

Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell

Yuichi Hongoh^{*†}, Vineet K. Sharma[‡], Tulika Prakash[‡], Satoko Noda^{*}, Todd D. Taylor[‡], Toshiaki Kudo^{*}, Yoshiyuki Sakaki[‡], Atsushi Toyoda^{†‡}, Masahira Hattori^{‡§}, and Moriya Ohkuma^{*}

^{*}Environmental Molecular Biology Laboratory, RIKEN, Saitama 351-0198, Japan; [‡]Genomic Sciences Center, RIKEN, Kanagawa 230-0045, Japan; and [§]Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, Chiba 277-8561, Japan

Communicated by Roy H. Doi, University of California, Davis, Davis, CA, February 11, 2008 (received for review October 30, 2007)

Termites harbor a symbiotic gut microbial community that is responsible for their ability to thrive on recalcitrant plant matter. The community comprises diverse microorganisms, most of which are as yet uncultivable; the detailed symbiotic mechanism remains unclear. Here, we present the first complete genome sequence of a termite gut symbiont—an uncultured bacterium named Rs-D17 belonging to the candidate phylum Termite Group 1 (TG1). TG1 is a dominant group in termite guts, found as intracellular symbionts of various cellulolytic protists, without any physiological information. To acquire the complete genome sequence, we collected Rs-D17 cells from only a single host protist cell to minimize their genomic variation and performed isothermal whole-genome amplification. This strategy enabled us to reconstruct a circular chromosome (1,125,857 bp) encoding 761 putative protein-coding genes. The genome additionally contains 121 pseudogenes assigned to categories, such as cell wall biosynthesis, regulators, transporters, and defense mechanisms. Despite its apparent reductive evolution, the ability to synthesize 15 amino acids and various cofactors is retained, some of these genes having been duplicated. Considering that diverse termite-gut protists harbor TG1 bacteria, we suggest that this bacterial group plays a key role in the gut symbiotic system by stably supplying essential nitrogenous compounds deficient in lignocelluloses to their host protists and the termites. Our results provide a breakthrough to clarify the functions of and the interactions among the individual members of this multilayered symbiotic complex.

gut bacteria | insect | phi29 | symbiosis

The termite gut harbors 10^6 – 10^8 microorganisms comprising >300 species of protists, bacteria, and archaea (1–5). These are mostly unique to termites and are essential, as a highly structured symbiotic community, for the host to survive on recalcitrant plant matter (1, 6–9). Although this symbiosis has long been attracting researchers for both basic and applied interests, the complexity and formidable unculturability of the gut microbiota have hampered clarification of the symbiotic mechanism. Among the as-yet-uncultivable gut symbionts, bacteria belonging to the candidate phylum Termite Group 1 (TG1) are common and often predominate in the gut microbial communities (8, 10, 11). Recently, these TG1 bacteria have been identified as intracellular symbionts of diverse cellulolytic gut protists (10, 12, 13). Because of its predominance, commonness, and specific localization confined to the gut protist cells, we assumed that the TG1 bacteria play a key role in the termite gut symbiotic system. However, TG1 is one of the ≈ 40 candidate phyla without isolated representatives (Fig. 1a) (5, 14), and no physiological information have been obtained thus far.

In the present study, we aimed to acquire a complete genome sequence of the TG1 bacteria to clarify their functions, which will provide a breakthrough to disentangle the complicated symbiotic web. We targeted phylotype Rs-D17, which is found specifically within the cells of the cellulolytic flagellate *Trichonympha agilis* in the gut of the termite *Reticulitermes speratus* (3, 10) (Fig. 1 b–d). Approximately 4,000 Rs-D17 cells are housed

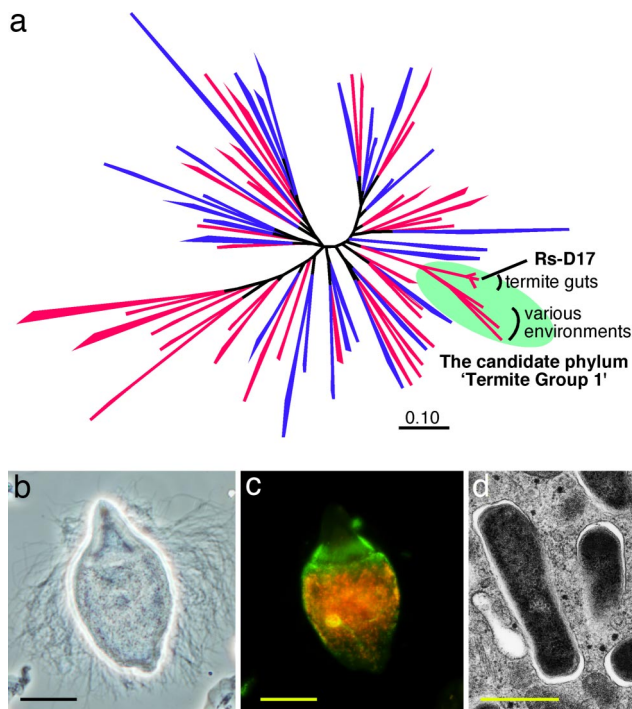


Fig. 1. Phylogenetic position, localization and morphology of phylotype Rs-D17. (a) A maximum-likelihood tree of the domain Bacteria based on 16S rRNA sequences. One hundred fifty-eight sequences from 66 phylum-level clusters were used for the analysis as described in ref. 5. Phyla with described isolates are shown in blue and the others are shown in red. (b) Phase-contrast microscopy of the parabasalid flagellate *T. agilis*. (c) Fluorescence *in situ* hybridization analysis, using Bacteria-universal (6-carboxyfluorescein-labeled, green) and Rs-D17-specific (Texas red-labeled, red) probes performed as described in ref. 10. (d) Transmission electron microscopy of Rs-D17 cells within a *T. agilis* cell, which was performed as described in ref. 10. The Rs-D17 cells have Gram-negative-type cell walls. (Scale bars: b and c, 10 μm ; d, 0.5 μm .)

within each host cell, accounting for 4%, in total, of all of the prokaryotic cells in the gut (10). TG1 bacteria as a whole occupy $\approx 10\%$ of the gut prokaryotic population in *R. speratus* (3, 8, 10). To obtain a sufficient amount of DNA for the genome analysis,

Author contributions: Y.H. and V.K.S. contributed equally to this work; Y.H., Y.S., A.T., M.H., and M.O. designed research; Y.H. and A.T. performed research; Y.H., V.K.S., T.P., S.N., and A.T. analyzed data; and Y.H., T.D.T., T.K., M.H., and M.O. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the DNA Data Bank of Japan [accession nos. AP009510 (chromosome), AP009511–3 (plasmids), and AB360878–AB360905 (others)].

[†]To whom correspondence may be addressed. E-mail: yhongoh@postman.riken.go.jp or toyoda@gsc.riken.jp.

This article contains supporting information online at www.pnas.org/cgi/content/full/0801389105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

we applied an isothermal whole genome amplification (WGA) technique (15) to the target bacteria housed within a single host protist cell. Here, we report the first complete genome sequence of a termite gut symbiont, which belongs to a hitherto uncharacterized phylum, TG1, in the domain bacteria.

Results

Purity of Rs-D17 Bacteria from a Single Host Protist Cell. Because the host protists *T. agilis* are also as yet uncultivable and are expected to comprise heterogeneous strains, we physically isolated only a single host cell by micromanipulation so as to minimize the genomic variation of the intracellular symbionts. Further, to obtain Rs-D17 cells as purely as possible, we ruptured the posterior part of the host cell in buffer and collected a small portion of the bacterial cells that leaked out [supporting information (SI) Fig. S1]. This procedure was necessary because the host *T. agilis* cells harbor other intracellular and extracellular symbionts mostly localized anteriorly (10) (Fig. 1c). The collected cells were directly subjected to isothermal whole genome amplification (WGA) (see *Materials and Methods*).

The purity of the amplified sample was checked by 16S rRNA clone analysis. Of 89 clones sequenced, 84 (94%) were affiliated to phylotype Rs-D17 with almost no sequence variation. The genomic variation within phylotype Rs-D17 was further investigated by clone analysis of the internal transcribed spacer (ITS) between the 16S and 23S rRNA genes. Of 48 clones sequenced, 40 had an identical sequence (771 bp), seven clones had a single base insertion to the sequence, and one clone had a single base mismatch. The insertions in the seven clones were identical: an additional guanine (G) to a 9-bp G stretch. These suggested that phylotype Rs-D17 consists of strains with only slight genome variations (genomovars) within a single host cell. In contrast, during the verification step of regions containing pseudogenes and duplicated genes (see below), significant differences were found in the Rs-D17 genomes between host cells, particularly in the length of their noncoding regions (0–15.8% difference) (see *SI Text*).

General Features of the Rs-D17 Genome. From the amplified sample, we successfully reconstructed a circular chromosome consisting of 1,125,857 bp (Fig. 2). The chromosome was completely covered by Sanger sequencing of shotgun clones and almost entirely covered by 454 pyrosequencing (giving 97.7% coverage and 42 × redundancy). The sequence has little ambiguity or variation, and the high sequence redundancy excluded chimeras produced during WGA (16). The chromosome encodes 761 putative protein-coding sequences (CDSs), one operon of rRNA genes, and 45 tRNA genes (Table 1). The tRNAs exhibited 42 anti-codon sequence variations corresponding to codons for all of the 20 aa (Dataset S1). Three circular plasmids (11,650 bp, 5,701 bp, and 5,362 bp), which encode only CDSs showing no significant homology to sequences in any databases, were additionally reconstructed (Fig. S2). Ninety percent of the fragments produced by 454 pyrosequencing were assigned to these Rs-D17 chromosome and plasmids. The replication origin of the chromosome was not readily determinable because of the absence of a clear transition point of the guanine/cytosine (GC) skew curve (Fig. 2). The genome additionally contains surplus pseudogenes, up to 121, which is comparable with the genomes of the intracellular parasites *Rickettsia* (17) (Dataset S2). One of the pseudogenes is that for *DnaA*, which regulates the chromosome replication. We set the genome sequence number 1 on this *dnaA* pseudogene. Several regions containing pseudogenes and duplicated genes were verified by PCR directly from single host cells, bypassing WGA, to confirm that they are not artifacts produced during WGA (see *SI Text*).

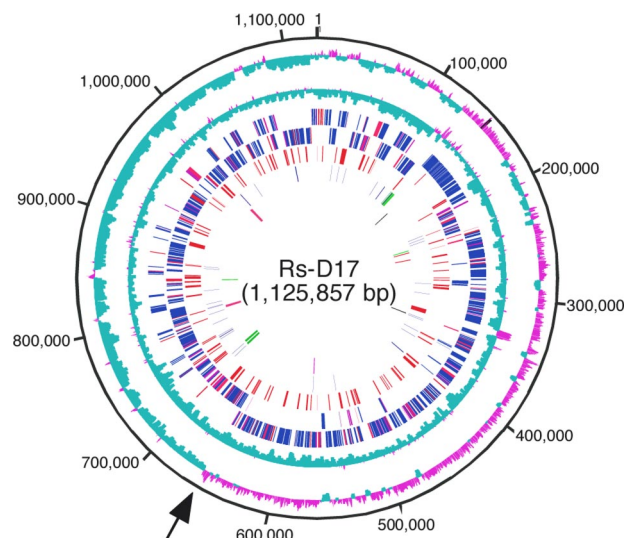


Fig. 2 Circular representation of the Rs-D17 chromosome. The concentric rings denote the following features (from outside): (i) Numbered base pair starting with base 1 of the *dnaA* pseudogene. (ii) GC skew as calculated by $(G - C)/(G + C)$. Pink indicates values >0 ; green–blue, <0 . (iii) G + C content. Pink indicates values $>50\%$; green–blue, $<50\%$. (iv) CDSs present on the forward strand (+). (v) CDSs present on the reverse strand (–). Blue, function assigned; purple, putative; green, conserved; dark red, predicted only. (vi) Pseudogenes. (vii) Green, rRNA genes; blue, tRNA genes; and red, CRISPR. (viii) Duplicated regions. Six pairs are colored differentially. Red, E1 and E2; green, F1 and F2; pink, G1 and G2; black, P1 and P2; light green, Q1 and Q2; purple, R1 and R2 (see also Fig. S4). Arrow indicates the replication termination site. In Silico Molecular Cloning software was used for constructing the map.

Energy Metabolism. The predicted metabolic pathways of Rs-D17 are shown in Fig. 3. Rs-D17 retains pathways for glycolysis, gluconeogenesis, and nonoxidative pentose phosphate biosynthesis but lacks the oxidative pentose phosphate pathway and most components of the tricarboxylic acid cycle. Glucose 6-phosphate (Glc-6P) in the host cell, which is imported by using a UhpC homolog, appears to be the major carbon and energy source. Thus, Rs-D17 spares its own adenosine 5'-triphosphates (ATPs) to phosphorylate glucose as in the intracellular parasite *Chlamydia pneumoniae* (18). The energy production of Rs-D17 solely depends on substrate-level phosphorylation through fermentation of sugar to acetate. Rs-D17 lacks all of the components of a respiratory chain or of known substitutive mechanisms. Therefore, the F_0F_1 -type ATPase is likely used to create the proton motive force and not for ATP synthesis. Nicotine amide dinucleotides (NADs) are recovered by producing hydrogen, ethanol and possibly D-lactate. The genome lacks genes for cytochrome oxidase, catalase and superoxide dismutase. These features clearly indicate that Rs-D17 is strictly an anaerobic fermenting bacterium. This is consistent with the physicochemical conditions inside the termite hindgut (19).

Rs-D17 may also import fructose or other sugars by using a phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS). However, because the HPr(Ser) kinase gene *hprK* is a pseudogene, regulation of carbohydrate assimilation does not work via this PTS. Interestingly, the allosteric pyruvate kinase in the glycolytic pathway lacks the carboxyl-terminal domain that binds to its activator. These findings exemplify the absence of most regulatory proteins or functions in Rs-D17.

Cell Wall Biogenesis and Defense Mechanisms. Many of the pseudogenes were categorized for cell wall biogenesis (Fig. 4). Although the pathways for peptidoglycan biosynthesis appear functional, the lipopolysaccharide (LPS) biosynthetic pathway

Table 1. General features of the Rs-D17 genome.

	Size, bp	GC content, %	Putative CDS	Function assigned	Conserved	Predicted only	CDS density, %	Average CDS size, bp	Pseudogenes	tRNA genes	rRNA operon
Chromosome	1,125,857	35.2	761	622	80	59	66.5	984	121	45	1
Plasmid											
pTGRD1	11,650	34.3	9	0	0	9	46.4	601	0	0	0
pTGRD2	5,701	35.4	3	0	0	3	60.2	1145	0	0	0
pTGRD3	5,362	32.6	3	0	0	3	37.0	656	0	0	0

has been collapsed, with 25 pseudogenes remaining. Several related genes were also found disrupted, including *lola* and *lolC* for lipoprotein transport and *pal* and *tolB* for outer membrane integration. Thus, Rs-D17 possesses a weak cell wall, which is obviously a result of adaptation to intracellular life protected by the host cell. Similarly, this bacterium has accumulated pseudogenes in defense mechanisms, such as restriction-modification systems and a multidrug exporter. Remarkably, only two of 26 restriction-modification systems remain intact; the others are merely pseudogenes or completely lack the restriction component. The presence of many remnants of restriction-modification systems and an apparently intact CRISPR system (20) implies that the ancestral genome may have suffered from invasion of exogenous genetic components, although only a few traces of phages in this genome remain.

Biosynthesis of Amino Acids, Cofactors, and Nucleotides. Despite its apparent reductive evolution, the Rs-D17 genome retains the ability to synthesize 15 amino acids and various cofactors (Fig. 3 and Fig. S3). The chromosome even contains duplicated genes for AroH, HisC, and PheA involved in the aromatic amino acid

biosynthesis and CoaD for CoA biosynthesis (Figs. S3 and S4). However, because the glutamine synthetase gene *glnA* is a pseudogene, Rs-D17 must import glutamine. With the loss of GlnA, the ability to incorporate ammonia has become limited in Rs-D17, and correspondingly, the ammonium (or ammonia) transporter gene *amtB* was found to be a pseudogene. The purine and pyrimidine biosynthetic pathways remain intact. Their precursor, 5-phospho- α -D-ribose 1-diphosphate (PRPP), is synthesized by ribose-phosphate diphosphokinase, whereas the phosphopentomutase gene in another biosynthetic pathway for PRPP was found to be disrupted. This exemplifies streamlining adaptation in this genome.

Genes Unique to Rs-D17. Of 761 putative CDSs on the chromosome, 59 (7.8%) showed no homology to any sequences in publicly available databases. Interestingly, one of them, gene TGRD.1 on the chromosome shares 56–76% amino acid identity with TGRD.2 and genes on each of the three plasmids, TGRD.P1-1, TGRD.P2-1, and TGRD.P3-1 (Dataset S3). These genes may play an important role in this bacterium and strongly suggest that the three plasmids derived from the Rs-D17 bac-

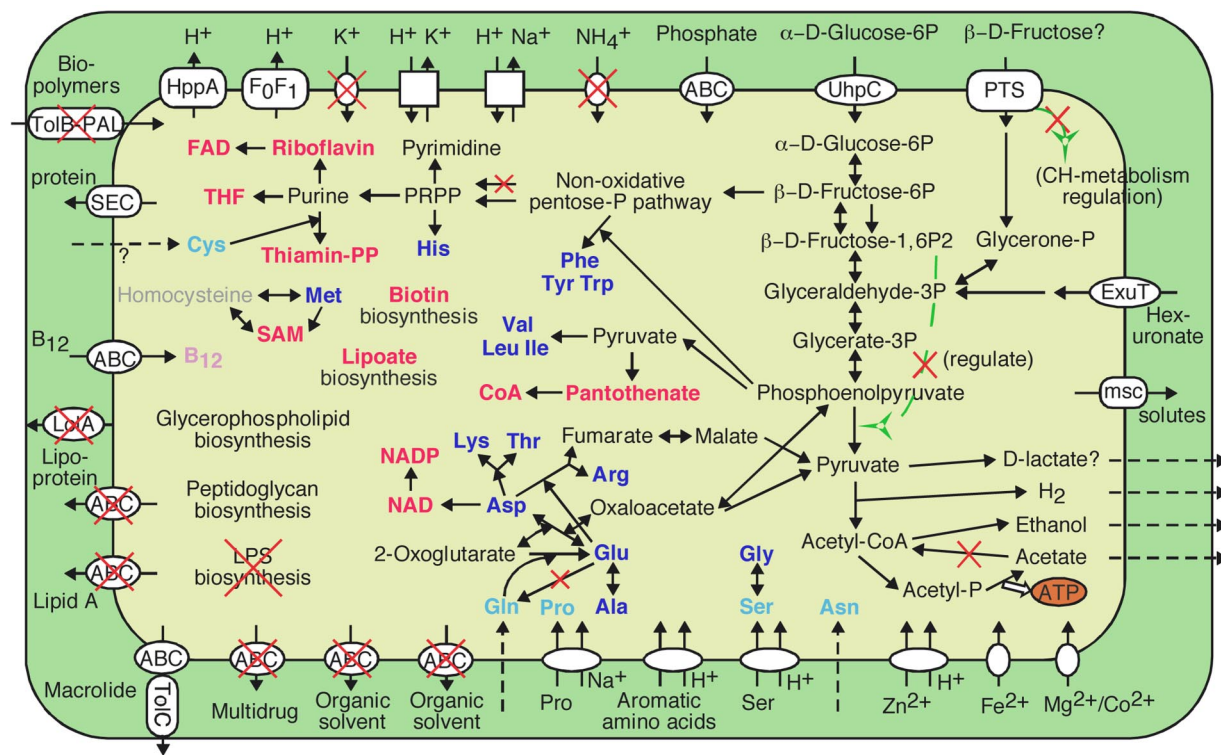


Fig. 3. Predicted metabolic pathways of phylotype Rs-D17. Periplasm (green) and cytoplasm (yellow) are shown bounded by outer and inner membranes, respectively. Blue, synthesized amino acids; red, cofactors. Compounds that must be imported are shown in pale colors. Nonfunctional pathways and transporters comprising pseudogenes are marked with red X's. Arrows with broken lines indicate diffusion or transport via some unidentified apparatus. ABC, ATP-binding cassette type transporter; msc, mechanosensitive channel; FAD, flavin adenine dinucleotide; SAM, S-adenosylmethionine; THF, tetrahydrofolate.

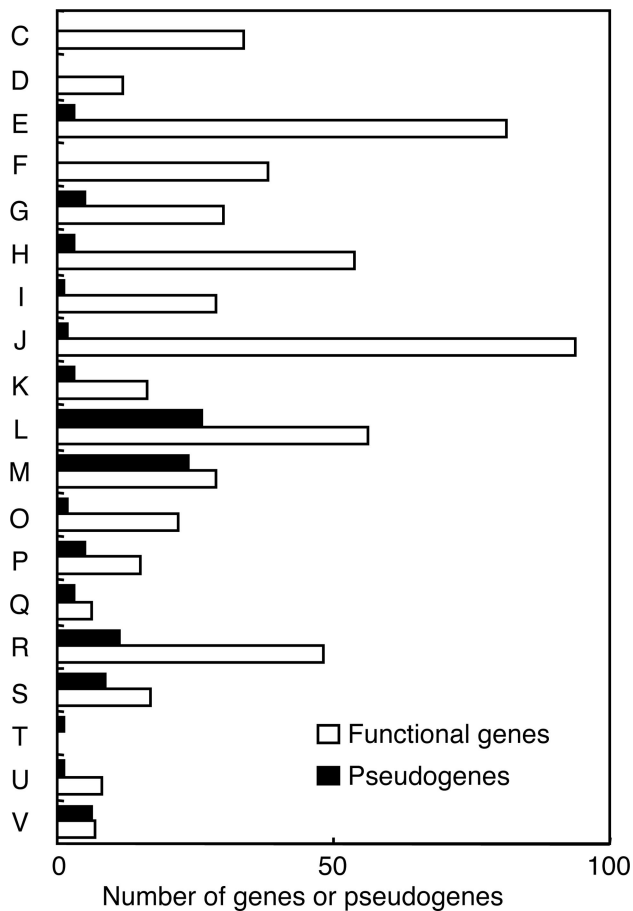


Fig. 4. Cluster of orthologous group of proteins (COGs) analysis of functional genes and pseudogenes in the Rs-D17 genome. C, energy production and conversion; D, cell cycle control, cell division chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; M, cell wall/membrane/envelope biogenesis; O, posttranslational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport, and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms.

teria. When the Rs-D17 genome was compared with those of the other known intracellular symbionts, 124 were unique to Rs-D17 and 430 were shared with one or more of the symbionts (Dataset S4). The majority of the unique orthologs were classified into the comparison of orthologous group of proteins (COG) categories [energy production and conversion], [replication, recombination and repair], [inorganic transport and metabolism], [general function predicted only], and [function unknown] (Fig. S5). These include genes for key enzymes in anaerobic metabolisms, such as iron hydrogenase (HydABC), pyruvate-ferredoxin oxidoreductase, and pyrophosphate-dependent phosphofructokinase. The Rs-D17 genome retains abundant genes for DNA recombination and repairs such as *xerCD*, *recA*, *ruvABC*, *uvrABCD*, *mutLS* and others, as an exception among the intracellular symbionts (Dataset S5).

Discussion

This complete genome from a termite gut symbiont revealed the importance of provision of amino acids and cofactors by gut

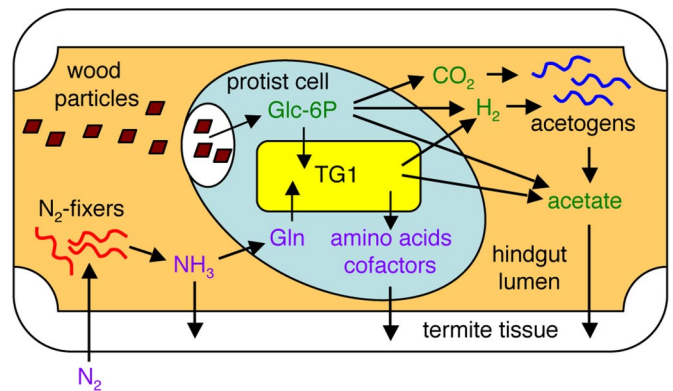


Fig. 5. Proposed schematic view of the function of TG1 bacteria in termite gut. Nitrogen is initially incorporated by some nitrogen-fixing bacteria, such as spirochetes (22). Ammonia and some amino acids in the hindgut lumen (28) are then imported by protist cells. Within a protist cell, those nitrogen compounds, including glutamine, are then converted into a variety of amino acids and cofactors by the intracellular TG1 bacteria. The protist cells probably digest TG1 cells to use these compounds because an exporting mechanism, such as that discovered in *B. aphidicola* (27), has not been identified in the Rs-D17 genome. The protist cells containing TG1 are then finally digested by termites during nutritional exchange via trophallaxis and by coprophagy among colony members (29). Ingested lignocelluloses are fermented by protists to CO_2 , H_2 , and acetate and by the symbiotic TG1 bacteria to H_2 and acetate. CO_2 and H_2 are then converted to acetate by reductive acetogenesis by acetogens, such as spirochetes (30). Thus, ingested carbon sources are finally absorbed in the form of acetates by termites.

bacteria; the Rs-D17 genome retains abundant genes for biosynthesis of those compounds despite its reduced genome size. This coincides with the ecology of wood-feeding termites, which thrive only on nitrogen-deficient food, i.e., lignocelluloses. The gut microorganisms must be responsible for both fixation of atmospheric dinitrogen (21–23) and supply of all of the essential nitrogenous compounds that termites cannot synthesize. The recycle of uric acid by gut bacteria has also been reported, which minimizes the loss of nitrogen (24). Moreover, a cellulolytic termite-gut protist reportedly requires dead bacterial cells of a specific strain as the nitrogen source for its growth when cultured axenically (25). Considering that diverse gut protists harbor TG1 phylotypes specific to the respective host protist species (10, 12, 13), we suggest that the cellulolytic protists have exploited them as essential symbionts that stably supply adequate repertoires of nitrogenous compounds required by each host species.

The presence of the *glnA* pseudogene in the Rs-D17 genome in turn implies the contribution of the host protist to its symbionts. Although the detailed nutritional properties of the termite gut protists are unknown, *glnA* and genes for cysteine and pyridoxine synthesis, which Rs-D17 lacks, have been found among the few gene repertoires related to amino acid and cofactor biosynthesis in a complementary DNA library of the protists from *R. speratus* gut (26). This is indicative of the complementary metabolism between the host cell and symbionts as in aphid-*Buchnera* symbiosis reported (27). Thus, the host protists likely provide the nitrogenous compounds that the symbiotic TG1 bacteria cannot synthesize, in addition to the safe habitat and the phosphorylated glucose. Then, these symbiotic complexes contribute to their host termites both by fermenting lignocelluloses and by supplying essential nitrogen sources (Fig. 5). Although one might expect that intracellular symbionts of termite gut protists may participate in hydrolyzing lignocellulose, no known related enzymes were found in the genome of Rs-D17. This is consistent with a report on the termite gut protists *Trichonympha sphaerica* that were exceptionally cultivated (31) although possibly not axenic (25). It was observed in this

previous study that symbionts-depleted *Trichonympha* cells were able to ferment cellulose to acetate, hydrogen, and carbon-dioxide (31).

Our strategy, acquiring a complete genome from a small number of almost pure genomovars, enabled the precise identification of pseudogenes and duplicated genes without ambiguity. Because this genome sequence is from the as-yet-uncultivable phylum TG1, there have been no reference genomes as confirmed by phylogenetic analysis based on the concatenated ribosomal protein sequences and by BLAST-based analysis of the putative CDSs (Fig. S6 and Dataset S6, respectively). Nevertheless, the surplus pseudogenes, which evidently indicate the lost functions, enabled us to elucidate how the adaptation process has been progressing in this bacterium. The Rs-D17 genome shares the known characteristics of obligately intracellular symbionts, such as small genome size, small number of RNA genes, streamlining adaptation, few repertoires of available carbon and energy sources, a weak cell wall and loss of regulators, defense mechanisms, and transporters. Besides, the retained ability to synthesize various amino acids and cofactors has been found to be a common feature of mutually symbiotic ones (27, 32–36) (Fig. S7). Thus, the present study demonstrated that a phylogenetically distinct intracellular symbiont of strictly anaerobic properties and of a unicellular eukaryotic host undergoes similar adaptation process with those of the known proteobacterial symbionts, which are aerobic and are harbored by multicellular eukaryotic hosts.

One of the known features occasionally found among the proteobacterial intracellular symbionts is absence of *dnaA* gene (32–36). Because *dnaA* was found to be a pseudogene in the Rs-D17 genome, the loss of DnaA protein is considered to be a common trait in intracellular symbionts irrespective of their phylogenetic lineages. This might imply control of replication by the host cell or might simply be a consequence of streamlining adaptation by using recombination-dependent replication of the chromosome as a substitutive mechanism (37). In any cases, Rs-D17 appears to have lost its ability to survive outside the host cell, indicating that the host protist has successfully domesticated the TG1 bacteria as highly specialized suppliers of the nitrogenous compounds. This protist-TG1 symbiosis seems also beneficial to the host termites because the net production efficiency of amino acids and cofactors of the TG1 bacteria with a reduced genome is probably higher than those of free-swimming gut bacteria.

The gene duplication on the chromosome is exceptional among the known intracellular bacteria. The duplicated genes for the amino acid and cofactor biosynthesis may enhance the ability to produce them as in *Buchnera aphidicola*, the primary symbionts of aphids that have plasmids encoding amplified genes for tryptophan and leucine biosynthesis (27). Interestingly, six regions containing 10 genes in total were found duplicated in the chromosome (Fig. 2), which share 100% sequence identity with their respective duplicates (Fig. S4). This clearly indicates that these duplication events have occurred recently. Taken together with the presence of numerous pseudogenes, it is likely that Rs-D17 is still in a dynamic process of adaptation as an intracellular symbiont of the gut protist. The gene duplications may have occurred by action of tyrosine recombinase XerCD and may have been maintained by the relatively abundant DNA repair systems (Dataset S5).

Reconstruction of a complete genome from uncultivable prokaryotes had previously been impossible, unless a large number of homogeneous target cells could be collected, as in the case of primary intracellular symbionts of insects. However, the present study demonstrates that only $\approx 10^3$ prokaryotic cells are sufficient to practically yield a complete genome sequence from a natural environment. The acquisition of a complete genome is essential for elucidating the precise functions of and interactions

among individual members of uncultivable microbial communities. Although conventional metagenomic analysis can provide a comprehensive view on a complex microbial community, it generally supplies poor information on symbiotic relationships among the community members. Our strategy will provide a breakthrough to clarify the detailed mechanism of the termite gut symbiosis, which is a model of multilayered symbiotic communities and may contain clues to industrial application, such as biomass energies.

Materials and Methods

Sample Collection. The wood-feeding termites *R. speratus* (Isoptera; Rhinotermitidae) were collected in Saitama, Japan. After being fed with cellulose powder for 3 days, the entire gut was removed from a worker termite and dissected in buffer (3). A single cell of *T. agilis* (Parabasalia: Trichonymphida) was physically isolated by using a micromanipulator (Eppendorf; Transferman NK2). After several washes, the protist cell was transferred to 1% Nonidet P-40 in buffer. Bacterial cells that leaked out were collected with a glass capillary (15- μ m diameter). WGA was performed with a Repli-g Midi kit (Qiagen) for 8 h.

Purity Check of WGA Sample. The purity of the amplified sample was checked by clone analysis of 16S rRNA gene. PCR was performed with primers 27F (5'-AGRGTGGATYMTGGCT-CAG) and 1492R (5'-GGHTACCTTGTTACGACTT), using a proof-reading DNA polymerase, Phusion (Finnzyme). The PCR program comprised an initial denaturation step at 98°C for 45 s, 18 cycles of denaturation (10 s at 98°C), annealing (1 min at 50°C) and extension (3 min at 72°C), and a final 10-min extension at 72°C. The product was purified by using a Monofas DNA purification kit (GL Science) and cloned by using a Zero Blunt TOPO PCR cloning kit (Invitrogen). For clone analysis of ITS, primers D17-TTS-FW (5'-AGTCGCCTAAGTTATGGTTG) and D17-ITS-RV (5'-AAGGCATTCACCATATGCAC) were used for PCR as described above with 15 cycles.

Genome Sequencing and Annotation. A hybrid sequencing approach was performed by using ABI 3730 sequencers and a GS20 pyrosequencing system (454; Life Sciences). Detailed sequencing procedures are described in *SI Text*. CDSs and tRNA genes were predicted and annotated by using the iMetaSys pipeline developed at RIKEN, as described in detail in *SI Text*. These computational predictions and annotations were then manually curated. Only sequences with ≥ 100 aa were regarded as putative CDSs unless significant homology was detected for the shorter ones. Pseudogenes were identified manually according to any of the following criteria: (i) corrections of mutations (frameshifts and/or nonsense) restored functional domains or conserved regions, (ii) lack of the initiation or termination codon, and (iii) lack of a large N-terminal or C-terminal conserved region. However, when a split or truncated CDS retained an intact functional domain, it was regarded as a functional CDS.

Verification of Pseudogenes and Duplicated Genes. PCR directly from single *T. agilis* cells, bypassing WGA, was performed to verify presence of pseudogenes and duplicated genes. PCR was performed for each host cell fixed in acetone, using Phusion with the following conditions: an initial denaturation step at 95°C for 2 min, 35 cycles of denaturation (10 s at 98°C), annealing (30 s at 52 or 60°C) and extension (2–4 min at 72°C), and a final 10-min extension at 72°C. PCR primers designed for these purposes are shown in Dataset S7, Dataset S8, and Fig. S4 (see also *SI Text*).

ACKNOWLEDGMENTS. We thank all of the technical staff, present and past, of the Sequencing Technology Team at RIKEN Genomic Sciences Center and the Environmental Molecular Biology lab for their assistance. We thank the

following RIKEN Sequence Technology Team members: Tomoyuki Aizu, Fumiwo Ejima, Tomoko Hasegawa, Hinako Ishizaki, Mikiko Iwatsu, Kaoru Kaida, Miho Kiyooka, Maho Naka, Saori Nakagawa, Emi Oomori, Rie Ryusui, Minami Shimizu, Noriko Yamamoto, Yuko Yamamoto, Takujiro Katayama, Wang Yingchun, Erika Iioka, Sayaka Kato, Keiko Kawano, Noriko Shiraishi, Taeko Takamori, Masako Tanaka, and Yoshiko Usami. We also thank the RIKEN Environmental Molecular Laboratory members who assisted with the experiments: Fumie Ohnuma, Tomoyuki Sato, and Jun-ichi

Inoue. This work was supported in part by a grant for the President's Discretionary Fund of RIKEN, a Special Fund for RIKEN Genomic Sciences Center, grants for the Bioarchitect Research Program and the Eco Molecular Science Research Program from RIKEN, a Discovery Research Institute Research Grant (to Y.H.) from RIKEN, and Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research Grants 18687002 (to Y.H.) and 19380055 (to M.O.).

- Nakajima H, Hongoh Y, Usami R, Kudo T, Ohkuma M (2005) Spatial distribution of bacterial phylotypes in the gut of the termite *Reticulitermes speratus* and the bacterial community colonizing the gut epithelium. *FEMS Microbiol Ecol* 54:247–255.
- Inoue T, Murashima K, Azuma J-I, Sugimoto A, Slaytor M (1997) Cellulose and xylan utilisation in the lower termite *Reticulitermes speratus*. *J Insect Physiol* 43:235–242.
- Hongoh Y, Ohkuma M, Kudo T (2003) Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiol Ecol* 44:231–242.
- Tokura M, Ohkuma M, Kudo T (2000) Molecular phylogeny of methanogens associated with flagellated protists in the gut and with the gut epithelium of termites. *FEMS Microbiol Ecol* 33:233–240.
- Hongoh Y, et al. (2006) Phylogenetic diversity, localization, and cell morphologies of members of the candidate phylum TG3 and a subphylum in the phylum *Fibrobacteres*, recently discovered bacterial groups dominant in termite guts. *Appl Environ Microbiol* 72:6780–6788.
- Cleveland LR (1924) The physiological and symbiotic relationships between the intestinal protozoa of termites and their host, with special reference to *Reticulitermes flavipes* Kollar. *Biol Bull* 46:178–227.
- Eutick ML, Veivers PC, O'Brien RW, Slaytor M (1978) Dependence of the higher termite, *Nasutitermes exitiosus* and the lower termite, *Coptotermes lacteus* on their gut flora. *J Insect Physiol* 24:363–368.
- Hongoh Y, et al. (2005) Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Appl Environ Microbiol* 71:6590–6599.
- Yang H, Schmitt-Wagner D, Stingl U, Brune A (2005) Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ Microbiol* 7:916–932.
- Ohkuma M, et al. (2007) The candidate phylum "Termite Group 1" of bacteria: Phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol Ecol* 60:467–476.
- Ohkuma M, Kudo T (1996) Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Appl Environ Microbiol* 62:461–468.
- Stingl U, Radek R, Yang H, Brune A (2005) "Endomicrobia": Cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol* 71:1473–1479.
- Ikeda-Ohtsubo W, Desai M, Stingl U, Brune A (2007) Phylogenetic diversity of "Endomicrobia" and their specific affiliation with termite gut flagellates. *Microbiology* 153:3458–3465.
- Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180:4765–4774.
- Dean FB, et al. (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 99:5261–5266.
- Lasken RS, Stockwell TB (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnology* 7:19.
- Blanc G, et al. (2007) Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet* 3:e14.
- Schwöppe C, Winkler HH, Neuhaus HE (2002) Properties of the glucose-6-phosphate transporter from *Chlamydia pneumoniae* (HPTcp) and the glucose-6-phosphate sensor from *Escherichia coli* (UhpC). *J Bacteriol* 184:2108–2115.
- Brune A (1998) Termites gut: The world's smallest bioreactor. *Trends Biotech* 16:16–21.
- Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 1:e60.
- Breznak JA, Brill WJ, Mertins JW, Coppel HC (1973) Nitrogen fixation in termites. *Nature* 244:577–580.
- Lilburn TG, et al. (2001) Nitrogen fixation by symbiotic and free-living spirochetes. *Science* 292:2495–2498.
- Slaytor M (2000) in *Termites: Evolution, Sociality, Symbioses, Ecology*, eds Abe T, Bignell DE, Higashi M (Kluwer Academic, Dordrecht, The Netherlands), pp 307–332.
- Potrikus CJ, Breznak JA (1980) Anaerobic degradation of uric acid by gut bacteria of termites. *Appl Environ Microbiol* 40:125–132.
- Odelson DA, Breznak JA (1985) Nutrition and growth characteristics of *Trichomitus termopsidis*, a cellulolytic protozoan from termites. *Appl Environ Microbiol* 49:614–621.
- Todaka N, et al. (2007) Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. *FEMS Microbiol Ecol* 59:592–599.
- Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 407:81–86.
- Slaytor M, Chappell (1994) Nitrogen metabolism in termites. *Comp Biochem Physiol* 107B:1–10.
- Fujita A, Shimizu I, Abe T (2001) Distribution of lysozyme and protease, and amino acid concentration in the gut of a wood-feeding termite, *Reticulitermes speratus* (Kolbe): Possible digestion of symbiont bacteria transferred by trophallaxis. *Physiol Entomol* 26:116–123.
- Leadbetter JR, Schmidt TM, Graber JR, Breznak JA (1999) Acetogenesis from H₂ plus CO₂ by spirochetes from termite guts. *Science* 283:686–689.
- Yamin MA (1981) Cellulose metabolism by the flagellate *Trichonympha* from a termite is independent of endosymbiotic bacteria. *Science* 211:58–59.
- Nakabachi A, et al. (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 314:267.
- Gil R, et al. (2003) The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proc Natl Acad Sci USA* 100:9388–9393.
- Akman L, et al. (2002) Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* 32:402–407.
- Degnan PH, Lazarus AB, Wernegreen JJ (2005) Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Res* 15:1023–1033.
- Wu D, et al. (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* 4:e188.
- Kogoma T (1997) Stable DNA replication: Interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* 61:212–238.