably cannot synthesize any nucleotide de novo, but apparently it can scavenge and recycle nucleotides from its environment or food, as has been reported for another oxymonad (30), metamonads, and trypanosomes (31). Similarly, most metabolic pathways involved in vitamin and cofactor biosynthesis seem incomplete. *S. strix* can synthesize de novo just *S*-adenosyl methionine (SAM) from methionine and NAD^+ and $NADP^+$ if provided with a source of nicotinate. The enzymes involved in synthesis of short-chain fatty acids were not identified, but most of the enzymes involved in synthesis of long-chain fatty acids are present. Several hydrogenases were identified; however, in the absence of maturases, it is not clear how active they are. It should be noted that even if no maturases were identified, the hydrogenases seem to be expressed, as they were identified in all 6 single-cell transcriptomes analyzed.

Genome of *Ca.* Ordinivivax streblomastigis. The genome assembly of *Ca*. O. streblomastigis consists of 247 scaffolds with an assembly size of ∼4.85 Mbp (Table 1). We identified 2 rRNA operons and 55 transfer RNA genes and predicted 4,043 partial and complete coding sequences. The 16S rRNA sequence reamplified manually from the whole-genome amplification sample was identical to the sequence present in the genome assembly. The genome completeness was estimated at 97.8% with a redundancy of 11.5%, but all redundant genes are identical in amino acid sequences.

The predicted metabolic pathways are shown in Fig. 5, and the list of annotated features can be found in Dataset S3. The genome encoded enzymes involved in glycolysis, gluconeogenesis, biosynthesis of various amino acids, nonoxidative pentose pathway, biosynthesis of nucleotides, vitamins, and cofactors. We also screened the genome for GHs potentially involved in cellulose and hemicellulose degradation using the Carbohydrate-Active enZYmes (http://www.cazy.org/). We identified a total of 212 GH genes, classified in 42 different families, many of them containing secretion signals (Dataset S4). The most represented family was the

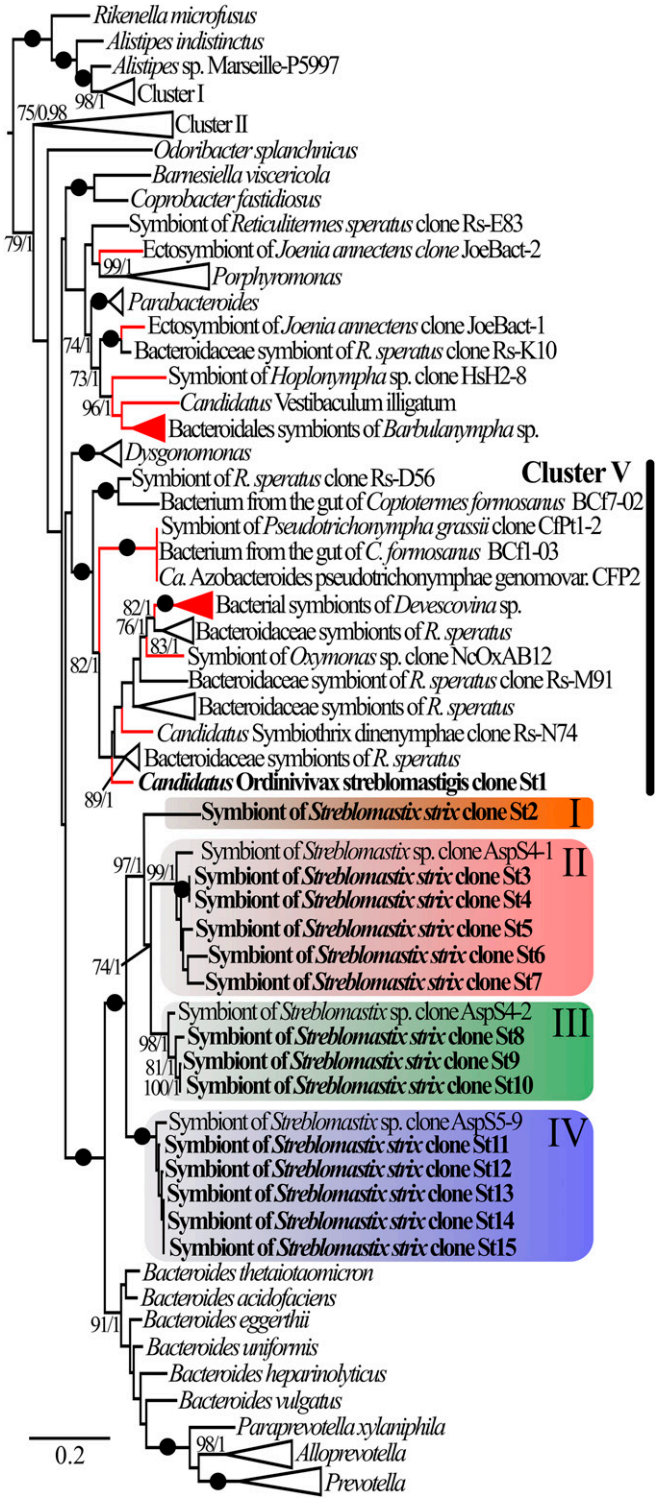

Fig. 2. Maximum likelihood (ML) tree showing the relationship of our isolated bacterial 16S rRNA phylotypes within Bacteroidetes. The tree was constructed using 16S rRNA gene sequences. Values at the nodes represent ML bootstrap support values (>70%) and bayesian inference posterior probabilities (>0.9) (BI). Dots represent full node support (ML: 100%; BI: 1). The tree was rooted with representatives of the genus *Flavobacterium* and *Capnocytophaga* (not shown). The symbionts of *S. strix* are shown in bold. The lineages within the miscellaneous Bacteroidales metagenome are highlighted and labeled from I to IV. The clusters of *Bacteroidales* bacteria symbionts from the termite gut are defined as Cluster I, II, and V according to Ohkuma et al. (26). Sequences originating from bacterial symbionts of other flagellates from the gut of the termites have their branches colored in red. (Scale bar, 0.2 expected substitutions per site.)

GH2, which were predicted to encode β-galactosidases and β-glucuronidases. The next most represented families are GH92, GH78, and GH95, which putatively encode α-mannosidases, α-L-rhamnosidases, and α-L-fucosidase. It is well known that rhamnose and mannose are components of many hemicelluloses. Representatives of GH43 and GH51 families, encoding α-L-arabinofuranosidases and β-xylosidases, were also present, suggesting that *Ca*. O. streblomastigis is involved in hemicellulose degradation. The GH5 family contained 5 representatives, and using dbCAN2 (32), we identified 3 enzymes belonging to subfamily 2 of GH5 family putatively encoding β-1,4-glucan cleaving enzymes (33). We also identified members of the GH94 family, which putatively encode cellobiose phosphorylases. The hypothesis that the bacterium is involved in the degradation of polysaccharides is also supported by the fact that we identified several copies of the starch utilization system (SUS) operon which is present in many saccharolytic members belonging to the phylum Bacteroidetes. With the exception of SusR, we have identified all components of this system, the most important SusD (54 partial and complete paralogues) and SusC (78 partial and complete paralogues) being responsible for binding of complex polysaccharides and translocation of oligosaccharides into the periplasm for further processing (34–36).

The cell contains transporters for most of the sugars resulting from the cellulose and hemicellulose degradation – xylose, rhamnose, glucose, and hexuronate, together with the enzymes required to funnel these substrates into the glycolysis (Fig. 5). These sugars are probably used as the primary source of carbon and energy, and they are converted to acetate or to succinate via fumarate reductase and subsequently to 2-oxoglutarate, which might serve as a substrate for amino acid biosynthesis. The reduced ferredoxin is regenerated using a sodium-translocating ferredoxin:NADH oxidoreductase (Rnf) complex (37), with subsequent production of NADH, which is reoxidized by the NADH-quinone oxidoreductase (NQR) complex. Just one subunit of a putative NADP-reducing hydrogenase was identified (HndC), so it is unclear if the bacterium can produce hydrogen. We also identified genes which encode a nitrogenase (NifHDK) together with various regulatory and cofactor proteins, suggesting that *Ca*. O. streblomastigis is able to fix nitrogen, but the genome also encodes an ammonia transporter (Amt). Biosynthetic pathways for de novo biosynthesis of 15 amino acids were identified. The threonine and tyrosine biosynthetic pathways lack a single enzyme, potentially due to incompleteness of the genome. A similar situation was found in the biosynthesis of chorismate, where we did not find any gene encoding for aroC, involved in the last step of chorismate biosynthesis. The genome encodes complete sets of genes for the synthesis of nucleotides and several cofactors (Fig. 5).

Metagenome of Miscellaneous Bacteroidales Symbionts. The metagenome of miscellaneous Bacteroidales symbionts consisted of 12,036 scaffolds, with an assembly size of ∼39.7 Mbp (Table 1), an estimated completeness of 97.8%, and a redundancy of 1,004%, which would suggest the presence of ∼10 to 11 bacterial genomes. We predicted 44,132 complete or partial coding sequences. The predicted metabolic pathways are shown in Fig. 6, and the annotated features can be found in Dataset S5. We considered a pathway to be present in all bacteria in the metagenome (shown in dark colors in Fig. 6) if at least 10 copies of all of the genes involved in the pathway were found. As in the case of *Ca*. O. streblomastigis, the metagenome contains an array of GH genes. We identified 747 genes that encode GHs, and we classified them in 61 GH families (Dataset S6). The most represented families are similar to the ones in the genome of *Ca*. O. streblomastigis, but the metagenome also contains representatives of GH9 annotated as endoglucanases and more representatives of GH5 family, in most cases, classified into GH5_46 subfamily, which contains enzymes active on model plant cell wall compounds (33). Based on the complex array of GH genes, it seems that the bacteria represented in the metagenome are also

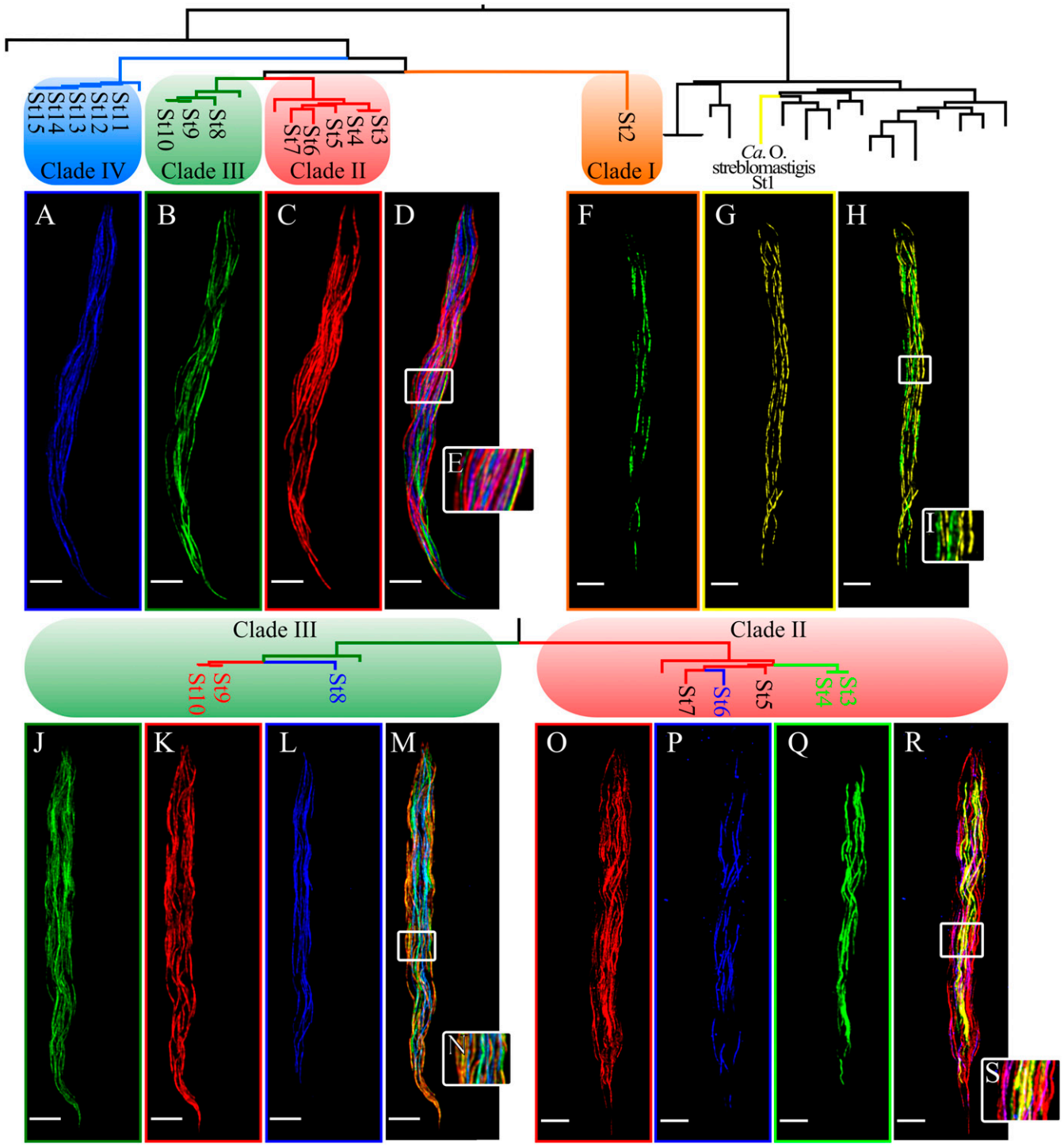

Fig. 3. FISH identification of the ectosymbionts of *Streblomastix strix*. (*Top Left*) Localization of the clades II, III, and IV of Bacteroidetes symbionts of *S. strix* using ST-BACT-850 (*A*), ST-BACT-137 (*B*), and ST-BACT-1261 (*C*). (*D*) Merged image of *A*–*C*. (*E*) Enlarged view of the region marked by a rectangle in *D*. (*Top Right*) Localization of St1 (*Ca*. O. streblomastigis) and St2 (Clade I) bacterial ribotypes on *S. strix* using ST-BACT-1020 (*F*) and ST-BACT-81 (*G*). Merged image of *F* and *G* is shown in *H*. (*I*) Enlarged view of the region marked by a rectangle in *H*. (*Bottom Left*) Specific localization of different ribotypes within clade III. (*J*) General localization of the bacteria from clade III using the general probe ST-BACT-137. (*K*) Specific localization of ST8 and ST9 ribotypes using the ST-BACT-734 probe. (*L*) Specific localization of ST8 ribotype using the ST-BACT-196 probe. (*M*) Merged picture of *J*–*L*. (*N*) Enlarged view of the region marked by a rectangle in *M*. (*Bottom Right*) Specific localization of different ribotypes within clade II. (*O*) General localization of the bacteria from clade II using the general probe ST-BACT-1261. (*P*) Specific localization of the ST6 ribotype using the ST-BACT-477 probe. (*Q*) Specific localization of ST3 and ST4 ribotypes using the ST-BACT-845 probe. (*R*) Merged picture of *O*–*Q*. (*S*) Enlarged view of the region marked by a rectangle in *R*. (Scale bars, 5 μm.)

adapted to digest lignocellulose. We also identified 273 and 187 complete or partial paralogs of SusC and SusD, respectively.

The metagenome also encodes several transporters for glucose, xylose, rhamnose, fucose, and hexuronate, together with the metabolic pathways to use these sugars as a substrate for glycolysis. Interestingly, we also identified a transporter for glycerol-3-phosphate, another substrate for glycolysis. Very importantly, in the entire metagenome, we did not identify any gene for the glycolytic

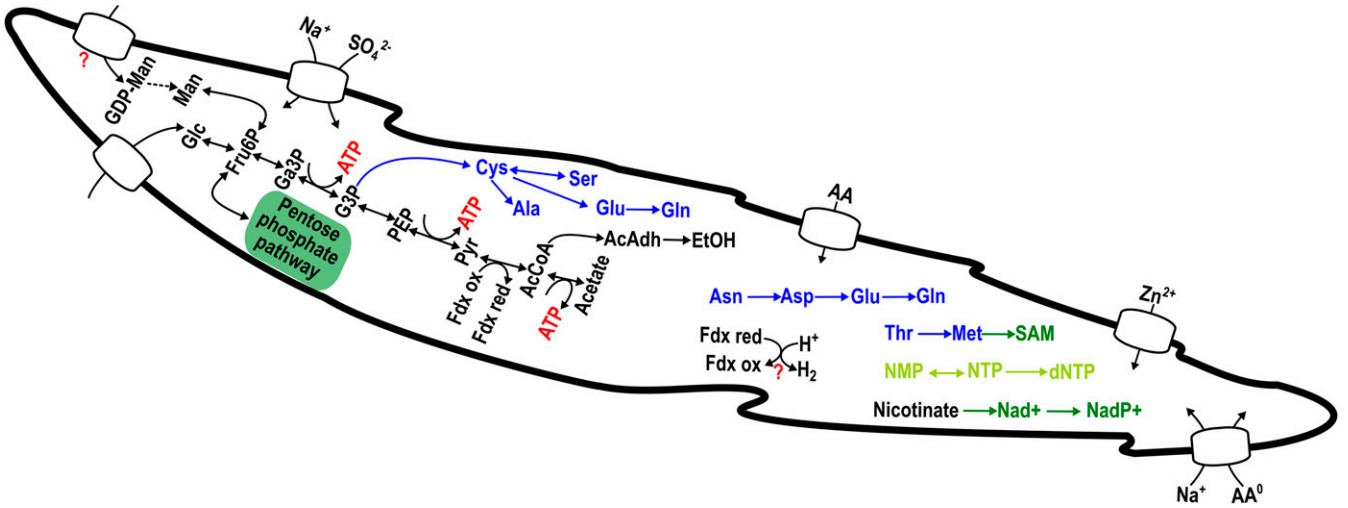

Fig. 4. Metabolic map of *Streblomastix strix* reconstructed from the draft genome. Links to the pathways of amino acid biosynthesis are shown in blue, and links to vitamin and cofactor biosynthesis are shown in green. AA, amino acid; AA^0, neutral amino acid; AcAdh, acetaldehyde; AcCoA, acetyl-coenzyme A; dNTP, deoxyribonucleotide triphosphate; Fru6P, fructose-6-phosphate; G3P, glycerate 3-phosphate; Ga3P, glyceraldehyde 3-phosphate; Glc, glucose; GDP-Man, guanosine diphosphate mannose; Man, mannose; NMP, nucleoside monophosphate; NTP, nucleoside triphosphate; Pyr, pyruvate.

enzyme enolase, suggesting that these bacteria are either unable to complete glycolysis and accumulate glycerate 2-phosphate or use an unknown alternative pathway. The former contrasts with the fact that all enzymes necessary to convert phosphoenolpyruvate (PEP) to pyruvate and further to acetate were identified (Fig. 6) and seem to be present in all bacteria from the metagenome. Pyruvate can also be converted to succinate using fumarate reductase, which can later be converted to 2-oxoglutarate and used in biosynthesis of amino acids. The methylglyoxal pathway (Fig. 6) may serve as an alternative to enolase, but it seems incomplete.

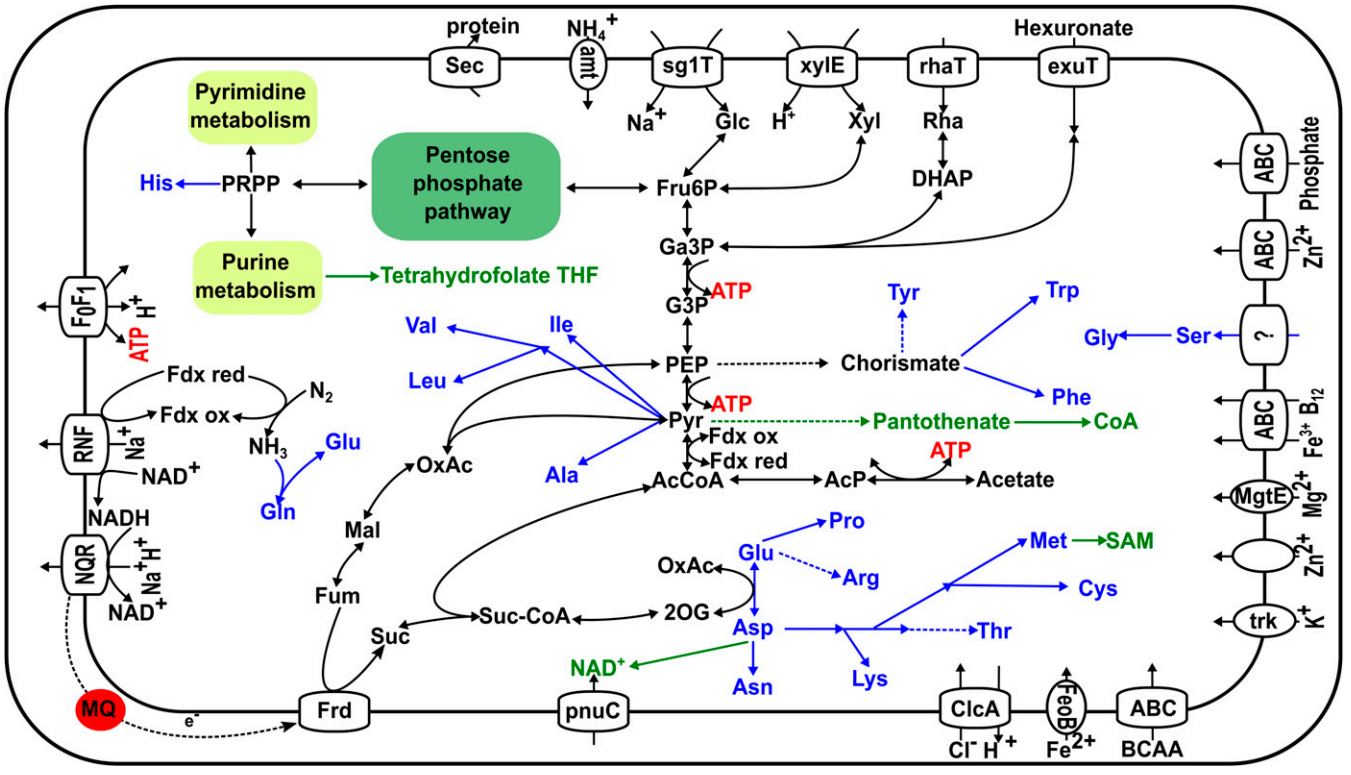

Fig. 5. Metabolic map of *Ca*. O. streblomastigis reconstructed from the draft genome. Links to the pathways of amino acid biosynthesis are shown in blue, and links to the vitamin and cofactor biosynthesis are shown in green. Dashed arrows indicate pathways that are lacking one or more essential genes. Menaquinone is shown in red because the biosynthetic pathway is incomplete. ABC, ATP-binding cassette transporter; AcCoA, acetyl-coenzyme A; AcP, acetyl phosphate; amt, ammonia transporter; BCAA, branched-chain amino acid; CoA, coenzyme A; DHAP, dihydroxyacetone phosphate; Fdx ox, oxidized ferredoxin; Fdx red, reduced ferredoxin; Frd, fumarate reductase; Fru6P, fructose-6-phosphate; Fum, fumarate; G3P, glycerate 3-phosphate; Ga3P, glyceraldehyde 3-phosphate; Glc, glucose; Mal, malate; MQ, menaquinone; 2OG, 2-oxoglutarate; OxAc, oxaloacetate; PRPP, phosphoribosyl pyrophosphate; Pyr, pyruvate; Rha, rhamnose; Sec, general secretion system; Suc, succinate; Xyl, xylose.

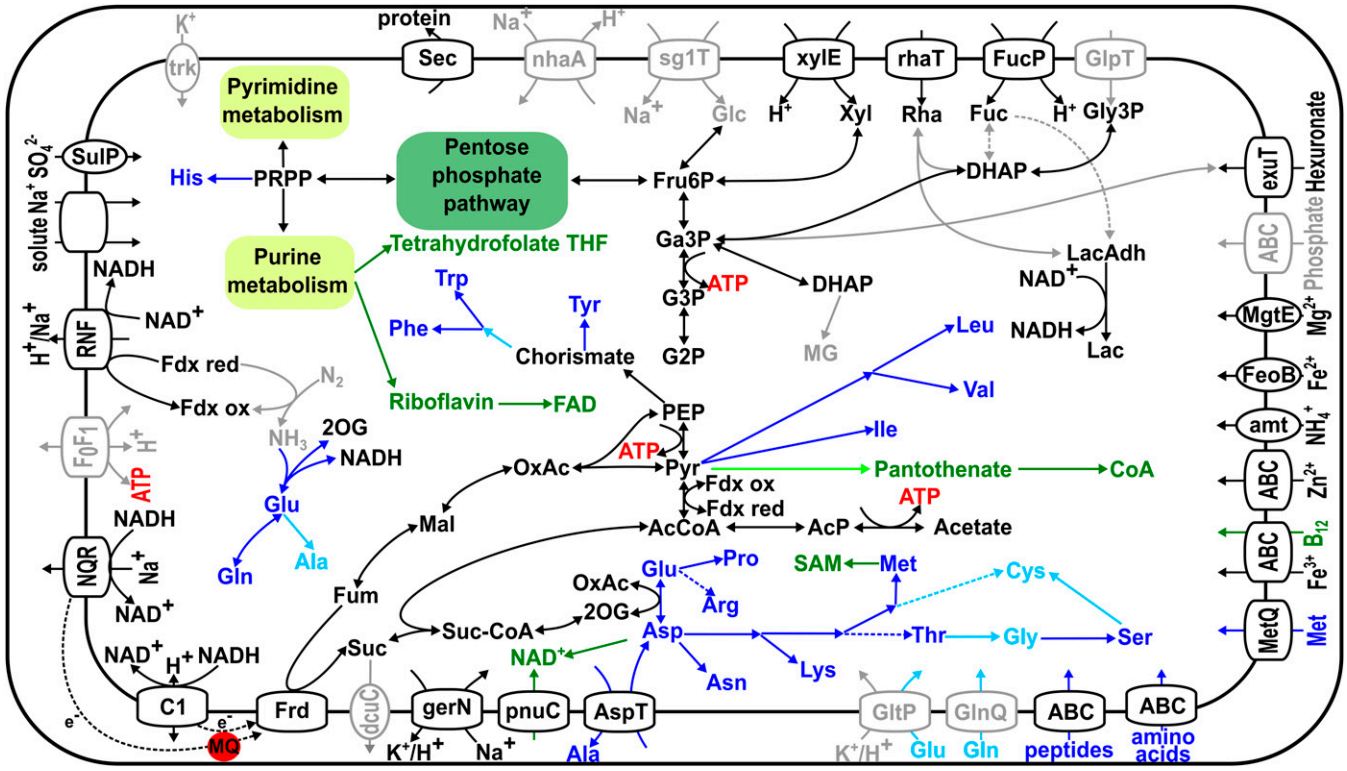

Fig. 6. Putative metabolic map of a bacterium within the miscellaneous Bacteroidales metagenome reconstructed from the metagenomic data. Links to the pathways of amino acid biosynthesis are shown in blue, and links to vitamin and cofactor biosynthesis are shown in green. Light colors (gray, light blue, or light green) represent steps in the metabolism that we think are not common for all bacteria from the metagenome. Dashed arrows indicate pathways that are lacking one or more essential genes. Menaquinone is shown in red because its biosynthetic pathway is incomplete. ABC, ATP-binding cassette transporter; AcCoA, acetyl-coenzyme A; AcP, acetyl phosphate; amt, ammonia transporter; C1, complex I; CoA, coenzyme A; DHAP, dihydroxyacetone phosphate; FAD, flavin adenine dinucleotide; Fdx ox, oxidized ferredoxin; Fdx red, reduced ferredoxin; Frd, fumarate reductase; Fru6P, fructose-6-phosphate; Fuc, fucose; Fum, fumarate; G2P, glycerate 2-phosphate; G3P, glycerate 3-phosphate; Ga3P, glyceraldehyde 3-phosphate; Glc, glucose; Gly3P, glycerol 3-phosphate; Lac, L-lactate; LacAdh, lactaldehyde; Mal, malate; MG, methylglyoxal; MQ, menaquinone; 2OG, 2-oxoglutarate; OxAc, oxaloacetate; PRPP, phosphoribosyl pyrophosphate; Pyr, pyruvate; Rha, rhamnose; Sec, general secretion system; Suc, succinate; Xyl, xylose.

The metagenome contains genes involved in the de novo biosynthesis of all amino acids except for threonine and arginine, and it seems that they are present in most members of the community. Within the metagenome, we identified several nitrogenase (NifHDK) operons, suggesting that some of the bacteria can fix nitrogen. Interestingly, we identified ammonia transporters as well as several transporters involved in transport of peptides and various amino acids.

Similar to *Ca*. O. streblomastigis, the reoxidation of the reduced ferredoxin is done by the Rnf complex, and the reoxidation of NADH is done by the NQR complex. However, these bacteria also have the complex 1 of the respiratory chain, which can reoxidize NADH and pass electrons to fumarate reductase (Fig. 6).

Discussion

By combining the sequence data of the whole genome-amplified DNA from a single cell of *S. strix*, with the single cell-amplified transcriptomes and 16S rRNA FISH, we have obtained a genome draft of *S. strix*, identified prokaryotic components of the community living in physical association with this eukaryote, mapped their diversity and spatial distribution, and propose their major metabolic roles.

The community inhabiting the surface of *S. strix* consists of a genetically complex population of Bacteroidetes bacteria. At least 15 types of Bacteroidetes 16S rRNA sequences have been recovered, and most importantly, FISH experiments with probes of various specificities confirmed the physical presence of at least 8 ribotypes. We were able to separate one individual genome described here as *Ca.* O. streblomastigis, while the remaining sequences remained in one miscellaneous Bacteroidales bin. Estimating the genome completeness of this metagenome revealed a redundancy of 1,004%, which means that it could contain 10 or 11 different genomes, which is similar to the confirmed number of ribotypes. A previous study reported 3 rRNA SSU sequences from Bacteroidetes symbionts of *Streblomastix* from the gut of *Archotermopsis* sp. and demonstrated that they form a distinct lineage sister to the genera *Bacteroides* and *Prevotella* (21). Our sequences form 4 specific clades in the same position, 3 of them being related to the previously published Bacteroidetes SSU sequences. Our FISH experiments indicate that the distribution of Bacteroidetes ribotypes on the surface of *S. strix* specimens is random. On some specimens, the 3 main clades of Bacteroidetes bacteria are equally represented, while on others, one of the clades seems dominant (*SI Appendix*, Fig. S3). In addition to Bacteroidetes bacteria, we also identified several other partial bacterial genomes, but FISH experiments failed to provide any proof that these bacteria are physically attached to *S. strix*.

From functional annotation of the individual genomes and the metagenome, a hypothesis for the overall metabolic network formed by this community emerges (Fig. 7). All *S. strix* symbionts encode a high functional diversity of GH enzymes, which degrade the complex polysaccharides present in the gut. Reconstructing the metabolism of *Ca*. O. streblomastigis is relatively straightforward; it most likely ferments sugars to acetate and succinate. However, for the remaining Bacteroidetes symbionts, the picture is not so clear. Within the miscellaneous Bacteroidales bin, we identified transporters for glucose, rhamnose, hexuronate,

and fucose, which strongly suggests that the bacteria are using these sugars as an energy source, but their glycolytic pathway is potentially incomplete since these genomes all lack enolase genes. While it would be possible to miss this gene from an individual single-cell genome, it is highly unlikely that we would specifically miss enolases from ∼10 genomes, which constitute the metagenome. So far, only 6 rumen bacterial genomes have been described that are lacking enolase; 5 from Firmicutes and 1 from *Prevotella* (38, 39). An alternative, the methylglyoxal shunt, has been suggested to bypass enolase, where dihydroxyacetone phosphate is converted to pyruvate or L-lactate via methylglyoxal (38). Use of methylglyoxal shunt has been suggested as a mechanism to decrease the ATP production while keeping the ability to dispose the excess carbohydrates (40), as this pathway leads to a net loss of ATP. In the metagenome, we identified a single methylglyoxal synthase but no enzymes that convert methylglyoxal to lactaldehyde. It should be noted that in *Escherichia coli*, it has been shown that lactaldehyde dehydrogenase can act with lower affinity on methylglyoxal and produce pyruvate (41), but the substrate specificity of this enzyme varies dramatically from organism to organism (42), and we cannot make the same conclusion without experimental evidence. A more likely function of this enzyme may be the conversion of lactaldehyde produced from degradation of rhamnose and fucose (Fig. 6) to L-lactate, which can probably be further metabolized by the cell, as it contains several operons of the lactate-utilization system (Dataset S5).

An alternative explanation for the absence of enolase is that these bacteria cooperate with *S. strix* or *Ca*. O. streblomastigis to obtain PEP or pyruvate (Fig. 7). We did not identify any specific transporters for phosphorylated trioses in the metagenome; however, we identified several EamA domain-containing proteins (Dataset S5), which are part of the Drug/Metabolite Exporter (DME) family of the DMT superfamily. While it is not clear what are the substrates for these transporters, it has been proposed that nucleotide sugar transporters, including triose phosphate transporters, evolved through differentiation of the EamA gene cluster (43), so it could be that some of these genes are involved in translocation of triose phosphates. In the genome of *S. strix*, we identified 6 putative nucleotide sugar transporters. These transporters are also part of the DMT superfamily, but phylogenetic analysis of our sequences showed that these transporters are not involved in triose phosphate translocation (*SI Appendix*, Fig. S2). Alternatively, *S. strix* could provide the bacteria with pyruvate, since the bacteria are able to convert pyruvate to acetate and produce ATP, but this would not explain the broad array of sugar transporters present in the Bacteroidetes metagenome, unless the bacteria were metabolizing these compounds.

In both the genome of *Ca*. O. streblomastigis and the metagenome, we identified several nitrogenases (NifHDK), suggesting that some of the bacteria can fix nitrogen and use ammonia for biosynthesis of amino acids (Fig. 5). Phylogenetic analysis of the *nifH* (*SI Appendix*, Fig. S4) gene suggested that they are part of the group II and group III-3a of *nifH* homologs (for nomenclature, see refs. 44 and 45), together with other sequences of *nifH* from the termite gut. Several other symbiotic bacteria of flagellates have been shown to have the ability to fix nitrogen (7–10), and it has been hypothesized that these bacteria also provide amino acids for the flagellate host. In our view, the same situation applies also to *S. strix* and its ectobionts. Moreover, since we hypothesize that not all symbionts of *S. strix* have the ability to fix nitrogen, amino acids and other nitrogenous compounds may be shared also between the bacteria. The means of the nutrient transport are unclear, but electron microscopy studies have identified pili connections between the symbiotic bacteria (19), which could provide a way for nutrient exchange.

The predicted metabolic pathways of the flagellate are very similar to the predicted metabolism of the oxymonad *Monocercomonoides exilis* strain PA203 (30). This is interesting because, while the 2 oxymonads seem to be closely related (46), they come from hosts with significantly different diets. It is well known that some flagellated protists in the gut of lower termites are important for the digestion of cellulose (47–50), and for some of them, the activity was even proven biochemically (17, 18, 51), or the cellulase genes were PCR-amplified from the genomic DNA (52). However, in the case of *S. strix*, we did not identify

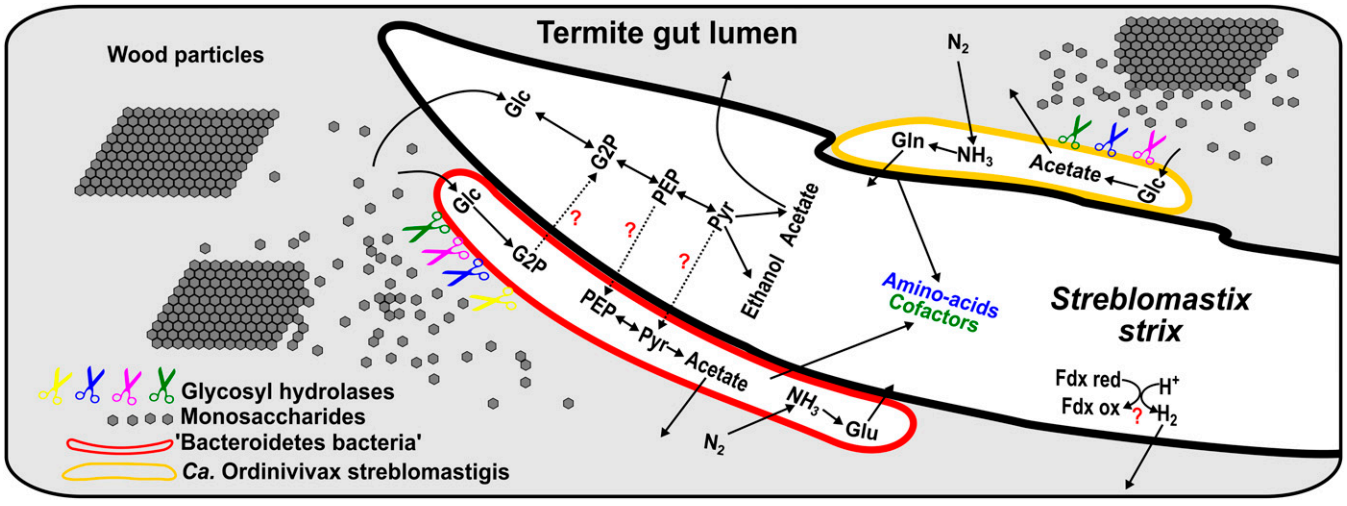

Fig. 7. Schematic view of the proposed interaction between the bacterial community and *S. strix*, in the gut of *Z. angusticollis*. The bacterial symbionts of *S. strix* use their complex array of GHs to digest the wood particles generating monosaccharides, which are taken up by the bacteria and protist. The eukaryote ferments the sugars to acetate and ethanol, which are most probably excreted into the termite gut. *Ca*. O. streblomastigis also ferments sugars to acetate and probably secretes the acetate in the gut. In the case of Bacteroidetes bacteria forming the metagenomic bin, the sugars cannot be fermented to acetate due to the lack of enolase. We hypothesize that these bacteria collaborate with *S. strix* to convert glycerate 2-phosphate (G2P) to PEP in order to complete the glycolysis or that *S. strix* provides a source of pyruvate (Pyr) for the bacteria, which would be used to produce ATP. The bacterial community also incorporates nitrogen into ammonia and some amino acids. Parts of the synthesized amino acids and various cofactors are most probably provided to *S. strix* by the bacteria, using some form of transport or via grazing on bacteria. The acetate that is secreted in the hindgut is used by the termite. Fdx ox, oxidized ferredoxin; Fdx red, reduced ferredoxin.

any genes for enzymes involved in lignocellulose digestion, but we identified several β-glucosidases, which could degrade cellobiose to glucose. This is also consistent with several studies that have suggested that *S. strix* does not ingest wood particles (14, 19, 21), and we also did not observe any wood particles inside *S. strix* in our electron microscopy preparations. Previous experiments by Cleveland (14) have shown that *S. strix* and its ectosymbionts cannot support themselves in the absence of other flagellates in the gut of the termite. If correct, it might be that partial digestion of the wood particles by other flagellates is a prerequisite for *S. strix* or that the flagellate requires some essential nutrient from other members of the gut community.

In the metabolism of *S. strix*, the monomeric sugars are converted to acetate and ethanol by extended glycolysis and fermentation (Fig. 3). We also identified several [FeFe]hydrogenases putatively involved in the reoxidation of ferredoxin; however, we did not identify any maturases, which usually help to assemble [FeFe]hydrogenases in the native conformation (53), but these maturases are missing also from the genomes of *Giardia intestinalis*, *Entamoeba histolytica*, and *Monocercomonoides exilis* (54, 55).

Based on these facts, we suggest that *S. strix* uses its bacterial ectosymbionts to at least partially degrade cellulose externally and thrives on the glucose and other simpler sugars transported into the cell using the specialized transporters. Also, as seen in the electron microscopy, *S. strix* is capable of internalizing and digesting the epibiotic bacteria (Fig. 1), which is supported by the presence of 2 peptidoglycan-degrading enzymes in the genome and transcriptomes (Dataset S2).

The single-cell metagenomic study presented here represents a detailed investigation of a superorganism composed of a single protist and its physically attached, complex community of prokaryotic cells. Reconstructing the metabolism revealed the roles of individual components as well as potential metabolic interactions. The symbiotic bacterial community covering the enlarged surface of *S. strix* is probably involved in the digestion of lignocellulose, as well as fixation of nitrogen, providing the flagellate host with a source of sugars, nitrogenous compounds, nucleotides, and amino acids. What this community gains from *S. strix* is less clear. Perhaps there is a benefit in its role as a physical scaffold to maintain the proximity of the community members, creating a suitable environment where the digestion of lignocellulose can happen efficiently. Alternatively, perhaps *S. strix* is selfishly planting the bacteria in order to digest them eventually, and it may also keep control over the majority of them by complementing their missing enolase.

Description of "*Candidatus* Ordinivivax streblomastigis gen. nov. et. sp. nov." *Ordinivivax streblomastigis* (L. n. *ordo*, gen. *ordinis*, a regular line, row, or series; L. masc. adj. *vivax*, tenacious of life, lively, vigorous, vivacious; N.L. masc. n. *Ordinivivax*, a regularly arranged living organism. N.L. gen. n. *streblomastigis*, of *Streblomastix*, a genus of flagellates). The bacterium attaches to the surface of *S. strix* from the gut of *Z. angusticollis.* This assignment is based on the genome fragments deposited in GenBank, https://www.ncbi.nlm.nih.gov/genbank (SNRX00000000), as well as specific 16S rRNA gene hybridization with probe ST-BACT-81 (*SI Appendix*, Table S2).

Materials and Methods

Cell Isolation and Whole-Genome and -Transcriptome Amplification. *Z. angusticollis* termites were collected from Galliano Island, British Columbia, Canada. The entire gut of a worker termite was removed, and the contents were resuspended in Trager U media (56). A single cell of *S. strix* was isolated by micromanipulation and washed 2 times in the same buffer, before being placed in a clear 0.2-mL PCR tube (Axygen). The genomic DNA of the isolated cell was amplified using the Illustra Single Cell GenomiPhi DNA Amplification Kit (GE Healthcare Life Sciences), according to the manufacturer's protocol, with an incubation time of 2 h. For transcriptomes, cells were picked by micromanipulation and the amplification was performed as described by Picelli et al. (57).

Genome Sequencing, Binning and Annotation. Sequencing was performed using Illumina MiSeq, Illumina HiSeq, and Oxford Nanopore technologies. Detailed sequencing procedures are described in *SI Appendix, SI Text*. Assembly of the metagenome was done using SPAdes 3.11.1 (58). Binning of the metagenomic data were performed using tetraESOM (23) and Anvi'o v4 (24). Prediction and annotation of the prokaryotic genes was performed using Prokka 1.12 (59). Predicted gene models and their annotations were then manually curated. For the eukaryotic gene predictions, we used Augustus 3.2.3 (60), and annotation was performed automatically using a combination of the blastp against a custom database and the National Center for Biotechnology Information nonredundant database. Detailed procedures of binning and annotation are available in *SI Appendix, SI Text*.

Amplification of the Bacteroidetes 16S rRNA Sequences. Specific primers (*SI Appendix*, Table S1) were designed to reamplify the full-length 16S SSU rRNA for each SSU sequence (including partial ones) found in the assembled Bacteroidetes bin. Amplifications were carried out using PrimeSTAR Max DNA polymerase premix (Clontech). The amplified fragments were gel-purified using the Qiaquick gel extraction kit (Qiagen) and cloned in the pJET vector, using the CloneJET PCR Cloning Kit (Thermo Scientific). Three individual clones were randomly selected and sequenced in both directions, using the primer walking strategy.

Sample Preparation and FISH. Several individuals of *Z. angusticollis* were transferred on cellulose powder to reduce the amount of wood particles, which cause autofluorescence (61). After 1 wk, the entire gut of a worker termite was removed, and the contents were fixed in 4% formaldehyde for 6 h. Probes for FISH were designed using ARB 6.0.4 (62) and the Silva REF NR 99 SSU database release 126. FISH preparations were performed on 10-well slides (1216690; Marienfeld GmbH) according to the protocol described in ref. 63. Detailed procedures for sample fixation and probe design are available in *SI Appendix, SI Text*.

ACKNOWLEDGMENTS. This work has received funding from the European Research Council under European Union's Horizon 2020 Research and Innovation Programme Grant Agreement 771592 (to V.H.); Grant Agency of Charles University Project 1292417 (to S.C.T.); Czech Science Foundation Project 15-16406S (to V.H.); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic (CR) within National Sustainability Program II Project (BIOCEV-FAR) LQ1604; the project "BIOCEV" (CZ.1.05/1.1.00/02.0109); the Centre for Research of Pathogenicity and Virulence of Parasites (CZ.02.1.01/0.0/0.0/16_019/0000759); the Natural Sciences and Engineering Research Council of Canada (RGPIN-2014-03994); and a grant to the University of British Columbia Centre for Microbial Diversity and Evolution from the Tula Foundation. F.H. is supported by a European Molecular Biology Organization Fellowship (ALTF 1260–2016). We acknowledge the Imaging Methods Core Facility in BIOCEV, supported by Czech-BioImaging large RI projects LM2015062 and CZ.02.1.01/0.0/0.0/16_013/0001775 Czech-BioImaging, funded by MEYS CR. Computational resources were provided by the CESNET LM2015042 and ERIT Scientific Cloud LM2015085, provided under the program "Projects of Large Research, Development, and Innovations Infrastructures." We also thank Prof. Dr. Ivan Čepička and Prof. Dr. Andreas Brune for their help with the taxonomy and fruitful discussion on the manuscript; and Prof. Dr. Ivan Hrdý for providing some of the termites used in FISH.

1. A. Brune, Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* **12**, 168–180 (2014).
2. A. Brune, "Methanogens in the digestive tract of termites" in *(Endo)Symbiotic Methanogenic Archaea*, J. H. P. Hackstein, Ed. (Springer Berlin Heidelberg, 2018), pp. 81–101.
3. Y. Hongoh, Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell. Mol. Life Sci.* **68**, 1311–1325 (2011).
4. J. F. H. Strassert, A. Mikaelyan, T. Woyke, A. Brune, 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 (2016).
5. Y. Hongoh *et al.*, 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 (2008).
6. H. Kuwahara, M. Yuki, K. Izawa, M. Ohkuma, Y. Hongoh, Genome of '*Ca.* Desulfovibrio trichonymphae', an H_2-oxidizing bacterium in a tripartite symbiotic system within a protist cell in the termite gut. *ISME J.* **11**, 766–776 (2017).
7. W. Ikeda-Ohtsubo *et al.*, '*Candidatus* Adiutrix intracellularis', an endosymbiont of termite gut flagellates, is the first representative of a deep-branching clade of Deltaproteobacteria and a putative homoacetogen. *Environ. Microbiol.* **18**, 2548–2564 (2016).
8. Y. Hongoh *et al.*, Genome of an endosymbiont coupling N_2 fixation to cellulolysis within protist cells in termite gut. *Science* **322**, 1108–1109 (2008).

9. V. Tai *et al*., Genome evolution and nitrogen-fixation in bacterial ectosymbionts of a protist inhabiting wood-feeding cockroaches. *Appl. Environ. Microbiol.* **82**, 4682–4695 (2016).
10. M. Ohkuma *et al*., Acetogenesis from H_2 plus CO_2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10224–10230 (2015).
11. M. Yuki *et al*., Dominant ectosymbiotic bacteria of cellulolytic protists in the termite gut also have the potential to digest lignocellulose. *Environ. Microbiol.* **17**, 4942–4953 (2015).
12. Y. D. Utami *et al*., 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 (2019).
13. V. Tai, E. R. James, S. J. Perlman, P. J. Keeling, Single-cell DNA barcoding using sequences from the small subunit rRNA and internal transcribed spacer region identifies new species of *Trichonympha* and *Trichomitopsis* from the hindgut of the termite *Zootermopsis angusticollis*. *PLoS One* **8**, e58728 (2013).
14. L. R. Cleveland, The effects of oxygenation and starvation on the symbiosis between the termite, *Termopsis,* and its intestinal flagellates. *Biol. Bull.* **48**, 309–326 (1925).
15. W. Ikeda-Ohtsubo, M. Desai, U. Stingl, A. Brune, Phylogenetic diversity of 'Endomicrobia' and their specific affiliation with termite gut flagellates. *Microbiology* **153**, 3458–3465 (2007).
16. W. Ikeda-Ohtsubo, A. Brune, Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol. Ecol.* **18**, 332–342 (2009).
17. M. A. Yamin, Cellulose metabolism by the flagellate *Trichonympha* from a termite is independent of endosymbiotic bacteria. *Science* **211**, 58–59 (1981).
18. M. A. Yamin, W. Trager, Cellulolytic activity of an axenically-cultivated termite flagellate, *Trichomitopsis termopsidis*. *J. Gen. Microbiol.* **113**, 417–420 (1979).
19. B. S. Leander, P. J. Keeling, Symbiotic innovation in the oxymonad *Streblomastix strix*. *J. Eukaryot. Microbiol.* **51**, 291–300 (2004).
20. C. A. Kofoid, O. Swezy, Studies on the parasites of the termites: On *Streblomastix strix*: A polymastigote flagellate with a linear plasmodial phase. *Univ. Calif. Publ. Zool.* **20**, 1–20 (1919).
21. S. Noda *et al*., 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 (2006).
22. B. D. Dyer, O. Khalsa, Surface bacteria of *Streblomastix strix* are sensory symbionts. *Biosystems* **31**, 169–180 (1993).
23. G. J. Dick *et al*., Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
24. A. M. Eren *et al*., Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
25. Y. Hongoh, M. Ohkuma, T. Kudo, Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiol. Ecol.* **44**, 231–242 (2003).
26. M. Ohkuma, S. Noda, Y. Hongoh, T. Kudo, Diverse bacteria related to the bacteroides subgroup of the CFB phylum within the gut symbiotic communities of various termites. *Biosci. Biotechnol. Biochem.* **66**, 78–84 (2002).
27. P. J. Keeling, B. S. Leander, Characterisation of a non-canonical genetic code in the oxymonad *Streblomastix strix*. *J. Mol. Biol.* **326**, 1337–1349 (2003).
28. G. Parra, K. Bradnam, I. Korf, CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
29. A. Karnkowska *et al*., A Eukaryote without a mitochondrial organelle. *Curr. Biol.* **26**, 1274–1284 (2016).
30. A. Karnkowska *et al*., The oxymonad genome displays canonical eukaryotic complexity in the absence of a mitochondrion. *Mol. Biol. Evol.*, 10.1093/molbev/msz147 (2019).
31. H. P. de Koning, D. J. Bridges, R. J. S. Burchmore, Purine and pyrimidine transport in pathogenic protozoa: From biology to therapy. *FEMS Microbiol. Rev.* **29**, 987–1020 (2005).
32. H. Zhang *et al*., dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
33. H. Aspeborg, P. M. Coutinho, Y. Wang, H. Brumer, 3rd, B. Henrissat, Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* **12**, 186 (2012).
34. E. C. Martens, N. M. Koropatkin, T. J. Smith, J. I. Gordon, Complex glycan catabolism by the human gut microbiota: The Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673–24677 (2009).
35. D. N. Bolam, N. M. Koropatkin, Glycan recognition by the Bacteroidetes Sus-like systems. *Curr. Opin. Struct. Biol.* **22**, 563–569 (2012).
36. N. T. Porter, A. S. Luis, E. C. Martens, *Bacteroides thetaiotaomicron*. *Trends Microbiol.* **26**, 966–967 (2018).
37. V. Hess *et al*., Occurrence of ferredoxin:NAD(+) oxidoreductase activity and its ion specificity in several Gram-positive and Gram-negative bacteria. *PeerJ* **4**, e1515 (2016).
38. W. J. Kelly *et al*., The glycobiome of the rumen bacterium *Butyrivibrio proteoclasticus* B316T highlights adaptation to a polysaccharide-rich environment. *PLoS One* **5**, e11942 (2010).
39. T. J. Hackmann, D. K. Ngugi, J. L. Firkins, J. Tao, Genomes of rumen bacteria encode atypical pathways for fermenting hexoses to short-chain fatty acids. *Environ. Microbiol.* **19**, 4670–4683 (2017).
40. J. B. Russell, Strategies that ruminal bacteria use to handle excess carbohydrate. *J. Anim. Sci.* **76**, 1955–1963 (1998).
41. L. Baldomà, J. Aguilar, Involvement of lactaldehyde dehydrogenase in several metabolic pathways of *Escherichia coli* K12. *J. Biol. Chem.* **262**, 13991–13996 (1987).
42. Y. Inoue, A. Kimura, Methylglyoxal and regulation of its metabolism in microorganisms. *Adv. Microb. Physiol.* **37**, 177–227 (1995).
43. Å. Västermark, M. S. Almén, M. W. Simmen, R. Fredriksson, H. B. Schiöth, Functional specialization in nucleotide sugar transporters through differentiation of the gene cluster EamA (DUF6) before the radiation of Viridiplantae. *BMC Evol. Biol.* **11**, 123 (2011).
44. M. S. Desai, A. Brune, Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. *ISME J.* **6**, 1302–1313 (2012).
45. A. Yamada, T. Inoue, S. Noda, Y. Hongoh, M. Ohkuma, Evolutionary trend of phylogenetic diversity of nitrogen fixation genes in the gut community of wood-feeding termites. *Mol. Ecol.* **16**, 3768–3777 (2007).
46. S. C. Treitli *et al*., Molecular and morphological diversity of the oxymonad genera *Monocercomonoides* and *Blattamonas* gen. nov. *Protist* **169**, 744–783 (2018).
47. M. Ohkuma, Termite symbiotic systems: Efficient bio-recycling of lignocellulose. *Appl. Microbiol. Biotechnol.* **61**, 1–9 (2003).
48. A. Brune, M. Ohkuma, "Role of the termite gut microbiota in symbiotic digestion" in *Biology of Termites: A Modern Synthesis*, D. E. Bignell, Y. Roisin, N. Lo, Eds. (Springer, Dordrecht, The Netherlands, 2011), pp. 439–475.
49. A. Brune, C. Dietrich, The gut microbiota of termites: Digesting the diversity in the light of ecology and evolution. *Annu. Rev. Microbiol.* **69**, 145–166 (2015).
50. J. Ni, G. Tokuda, Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol. Adv.* **31**, 838–850 (2013).
51. D. A. Odelson, J. A. Breznak, Nutrition and growth characteristics of *Trichomitopsis termopsidis*, a cellulolytic Protozoan from termites. *Appl. Environ. Microbiol.* **49**, 614–621 (1985).
52. K. Ohtoko *et al*., Diverse genes of cellulase homologues of glycosyl hydrolase family 45 from the symbiotic protists in the hindgut of the termite *Reticulitermes speratus*. *Extremophiles* **4**, 343–349 (2000).
53. S. Pütz *et al*., Fe-hydrogenase maturases in the hydrogenosomes of *Trichomonas vaginalis*. *Eukaryot. Cell* **5**, 579–586 (2006).
54. J. E. J. Nixon *et al*., Iron-dependent hydrogenases of *Entamoeba histolytica* and *Giardia lamblia*: Activity of the recombinant entamoebic enzyme and evidence for lateral gene transfer. *Biol. Bull.* **204**, 1–9 (2003).
55. D. Lloyd, J. R. Ralphs, J. C. Harris, *Giardia intestinalis*, a eukaryote without hydrogenosomes, produces hydrogen. *Microbiology* **148**, 727–733 (2002).
56. W. Trager, The cultivation of a cellulose-digesting flagellate, *Trichomonas termopsidis*, and of certain other termite protozoa. *Biol. Bull.* **66**, 182–190 (1934).
57. S. Picelli *et al*., Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
58. A. Bankevich *et al*., SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
59. T. Seemann, Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
60. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (suppl. 2), ii215–ii225 (2003).
61. R. Radek *et al*., Phylogeny and ultrastructure of *Oxymonas jouteli*, a rostellum-free species, and *Opisthomitus longiflagellatus* sp. nov., oxymonadid flagellates from the gut of *Neotermes jouteli*. *Protist* **165**, 384–399 (2014).
62. W. Ludwig *et al*., ARB: A software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
63. W. Manz, R. Amann, W. Ludwig, M. Wagner, K.-H. Schleifer, Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria: Problems and solutions. *Syst. Appl. Microbiol.* **15**, 593–600 (1992).